Optimal *Sequential* Resource Sharing and Exchange in *Multi*-Agent Systems

Yuanzhang Xiao

Advisor: Prof. Mihaela van der Schaar

Electrical Engineering Department, UCLA

Ph.D. defense, March 3, 2014

Research agenda

Sequential resource sharing/exchange in ***multi***-agent systems

- Sequential:
  - Agents interact over a long time horizon
  - Agents' current decisions affect future
  - Agents aim to maximize long-term payoffs
  - Different from standard myopic optimization problems
- Multi-agent:
  - Multiple agents influencing each other
  - Different from standard Markov decision processes (MDPs)

New tools and formalisms!

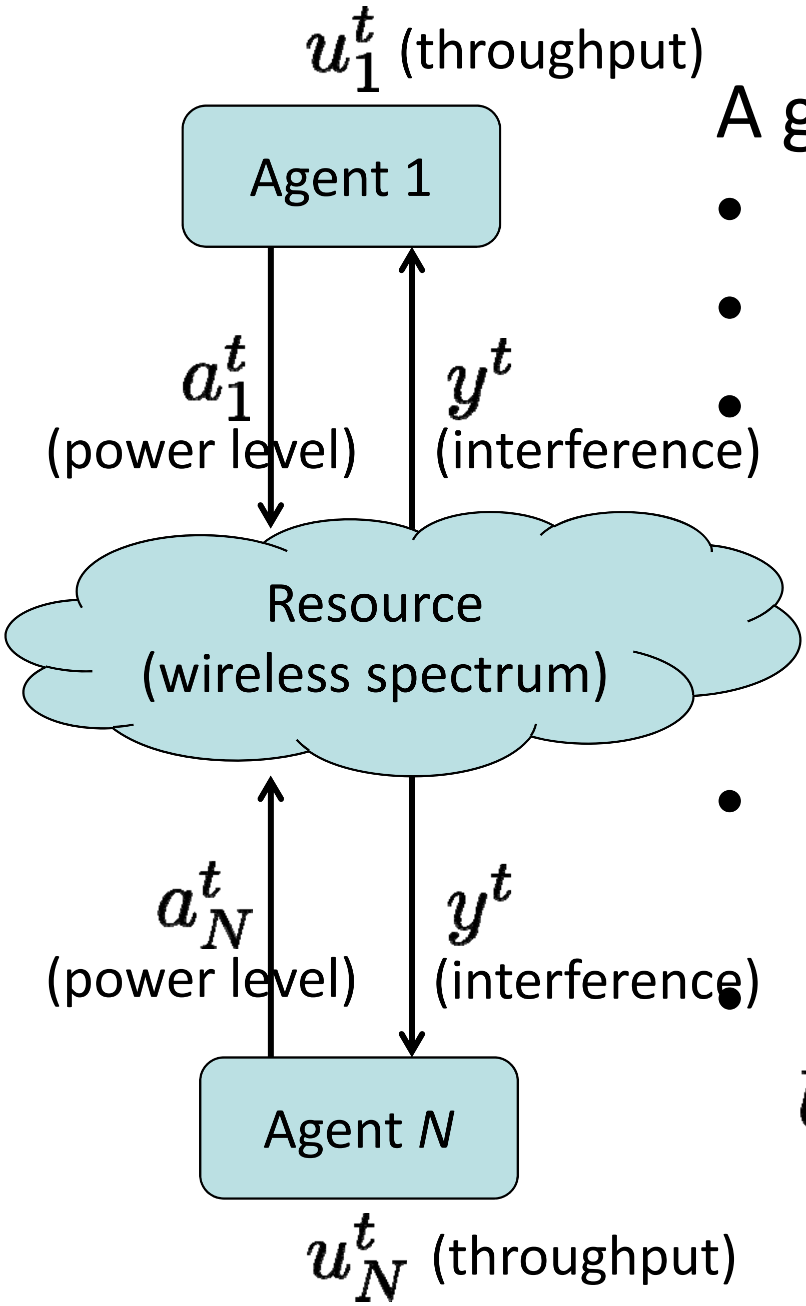
Research dimensions

- Interactions
  - agents interact with all other agents
  - agents interact in pairs
- Externalities
  - one’s action affects the others’ payoffs directly and negatively
  - one’s action affects the others’ payoffs directly and positively
  - one’s action does not affect the others’ payoffs, but is coupled with the others’ actions through constraints
- Monitoring
  - perfect / imperfect
- State
  - none (system stays the same) / public / private
- Deviation-proof
  - no / yes

Resource sharing with strong negative externality

- Interactions
  - **everybody interacts with everybody**
  - agents interact in pairs
- Externalities
  - **one's action affects the others' payoffs directly and negatively**
  - one's action affects the others' payoffs directly and positively
  - one's action does not affect the others' payoffs, but is coupled with the others' actions through constraints
- Monitoring
  - perfect / **imperfect**
- State
  - **none (system stays the same)** / public / private
- Deviation-proof
  - no / **yes**

A general resource sharing problem



A general resource sharing scenario:

- A resource shared by agents 1, ..., *N*
- Time is slotted *t* = 0, 1, 2, ...
- At each time slot *t*:
  1. Agent *i* chooses action a_i^t
  2. Receives monitoring signal y^t
  3. Receives payoff $u_i^t = u_i(a_i^t, \boldsymbol{a}_{-i}^t)$
- Strategy:

$$\pi_i : (y^0, \ldots, y^{t-1}) \mapsto a_i^t$$

- Long-term payoff:

$$U_i(\pi_i, \boldsymbol{\pi}_{-i}) = \mathbb{E}\{(1-\delta)\sum_{t=0}^{\infty} \delta^t u_i^t\}$$

Design optimal resource sharing policies

Design problem:

$$\max_{\boldsymbol{\pi}} \; W(U_1(\boldsymbol{\pi}), \ldots, U_N(\boldsymbol{\pi}))$$ ← Social welfare function

$$s.t. \quad U_i(\boldsymbol{\pi}) \geq \underline{v}_i, \; \forall i \in \mathcal{N}$$ ← Minimum payoff guarantees

$$\boldsymbol{\pi} \text{ is deviation} - \text{proof}$$

Formally, $\boldsymbol{\pi}$ is deviation-proof, if for all $i \in \mathcal{N}$, we have

$$U_i(\pi_i, \boldsymbol{\pi}_{-i}) \geq U_i(\pi_i', \boldsymbol{\pi}_{-i}), \; \forall \pi_i'$$

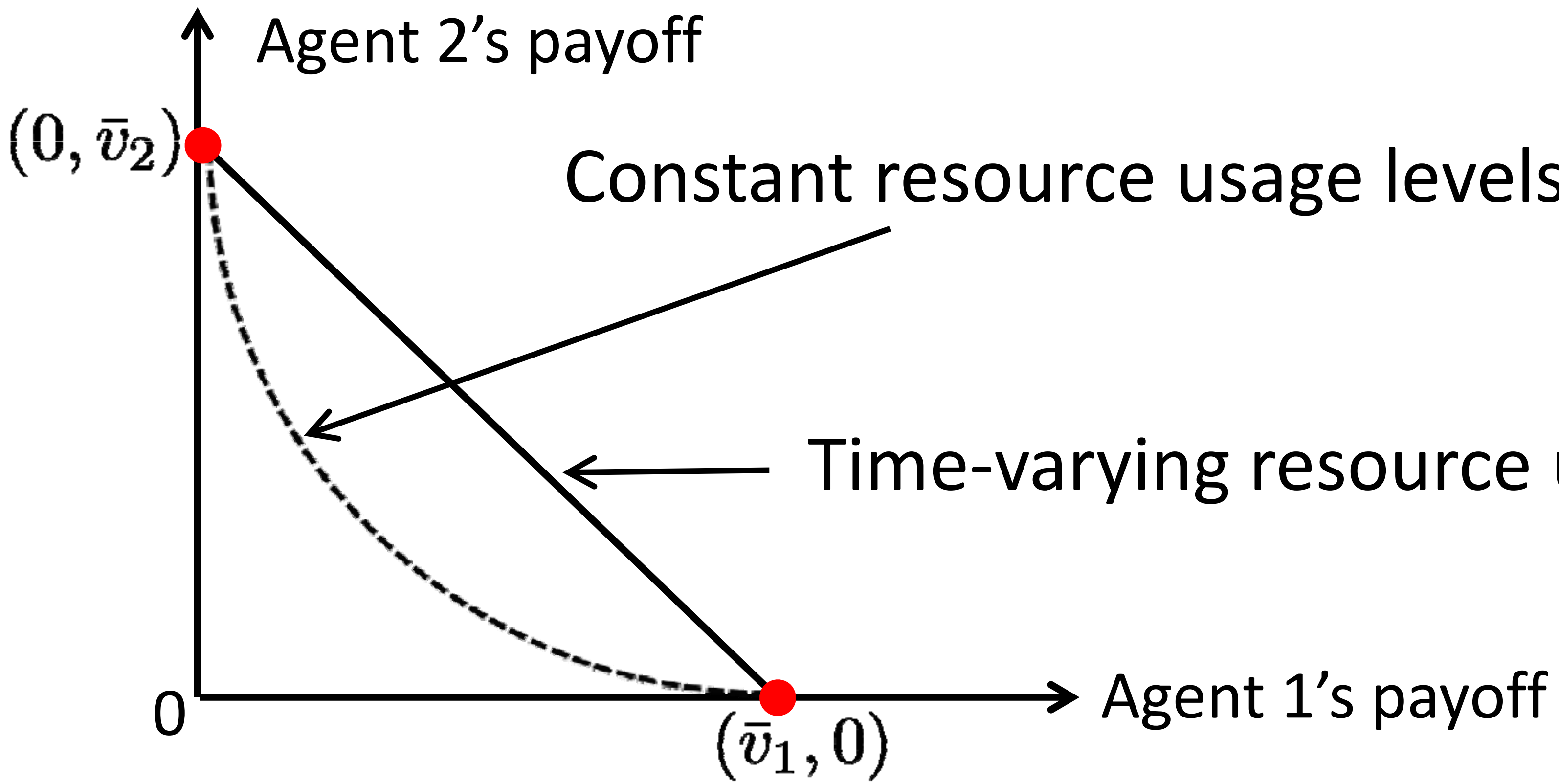
A special (but large) class of problems

Resource sharing with strong negative externalities

Definition of Strong Negative Externalities (Intuitive)

Negative Externalities: $\frac{\partial u_i}{\partial a_j} < 0, \ \forall j \neq i$

Strong Negative Externalities: $\left|\frac{\partial u_i}{\partial a_j}\right|$ is large



Many resource sharing scenarios

Communication networks

- power control

$$u_i = \log_2 \left(1 + \frac{g_{ii} a_i}{\sum_{j \neq i} g_{ji} a_j + \sigma_i}\right)$$

- Medium Access Control (MAC)

$$u_i = a_i \cdot \prod_{j \neq i} (1 - a_j), \ \forall j : \ a_j \in [0, 1]$$

- flow control

$$u_i = a_i^{\beta_i} \cdot \left(\mu - \sum_{j \neq i} a_j\right)^+$$

Residential demand-side management, etc.

Engineering literature - I

Network Utility Maximization

(F. Kelly, M. Chiang, S. Low, etc.)

- No externality $U_i(a_i)$,
  or $U_i(\boldsymbol{a})$ jointly concave
- Short-term performance

Inefficient

- Myopic optimization (find the optimal action)

Our work

- Negative externality,
  not jointly concave in general
- Long-term performance
- Foresighted optimization (find the optimal policy)

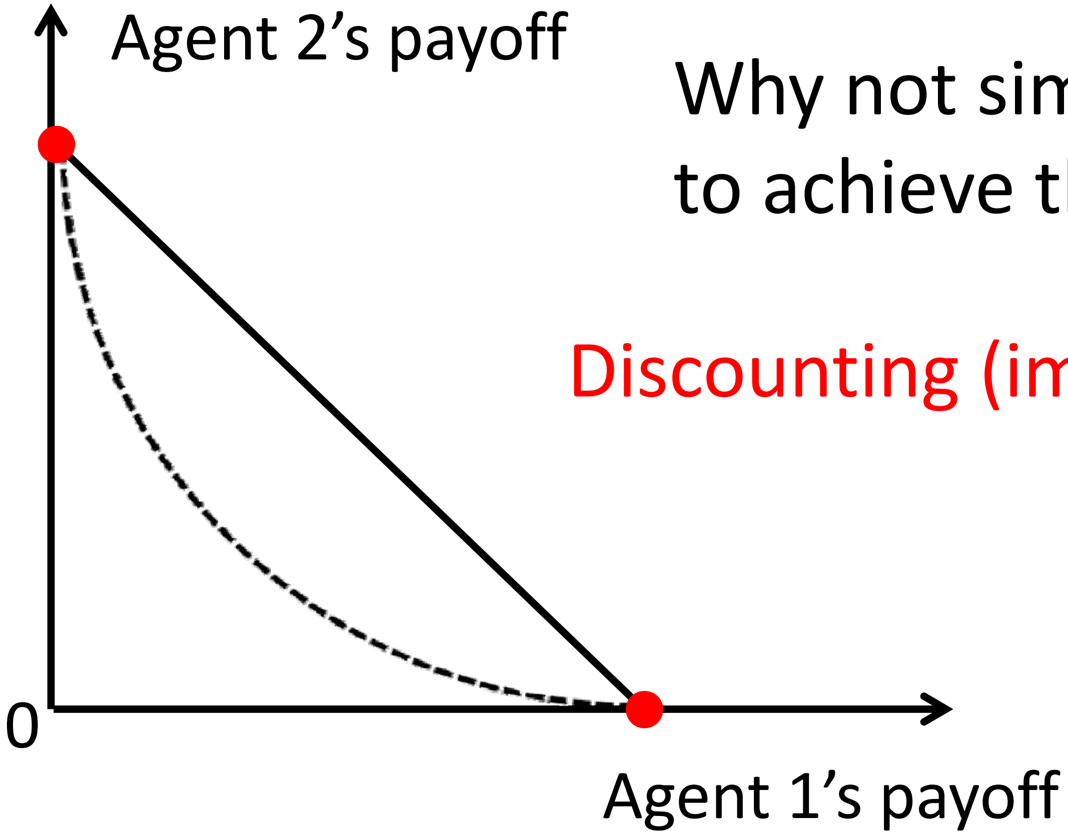
Engineering literature - II

Markov decision processes (D. Bertsekas, J. Tsitsiklis, E. Altman, etc.)	Our work
• Single agent	• Multiple agents
• Stationary policy is optimal	• Nonstationary policy

Economics literature

Existing theory (Fudenberg, Levine, Maskin 1994)	Our work
• Folk theorem-type results ***Not* constructive**	• Constructive
• Cardinality of feedback signals proportional to the cardinality of action sets ***High* overhead**	• Binary feedback regardless of the cardinality of action sets (exploit strong externality)
• Discount factor → 1	• Discount factor lower bounded
• Interior	• Pareto boundary

Challenge 1 – Why not round-robin TDMA?



Why not simply use round-robin TDMA to achieve the Pareto boundary?

Discounting (impatience, delay-sensitivity)

Challenge 1 – Illustrating Example

A simple example abstracted from wireless communication:

- 3 homogeneous agents, discount factor 0.7
- maximum payoff of each agent is 1
- max-min fairness: $\max \min_i U_i$ → optimal (1/3, 1/3, 1/3)

Round-robin TDMA policies (and variants):

- cycle length of 3: 123 123 123 → 0.18 (46% loss)
- cycle length of 4: 1233 1233 1233 → 0.26 (22% loss)
- cycle length of 8: 12332333 → 0.29 (13% loss)

Longer cycles to approach the optimal policy?

Computational Complexity

Longer cycles to approach the optimal nonstationary policy?

\# of non-trivial policies (each user has at least one slot) grows *exponentially* with # of users!
Lower bounded by N^{L-N} (N: # of users, L: cycle length)

In the 3-user example, to achieve within ~10% of optimal nonstationary policy, we need a cycle length 8 → 5796 policies

Under moderate number of users (N=10), for a good performance (L=20), more than 10^{10} (ten billion!) policies

Optimal nonstationary policy: complexity *linear* with # of users

Moral:

– Optimal policy is not cyclic

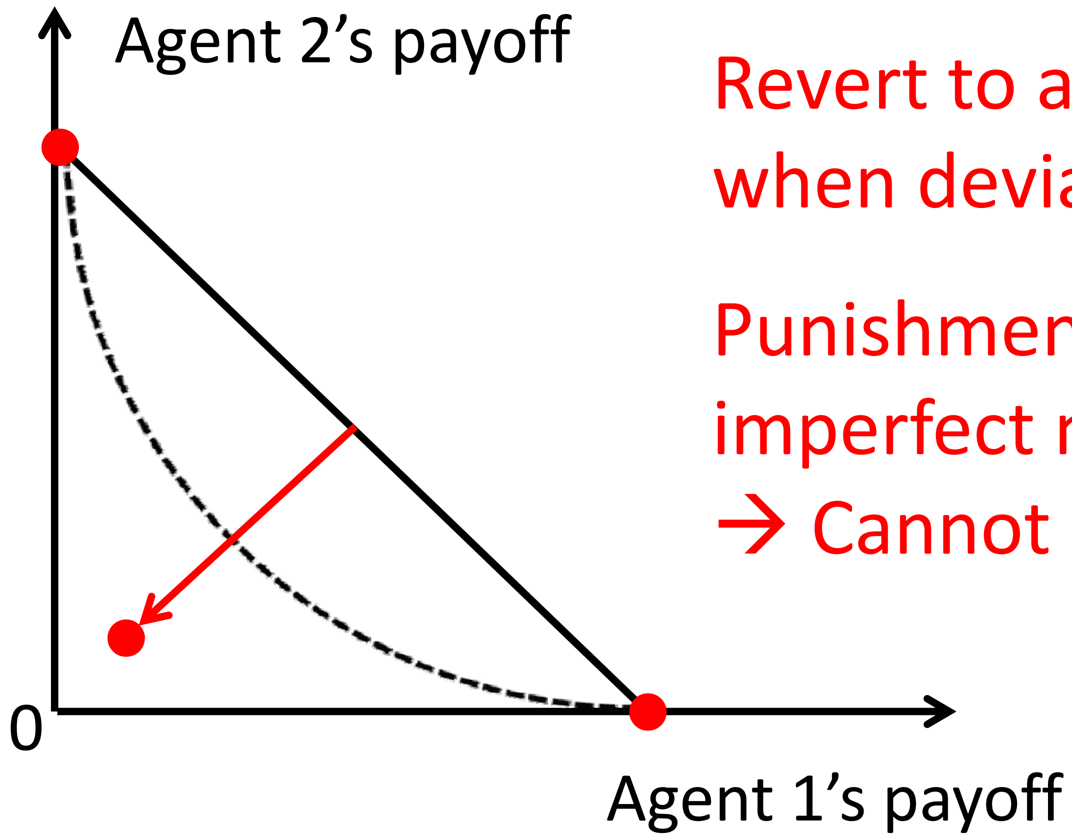
Good news:

- We construct a simple, intuitive and general algorithm to build such policies

- Complexity: *linear* vs *exponential* of round-robin

Challenge 2 – Imperfect monitoring

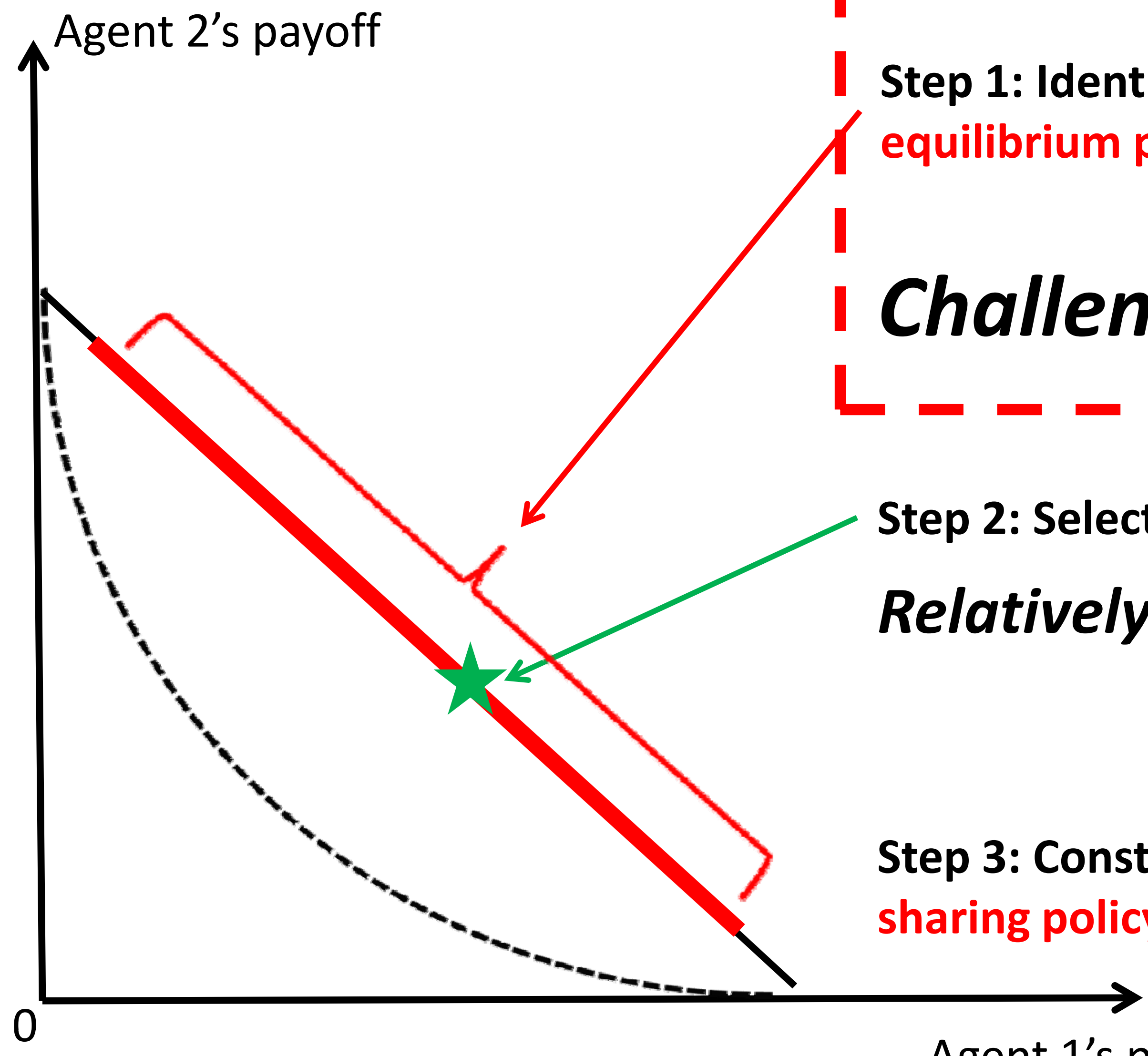
How to make the schedule deviation-proof?
(e.g. 122 122 122 may be,
but 1122222 1122222 may not)



Revert to an inefficient Nash equilibrium when deviation is detected?

Punishment will be triggered due to imperfect monitoring.
→ Cannot stay on Pareto boundary!

The design framework



Step 1: Identify the set of Pareto optimal equilibrium payoffs

Challenging!

Step 2: Select the optimal operating point

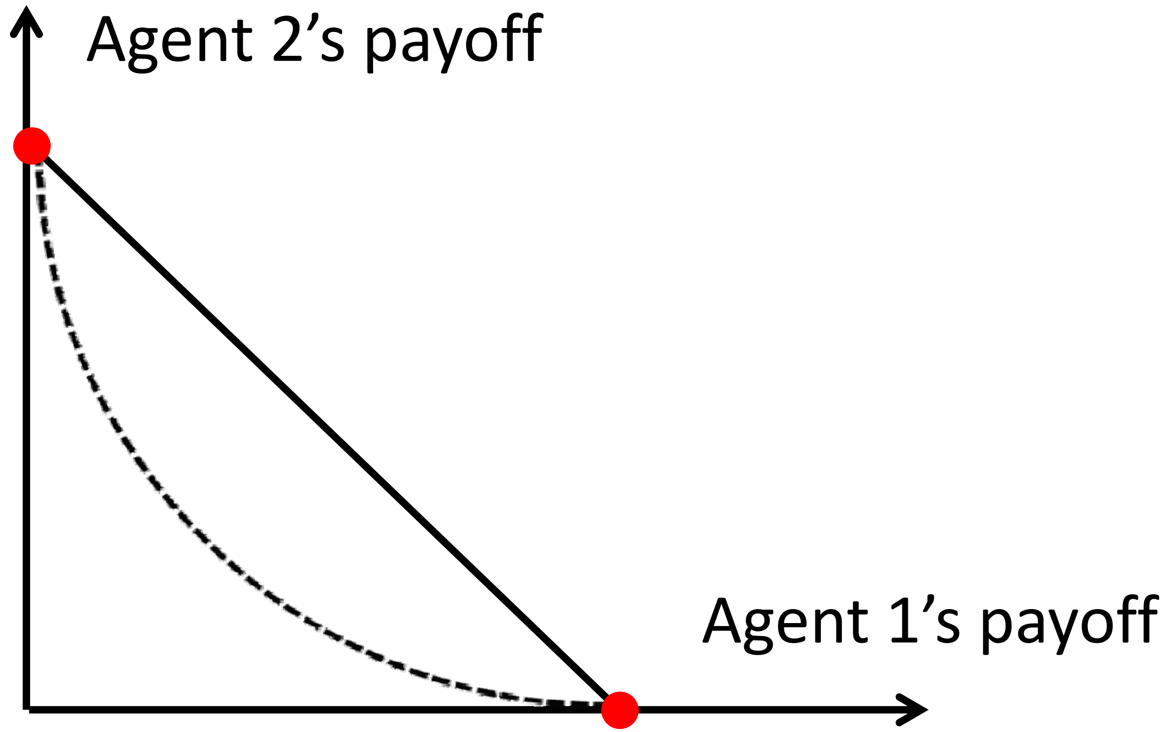
Relatively easy given step 1.

Step 3: Construct the optimal spectrum sharing policy

Challenging!

A typical scenario

- Action set: compact or finite
- Agent *i*'s preferred action profile: $\tilde{\boldsymbol{a}}^i = \arg\max_{\boldsymbol{a}} u_i(\boldsymbol{a})$
- $u_j(\tilde{\boldsymbol{a}}^i) = 0, \ \forall j \neq i$
- Strong negative externality: for any action profile $\boldsymbol{a} \neq \tilde{\boldsymbol{a}}^i, \ \forall i$, the payoff vector $\boldsymbol{u}(\boldsymbol{a})$ lies below the hyperplane determined by $\boldsymbol{u}(\tilde{\boldsymbol{a}}^i), \ \forall i$



A typical scenario

- Action set: compact or finite
- Agent *i*'s preferred action profile: $\tilde{\boldsymbol{a}}^i = \arg\max_{\boldsymbol{a}} u_i(\boldsymbol{a})$
- $u_j(\tilde{\boldsymbol{a}}^i) = 0, \ \forall j \neq i$
- Strong negative externality: for any action profile $\boldsymbol{a} \neq \tilde{\boldsymbol{a}}^i, \ \forall i$, the payoff vector $\boldsymbol{u}(\boldsymbol{a})$ lies below the hyperplane determined by $\boldsymbol{u}(\tilde{\boldsymbol{a}}^i), \ \forall i$
- $u_i(\boldsymbol{a})$ increasing in a_i and decreasing in a_j
- Binary noisy signal:

$$y = \begin{cases} 1, & f(\boldsymbol{a}) + \varepsilon > threshold \\ 0, & otherwise \end{cases}$$

$f(\boldsymbol{a})$: resource usage status, increasing in each a_i

ε : noise, infinite support

Step 1 – Identification

When agent i is active, agent j's *relative benefit from deviation*:

$$\alpha_{ij} \triangleq \sup_{a'_j \neq \tilde{a}^i_j} \frac{u_j(a'_j, \tilde{\boldsymbol{a}}^i_{-j}) - u_j(\tilde{\boldsymbol{a}}^i)}{\rho(y=1|a'_j, \tilde{\boldsymbol{a}}^i_{-j}) - \rho(y=1|\tilde{\boldsymbol{a}}^i)}$$

Payoff gain from deviation

Probability of detecting deviation

Theorem (Analytical identification of Pareto optimal equilibrium payoffs)

The set of Pareto optimal equilibrium payoffs is

$$\mathscr{P}_{\underline{\mu}} = \{\mathbf{v} : \sum_i v_i/\bar{v}_i = 1,\ v_j/\bar{v}_j \geq \mu_j\},\ \textit{where}\ \mu_j \triangleq \max_{i \neq j} \alpha_{ij} \cdot \left[1 - \rho(y = 1|\tilde{\mathbf{a}}^i)\right],$$

and any payoff in $\mathscr{P}_{\underline{\mu}}$ *can be achieved if the discount factor* δ *satisfies*

$$\delta \geq \underline{\delta}_{\mu} \triangleq \left(1 + \frac{1 - \sum_i \mu_i}{N - 1 + \sum_i \sum_{j \neq i} \alpha_{ij} \rho(y = 1|\tilde{\mathbf{a}}^i)}\right)^{-1}.$$

Step 1 – Identification

When agent i is active, agent j's *relative benefit from deviation*:

$$\alpha_{ij} \triangleq \sup_{a'_j \neq \tilde{a}^i_j} \frac{u_j(a'_j, \tilde{\mathbf{a}}^i_{-j}) - u_j(\tilde{\mathbf{a}}^i)}{\rho(y=1|a'_j, \tilde{\mathbf{a}}^i_{-j}) - \rho(y=1|\tilde{\mathbf{a}}^i)}$$

Payoff gain from deviation
Probability of detecting deviation

Theorem (Analytical identification of Pareto optimal equilibrium payoffs)

Hyperplane (strong externalities) + Constraints → Part of hyperplane (easily computed)

$$\mathscr{P}_{\underline{\mu}} = \{\mathbf{v} : \sum_i v_i/\bar{v}_i = 1, \; v_j/\bar{v}_j \geq \mu_j\}, \quad \textit{where } \mu_j \triangleq \max_{i \neq j} \alpha_{ij} \cdot \left[1 - \rho(y = 1|\tilde{\mathbf{a}}^i)\right],$$

and any payoff in $\mathscr{P}_{\underline{\mu}}$ *can be achieved if the discount factor* δ *satisfies*

Conditions on the discount factor (delay sensitivity):

$$\delta \geq \underline{\delta}_{\underline{\mu}} \triangleq \left(1 + \frac{1 - \sum_i \mu_i}{N - 1 + \sum_i \sum_{j \neq i} \alpha_{ij} \rho(y = 1|\tilde{\mathbf{a}}^i)}\right)^{-1}.$$

Step 1 - Key ideas

Decompose the target payoff profile $[v_1^\star, \ldots, v_N^\star]^T$ by $\boldsymbol{a}$

- decomposition: $\underbrace{v_i^\star}_{\text{Target payoff}} = (1-\delta) \cdot \underbrace{u_i(\boldsymbol{a})}_{\text{Instantaneous payoff}} + \delta \cdot \left[\sum_{y=0}^{1} \rho(y|\boldsymbol{a}) \underbrace{\gamma_i(y)}_{\text{Continuation payoff}}\right]$
- incentive constraints (IC): for all a_i' , we have

$$v_i^\star \geq (1-\delta) \cdot u_i(a_i', \boldsymbol{a}_{-i}) + \delta \cdot \left[\sum_{y=0}^{1} \rho(y|a_i', \boldsymbol{a}_{-i}) \gamma_i(y)\right]$$

Comparison with Bellman equations in MDPs

MDPs	Repeated Games
one agent → actions	multi-agent → action profiles
values	value profiles
value functions single-valued	value functions set-valued

Step 1 - APS

Consider a set $W \subset \mathbb{R}^n$ and a discount factor δ.

A pair $(\boldsymbol{v}, \boldsymbol{a})$ is admissible with respect to W and δ, if $\exists \boldsymbol{\gamma}(y) \in W$:

$$v_i = (1-\delta) \cdot u_i(\boldsymbol{a}) + \delta \cdot \left[\sum_{y=0}^{1} \rho(y|\boldsymbol{a})\gamma_i(y)\right]$$
$$\geq (1-\delta) \cdot u_i(a_i', \boldsymbol{a}_{-i}) + \delta \cdot \left[\sum_{y=0}^{1} \rho(y|a_i', \boldsymbol{a}_{-i})\gamma_i(y)\right]$$

$$B_\delta(W) = \{\boldsymbol{v} : \text{exists } \boldsymbol{a} \text{ such that } (\boldsymbol{v}, \boldsymbol{a}) \text{ admissible}\}$$

W self-generating: $W \subset B(W)$

All payoffs in the self-generating set are equilibrium payoffs!

By Abreu, Pearce, Stacchetti 1990 (APS)

Step 1 – APS is not constructive

APS proposed a *set-valued* value iteration to compute *W*:

Given a discount factor δ : ← Is it even feasible??

choose an initial $W_0 \subset \mathbb{R}^n$ all equilibrium payoffs $\subset B_\delta(W_0) \subset W_0$

How??

$W_1 = B_\delta(W_0)$ ← How??

$\vdots$

$W_\infty = B_\delta(W_\infty)$

Check whether $\boldsymbol{v} \in B_\delta(W_0)$: find $\boldsymbol{a}$ such that

$$\exists \gamma(y) \in W_0 \quad v_i = (1-\delta) \cdot u_i(\boldsymbol{a}) + \delta \cdot \left[\sum_{y=0}^{1} \rho(y|\boldsymbol{a})\gamma_i(y)\right]$$
$$\geq (1-\delta) \cdot u_i(a_i', \boldsymbol{a}_{-i}) + \delta \cdot \left[\sum_{y=0}^{1} \rho(y|a_i', \boldsymbol{a}_{-i})\gamma_i(y)\right]$$

A feasibility checking problem; May explore entire action space

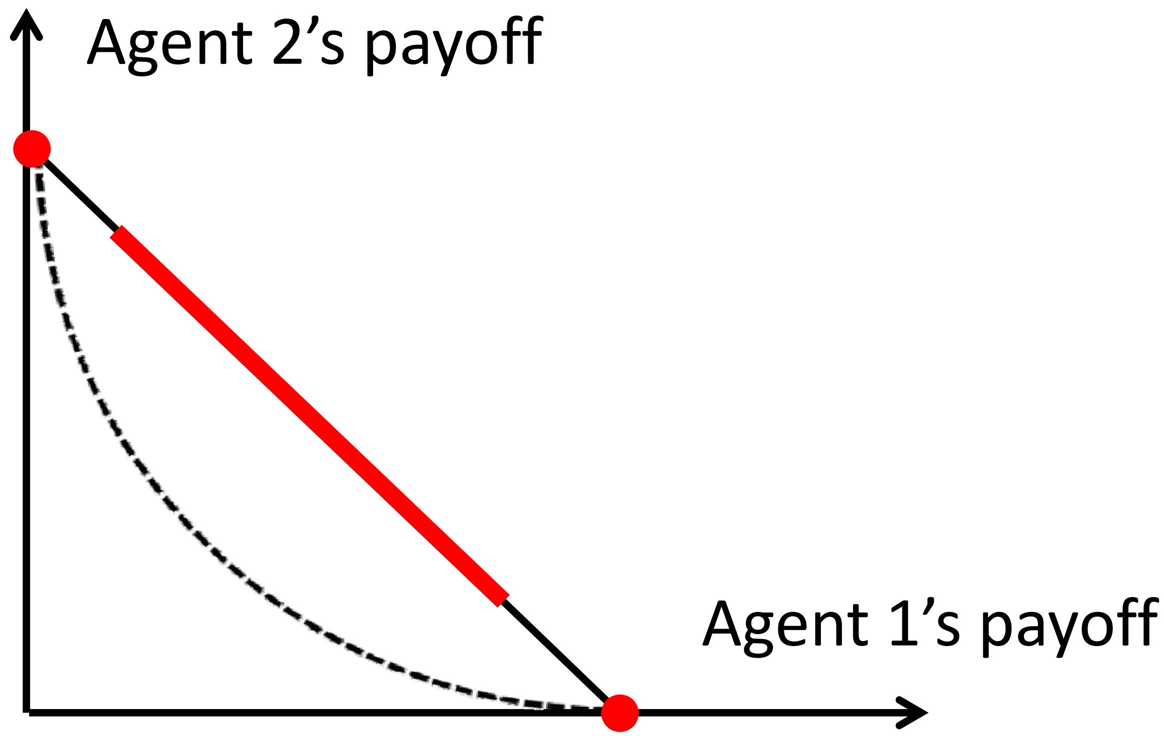
Even if we could compute *W*, how to construct the policy??

Step 1 – Our approach

We *analytically* determine *W*!

Consider *W* of the following form:

$$\mathscr{P}_{\boldsymbol{\mu}} = \{\boldsymbol{v} : \sum_i v_i/\bar{v}_i = 1, v_i/\bar{v}_i \geq \mu_i\}$$



Step 1 – Our approach

We *analytically* determine *W*!

Consider *W* of the following form:

$$\mathscr{P}_{\boldsymbol{\mu}} = \{\boldsymbol{v} : \sum_i v_i/\bar{v}_i = 1, v_i/\bar{v}_i \geq \mu_i\}$$

Check whether $\boldsymbol{v} \in B_\delta(\mathscr{P}_{\boldsymbol{\mu}})$: find ***a*** such that $\exists \gamma(y) \in \mathscr{P}_{\boldsymbol{\mu}}$

$$v_i = (1-\delta) \cdot u_i(\boldsymbol{a}) + \delta \cdot \left[\sum_{y=0}^{1} \rho(y|\boldsymbol{a})\gamma_i(y)\right]$$
$$\geq (1-\delta) \cdot u_i(a_i', \boldsymbol{a}_{-i}) + \delta \cdot \left[\sum_{y=0}^{1} \rho(y|a_i', \boldsymbol{a}_{-i})\gamma_i(y)\right]$$

linear constraints

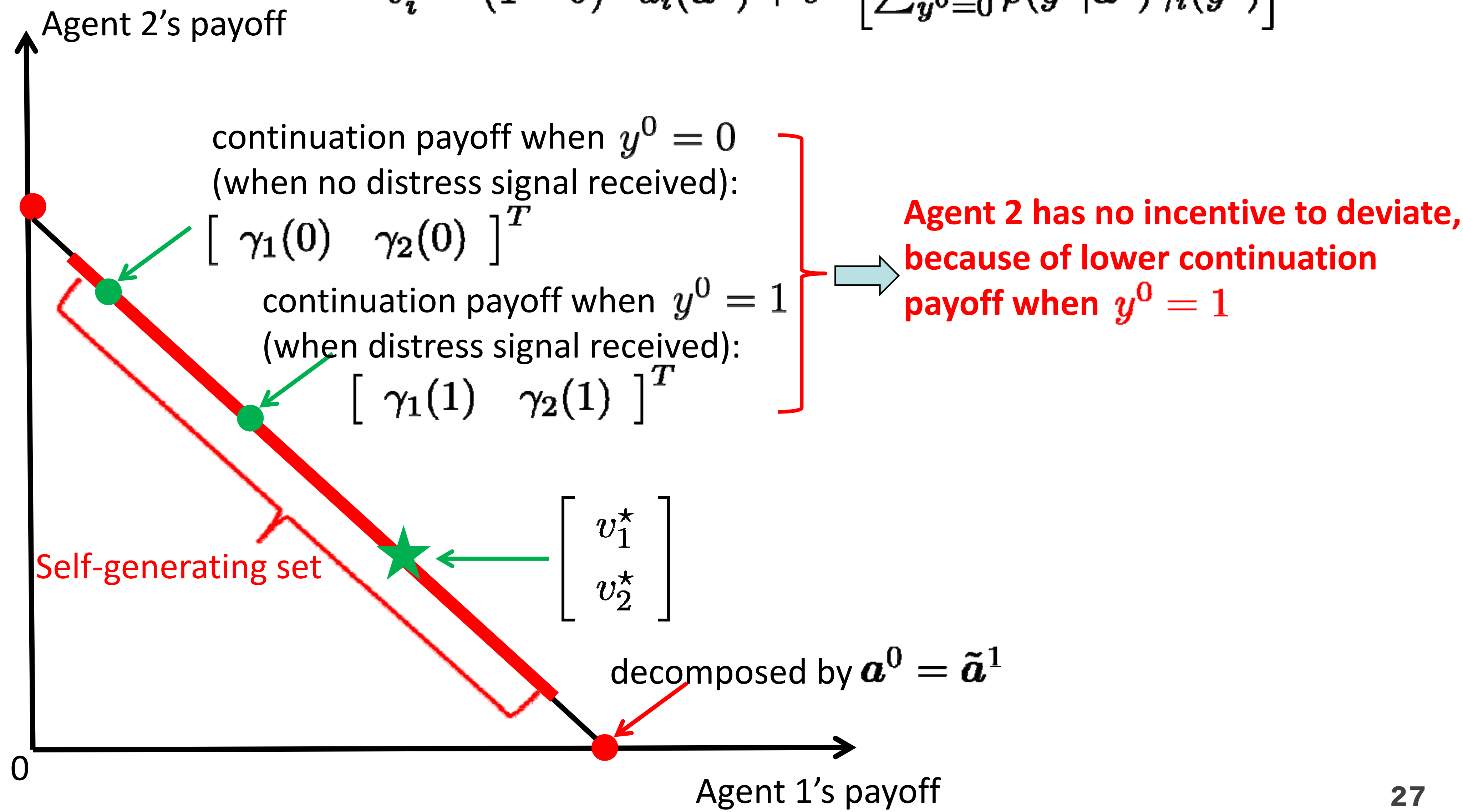
Find the lower bound on δ :

$$\underline{\delta}_{\boldsymbol{\mu}} = \max_{\boldsymbol{v} \in \mathscr{P}_{\boldsymbol{\mu}}} \min_{i \in \mathcal{N}} \max_{j \in \mathcal{N}} \beta_{ij}(\boldsymbol{v})$$

Step 1 – Illustrate self-generating sets

Decompose the target payoff profile $[v_1^\star, \ldots, v_N^\star]^T$:

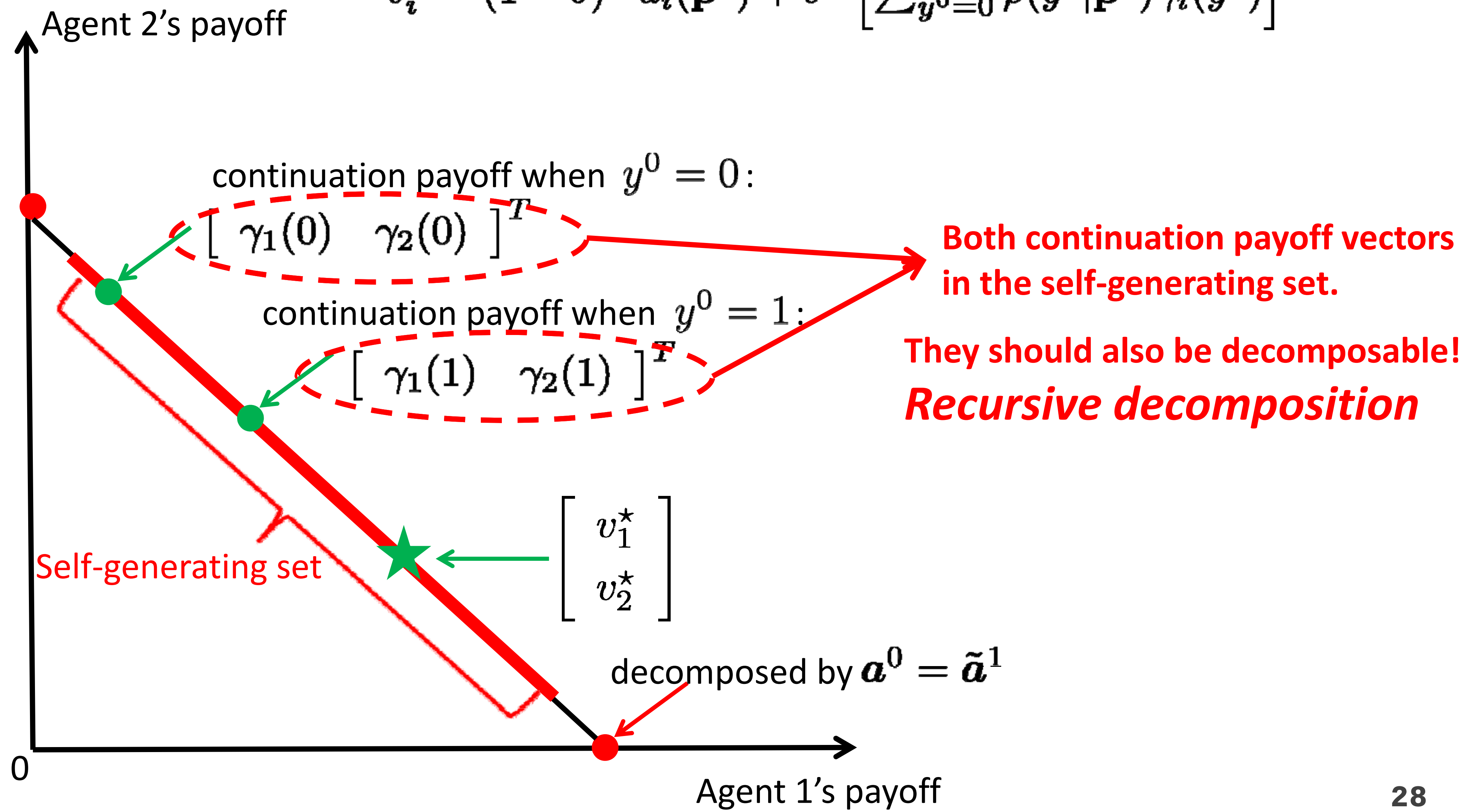
$$v_i^\star = (1-\delta)\cdot u_i(\boldsymbol{a}^0) + \delta \cdot \left[\sum_{y^0=0}^{1} \rho(y^0|\boldsymbol{a}^0)\gamma_i(y^0)\right]$$



Step 1 – Illustrate self-generating sets

Decompose the target payoff profile $[v_1^\star, \ldots, v_N^\star]^T$:

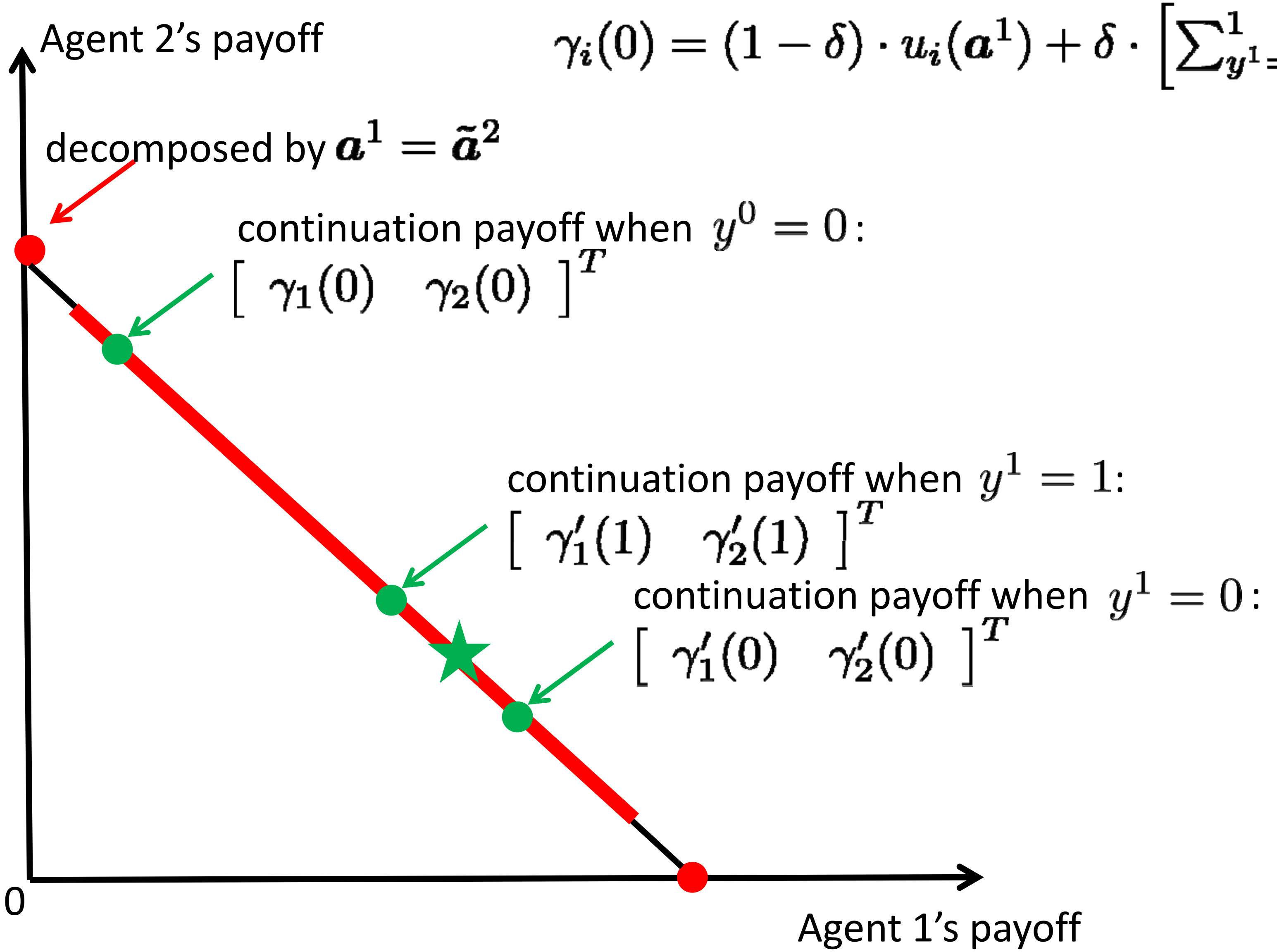
$$v_i^\star = (1-\delta) \cdot u_i(\mathbf{p}^0) + \delta \cdot \left[\sum_{y^0=0}^{1} \rho(y^0|\mathbf{p}^0)\gamma_i(y^0)\right]$$



Step 1 – Illustrate self-generating sets

For example, decompose $\begin{bmatrix} \gamma_1(0) & \gamma_2(0) \end{bmatrix}^T$:

$$\gamma_i(0) = (1-\delta)\cdot u_i(\boldsymbol{a}^1) + \delta \cdot \left[\sum_{y^1=0}^{1} \rho(y^1|\boldsymbol{a}^1)\gamma_i'(y^1)\right]$$



Step 2 – Select optimal operating point

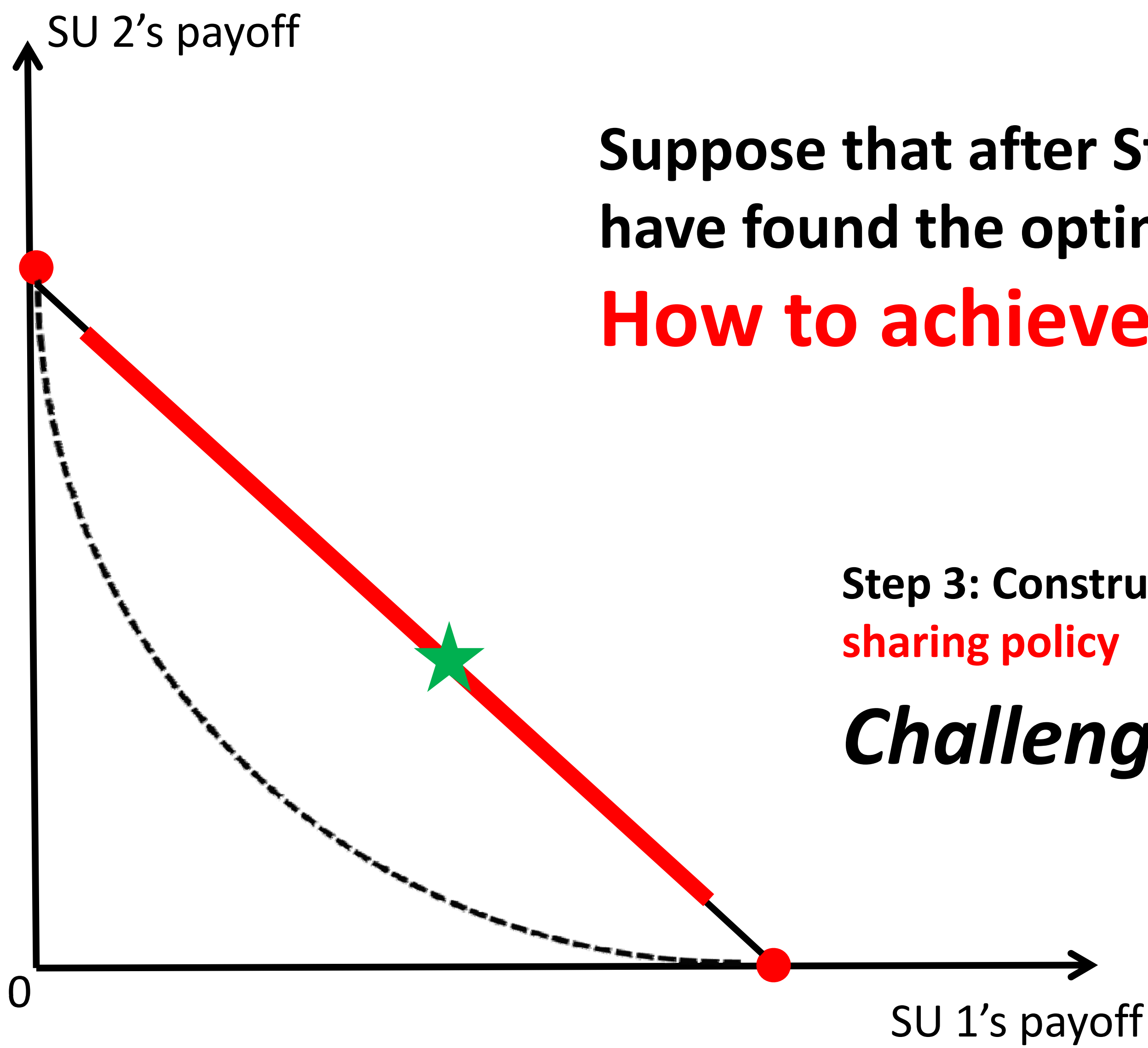
- Designer selects optimal operating point $\mathbf{v}^\star$:

$$\begin{aligned} \mathbf{v}^\star = \arg\max_{\mathbf{v}} \; & W(v_1, \ldots, v_N) \\ s.t. \; & \mathbf{v}^\star \in \mathscr{P}_{\underline{\boldsymbol{\mu}}} \\ & v_i \geq \underline{v}_i, \; \forall i \in \mathcal{N} \end{aligned}$$

Linear equalities and inequalities

- The above problem is easy to solve
  - $W(v_1, \ldots, v_N)$ usually jointly concave → convex optimization
  - Constraints are linear → dual decomposition, distributed algorithms

Step 3



Suppose that after Steps 1 and 2, we have found the optimal operating point.

How to achieve it?

Step 3: Construct the optimal spectrum sharing policy

Challenging!

Step 3 – Low-complexity online algorithm

The low-complexity online algorithm run by each user:

- A longest-"distance"-first (LDF) scheduling
- No message exchanges are needed at run-time

Input: the target payoff $\mathbf{v}^\star \in \mathscr{P}_{\underline{\mu}}$, the discount factor $\delta \geq \underline{\delta}$

Initialization: Set $t = 0$, $\mathbf{v}(0) = \mathbf{v}^\star$.
repeat
  "Distance from target" $d_i(t) = \frac{v_i(t)/\bar{v}_i - \underline{\mu}_i}{1 - v_i(t)/\bar{v}_i + \sum_{j \neq i} \alpha_{ji} \rho(y=1|\tilde{\mathbf{a}}^j)}, \forall i$
  $i^* \triangleq \arg\max_{j \in \mathcal{N}} d_j(t)$, $\mathbf{a}(t) = \tilde{\mathbf{a}}^{i^*}$
  if $y^t = 0$ (indicating no deviation) **then**
    $v_{i^*}(t+1) = v_{i^*}(t) - (\frac{1}{\delta} - 1) \cdot (1 - \frac{v_{i^*}(t)}{\bar{v}_{i^*}} + \sum_{j \neq i^*} \alpha_{i^* j} \rho(y = 1|\tilde{\mathbf{a}}^{i^*})) \cdot \bar{v}_{i^*}$
    $v_i(t+1) = \frac{1}{\delta} \cdot v_i(t) + (\frac{1}{\delta} - 1) \cdot \alpha_{i^* i} \rho(y = 1|\tilde{\mathbf{a}}^{i^*}) \cdot \bar{v}_i, \forall i \neq i^*$
  else
    $v_{i^*}(t+1) = v_{i^*}(t) - (\frac{1}{\delta} - 1) \cdot (1 - \frac{v_{i^*}(t)}{\bar{v}_{i^*}} - \sum_{j \neq i^*} \alpha_{i^* j} \rho(y = 0|\tilde{\mathbf{a}}^{i^*})) \cdot \bar{v}_{i^*}$
    $v_i(t+1) = \frac{1}{\delta} \cdot v_i(t) + (1 - \frac{1}{\delta}) \cdot \alpha_{i^* i} \rho(y = 0|\tilde{\mathbf{a}}^{i^*}) \cdot \bar{v}_i$
  end if
  $t \leftarrow t + 1$
until $\mathbf{v}(t) = \mathbf{v}^\star$

Define "distance from target"

User with the longest distance transmits

Distances updated analytically

Theorem: this algorithm achieves the desired Pareto optimal point $\mathbf{v}^\star$

Convergence

Theorem: The algorithm converges to the desired Pareto optimal point in logarithmic time.

Details: $\left|(1-\delta)\sum_{\tau=0}^{t}\delta^{\tau}\cdot u_i^{\tau}-v_i^{\star}\right|\le v_i^{\star}\cdot\delta^{t+1}$

Distance decreases exponentially
→ Convergence in log. time

Throughput achieved at time *t*

Target operating point

Theorem: Dynamic entry and exit of agents does not affect the convergence rate of existing agents!

Implementation

Message exchange *before* run-time

- Each user i needs to know:
  - maximum payoffs of all the users: $\{\bar{v}_j\}_{j \in \mathcal{N}}$
  - boundary of $\mathcal{P}_{\underline{\mu}}$: $\{\underline{\mu}_j\}_{j \in \mathcal{N}}$
  - relative benefits from deviation: $\{b_{ij}\}_{j \in \mathcal{N}}$
  - probability of distress signal: $\{\rho(y = 1 | \tilde{\mathbf{p}}^j)\}_{j \in \mathcal{N}}$
- Total amount: $N^2 + 2N$

Message exchange *at* run-time

- ***None!***

Total amount of message exchange bounded, does not increase with time!

Other algorithms (e.g. NUM): $O(N) \times \#$ of steps to converge

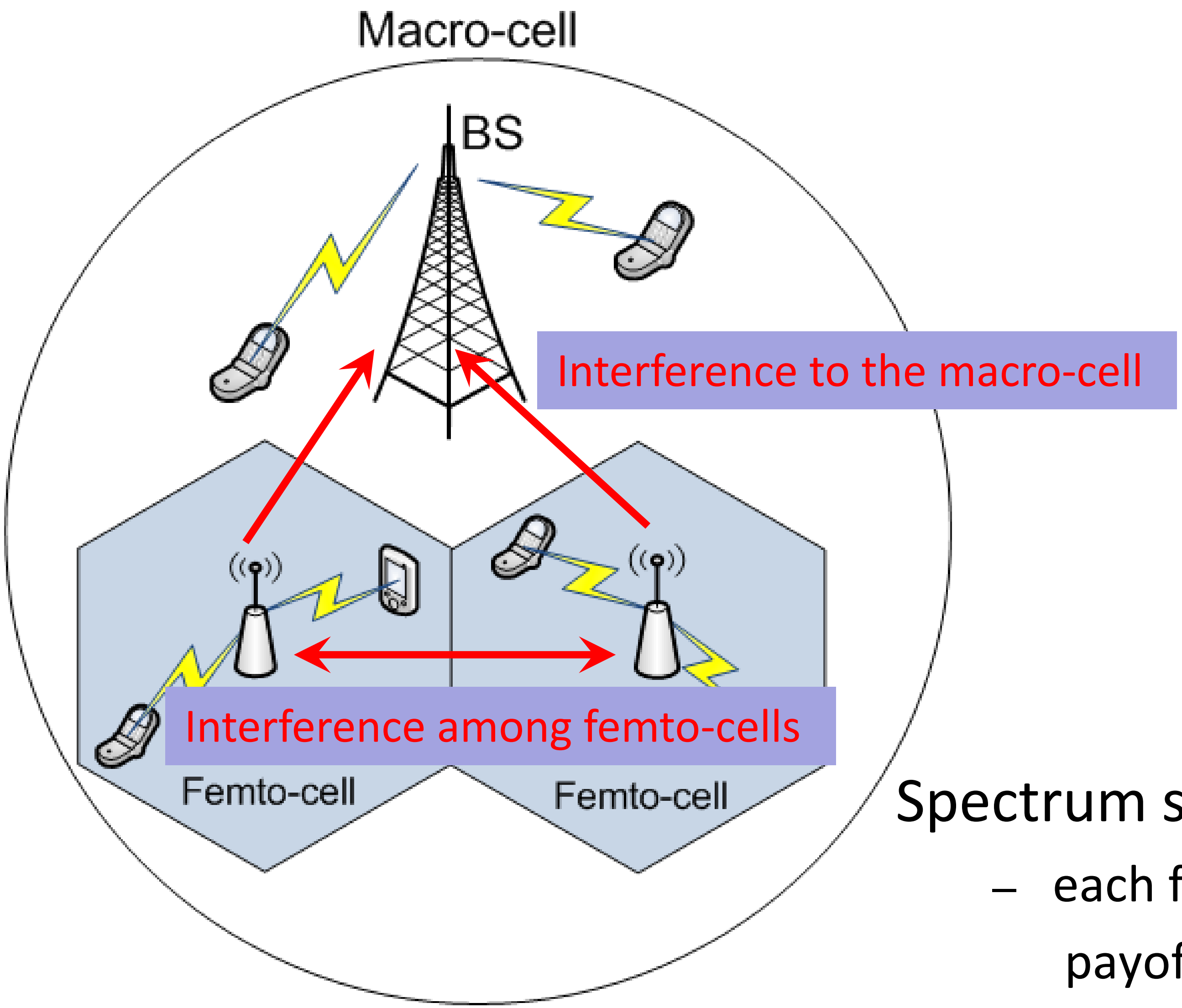
unbounded

Applications

Nonstationary spectrum sharing - Utility Maximization

Y. Xiao and M. van der Schaar, "Dynamic Spectrum Sharing Among Repeatedly Interacting Selfish Users With Imperfect Monitoring," *JSAC special issue on Cognitive Radio Systems*, vol. 30, no. 10, pp. 1890-1899, Nov. 2012.

System Model - Illustration



Spectrum sharing among femto-cells

- each femto-cell maximizes its own payoff (e.g. throughput)
- subject to interference temperature constraints imposed by the macro-cell

Simulation results - benchmarks

Constant policies: transmit at fixed power levels simultaneously

Jianwei Huang, Randall Berry, and Michael Honig, "Distributed interference compensation for wireless networks," *IEEE JSAC*, 2006.

C. W. Tan and Steven Low, "Spectrum management in multiuser cognitive wireless networks: Optimality and algorithm," *IEEE JSAC*, 2011.

Punish-forgive (PF) policies:

- deviation-proof
- same as constant policies when no distress signal
- transmit at maximum power levels forever once distress signal received

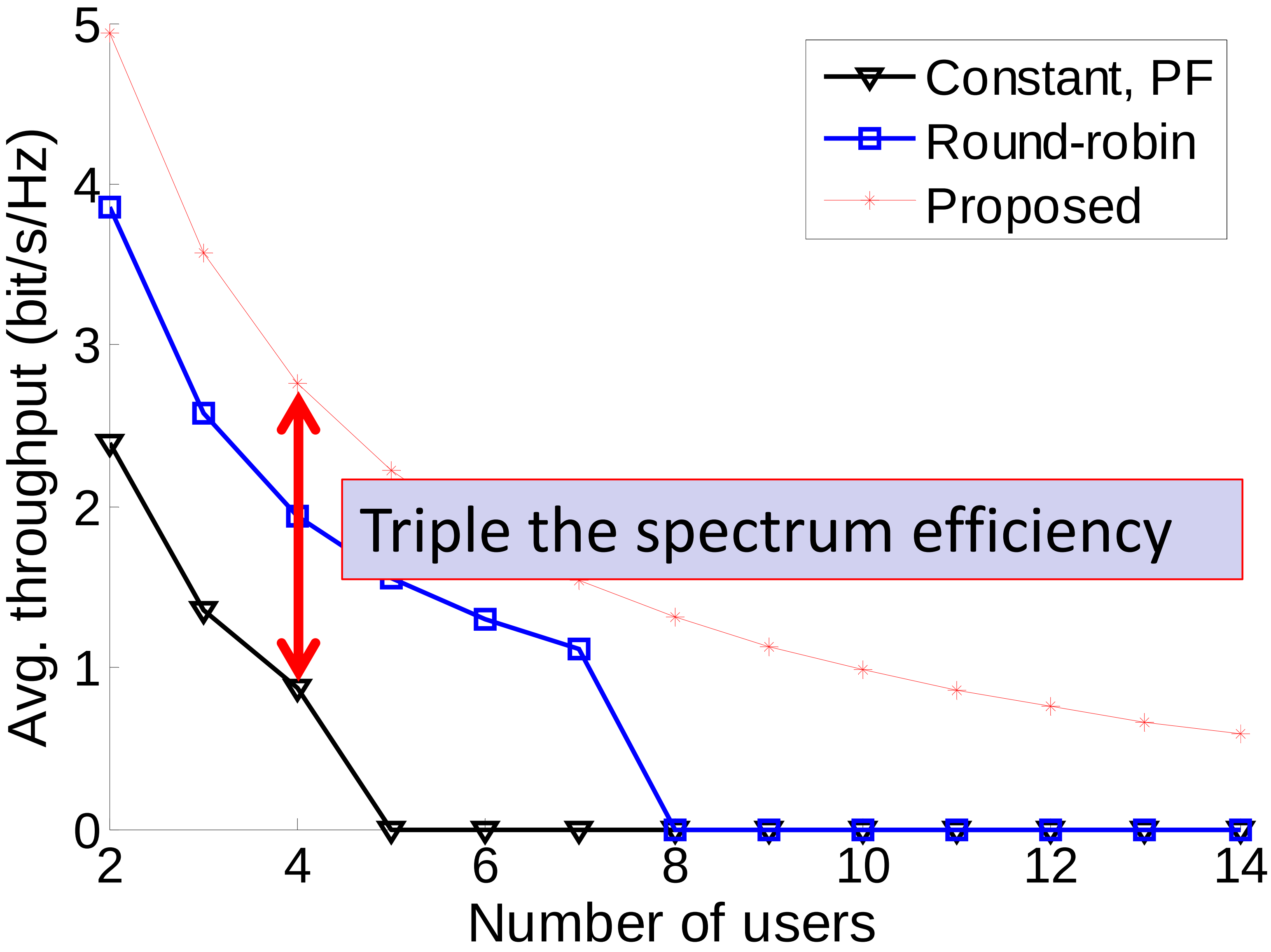
R. Etkin, A. Parekh, and David Tse, "Spectrum sharing for unlicensed bands," *JSAC*, 2007.

Y. Wu, B. Wang, Ray Liu, and T. C. Clancy, "Repeated open spectrum sharing game with cheat-proof strategies," *IEEE Trans. Wireless Commun.*, 2009.

Round-robin TDMA policies

Simulation results

Fixed minimum throughput guarantees: 0.5 bits/s/Hz



Extensions

A framework of cost minimization:

- Each agent i incurs a cost $c_i(a_i)$
- Each agent minimizes its cost subject to minimum payoff requirement

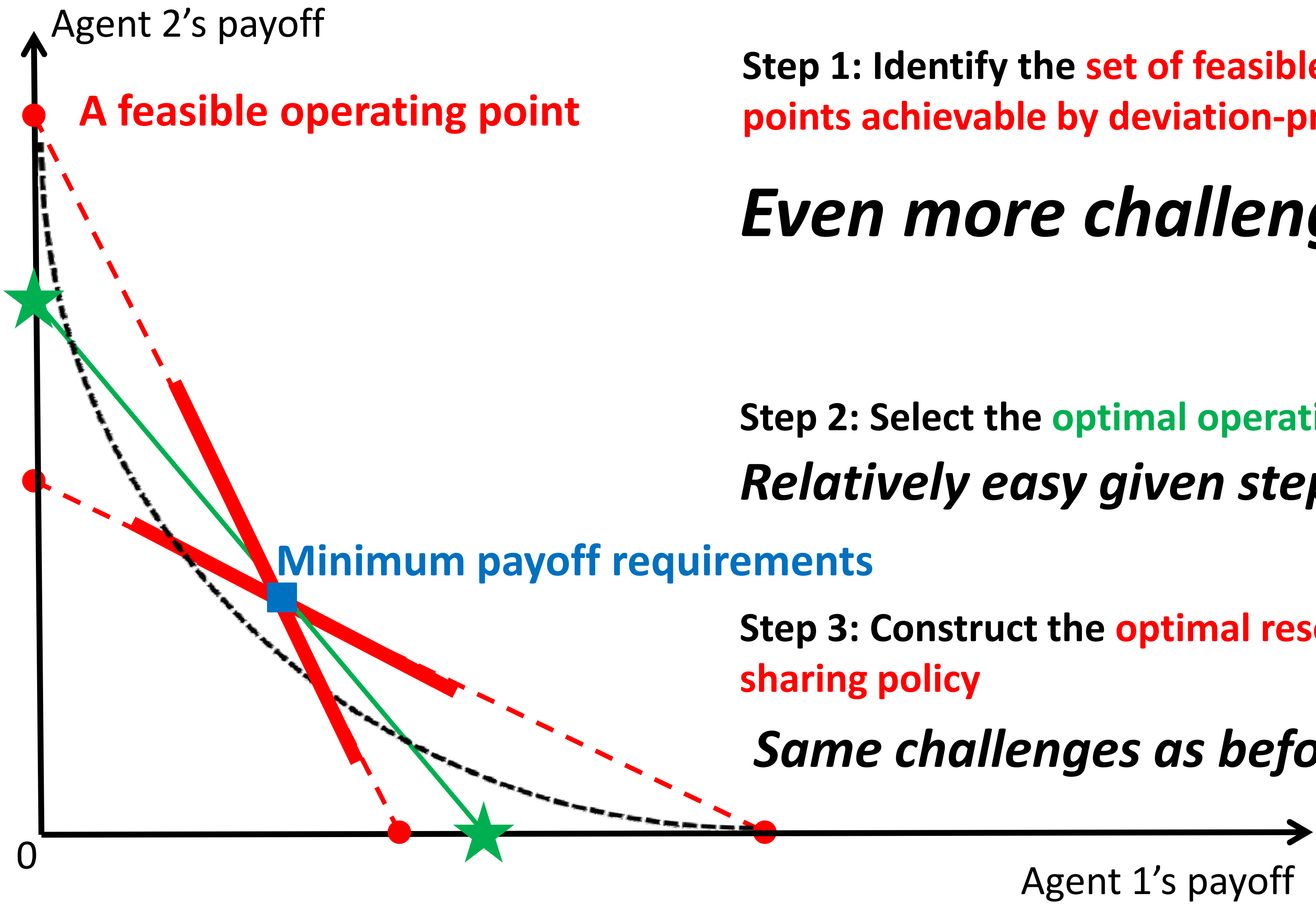
Design problem:

$$\min_{\boldsymbol{\pi}} \; W(C_1(\boldsymbol{\pi}), \ldots, C_N(\boldsymbol{\pi})) \quad \leftarrow \text{Social welfare function}$$

$$s.t. \quad U_i(\boldsymbol{\pi}) \geq \underline{v}_i, \; \forall i \in \mathcal{N} \quad \leftarrow \text{Minimum payoff guarantees}$$

$$\boldsymbol{\pi} \text{ is deviation} - \text{proof}$$

NOT a trivial extension



Step 1: Identify the **set of feasible operating points achievable by deviation-proof policies**

Even more challenging!

Step 2: Select the **optimal operating point**

Relatively easy given step 1.

Step 3: Construct the **optimal resource sharing policy**

Same challenges as before.

Nonstationary spectrum sharing – Energy consumption minimization

Y. Xiao and M. van der Schaar, “Energy-efficient nonstationary spectrum sharing,” Accepted by *IEEE Transactions on Communications.* Available at: http://arxiv.org/abs/1211.4174

Energy efficiency

Benchmarks:

1. Stationary policies: transmit at fixed power levels simultaneously

Jianwei Huang, Randall Berry, and Michael Honig, "Distributed interference compensation for wireless networks," *IEEE JSAC*, 2006.

R. Etkin, A. Parekh, and David Tse, "Spectrum sharing for unlicensed bands," *JSAC*, 2007.

Y. Wu, B. Wang, Ray Liu, and T. C. Clancy, "Repeated open spectrum sharing game with cheat-proof strategies," *IEEE Trans. Wireless Commun.*, 2009.

C. W. Tan and Steven Low, "Spectrum management in multiuser cognitive wireless networks: Optimality and algorithm," *IEEE JSAC*, 2011.

S. Sorooshyari, C. W. Tan, and Mung Chiang, "Power control for cognitive radio networks: Axioms, algorithms, and analysis", *ACM/IEEE Trans. Netw.*, 2012.

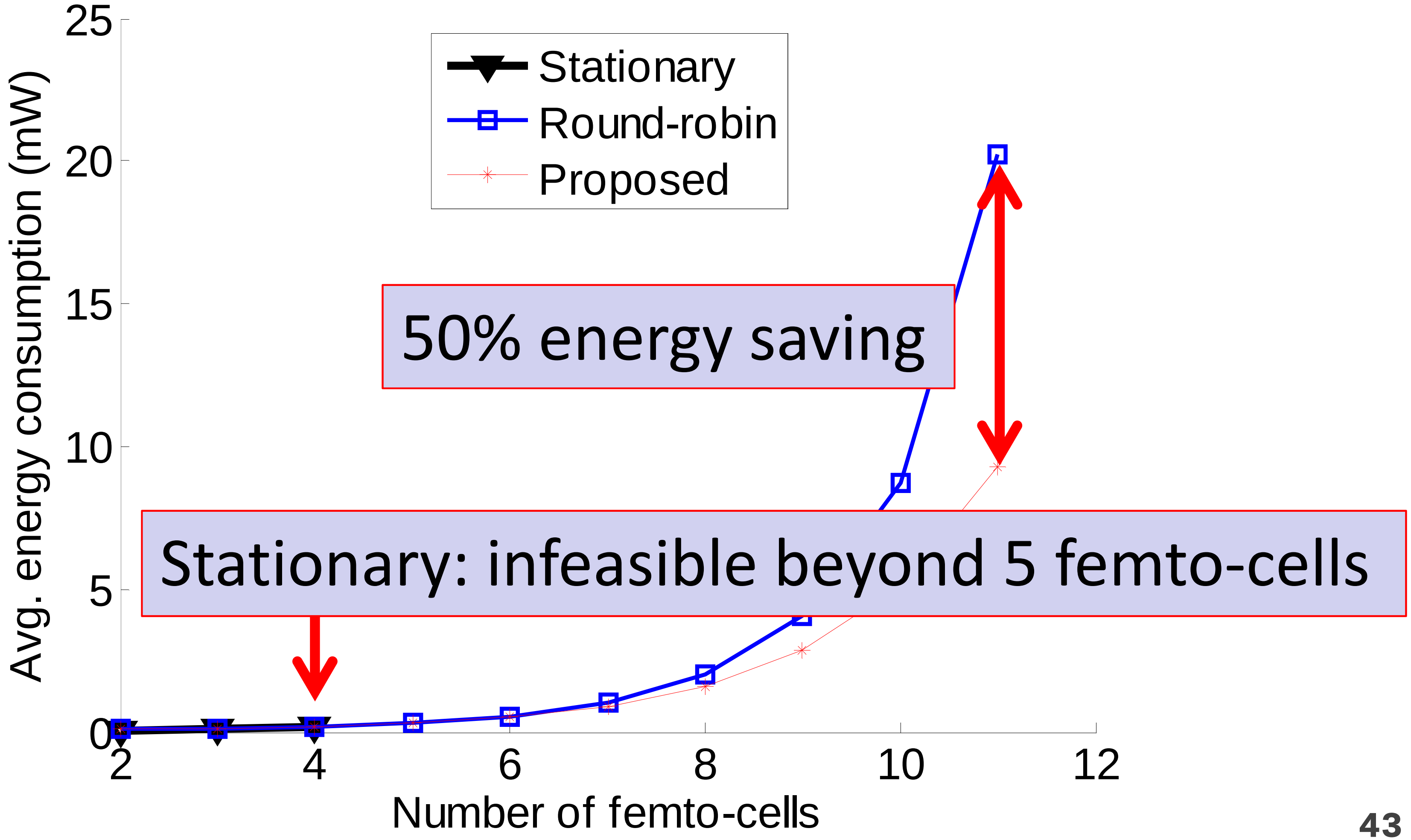
2. Round-robin TDMA policies

Energy efficiency

1 BS with minimum throughput requirement of 1 bit/s/Hz

2-15 femto-cells with minimum throughput requirement of 0.5 bit/s/Hz

Small number of femto-cells:

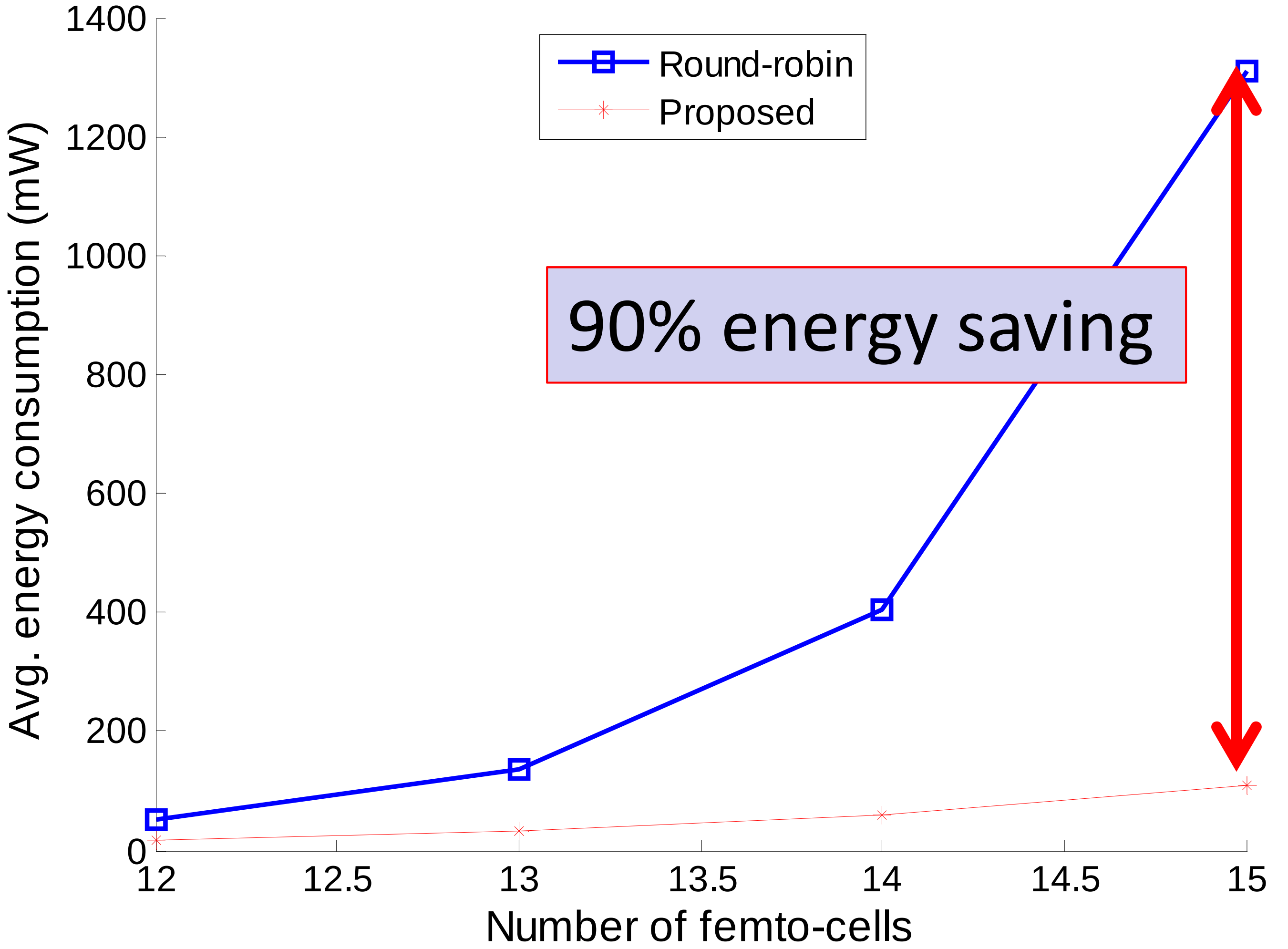


Energy efficiency

1 BS with minimum throughput requirement of 1 bit/s/Hz

2-15 femto-cells with minimum throughput requirement of 0.5 bit/s/Hz

Large number of femto-cells:



The general scenario

- Action set: compact or finite
- Agent *i*'s preferred action profile: $\tilde{\boldsymbol{a}}^i = \arg\max_{\boldsymbol{a}} u_i(\boldsymbol{a})$
- $u_j(\tilde{\boldsymbol{a}}^i) = 0, \ \forall j \neq i$ **Not necessary**
- Strong negative externality: for any action profile $\boldsymbol{a} \neq \tilde{\boldsymbol{a}}^i, \ \forall i$, the payoff vector $\boldsymbol{u}(\boldsymbol{a})$ lies below the hyperplane determined by $\boldsymbol{u}(\tilde{\boldsymbol{a}}^i), \ \forall i$
- $u_i(\boldsymbol{a})$ increasing in a_i and decreasing in a_j **Not necessary**
- Binary noisy signal:

$$y = \begin{cases} 1, & f(\boldsymbol{a}) + \varepsilon > threshold \\ 0, & otherwise \end{cases}$$

More general, still binary

$f(\boldsymbol{a})$: resource usage status, increasing in each a_i

ε : noise

Conclusions so far

Proposed

- Optimal ***nonstationary*** resource sharing policies
- Efficiency is achieved ***even under binary* feedback with errors**

Huge performance gain in spectrum sharing

- 3x spectrum efficiency
- 90% energy saving

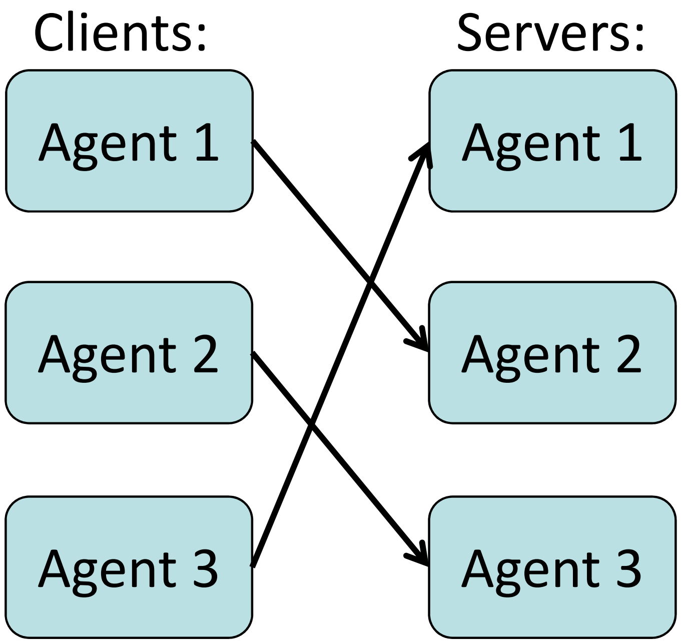
Solutions applicable to many engineering systems

- Decentralized users sharing a common resource
- Imperfect knowledge about the resource usage status

Resource exchange with imperfect monitoring

- Interaction
  - everybody interacts with everybody
  - **agents interact in pairs**
- Externality
  - one's action affects the others' payoffs directly and negatively
  - **one's action affects the others' payoffs directly and positively**
  - one's action does not affect the others' payoffs, but is coupled with the others' actions through constraints
- Monitoring
  - perfect / **imperfect**
- State
  - none (the system stays the same) / **public** / private
- Deviation-proof
  - no / **yes**
-

A resource exchange problem



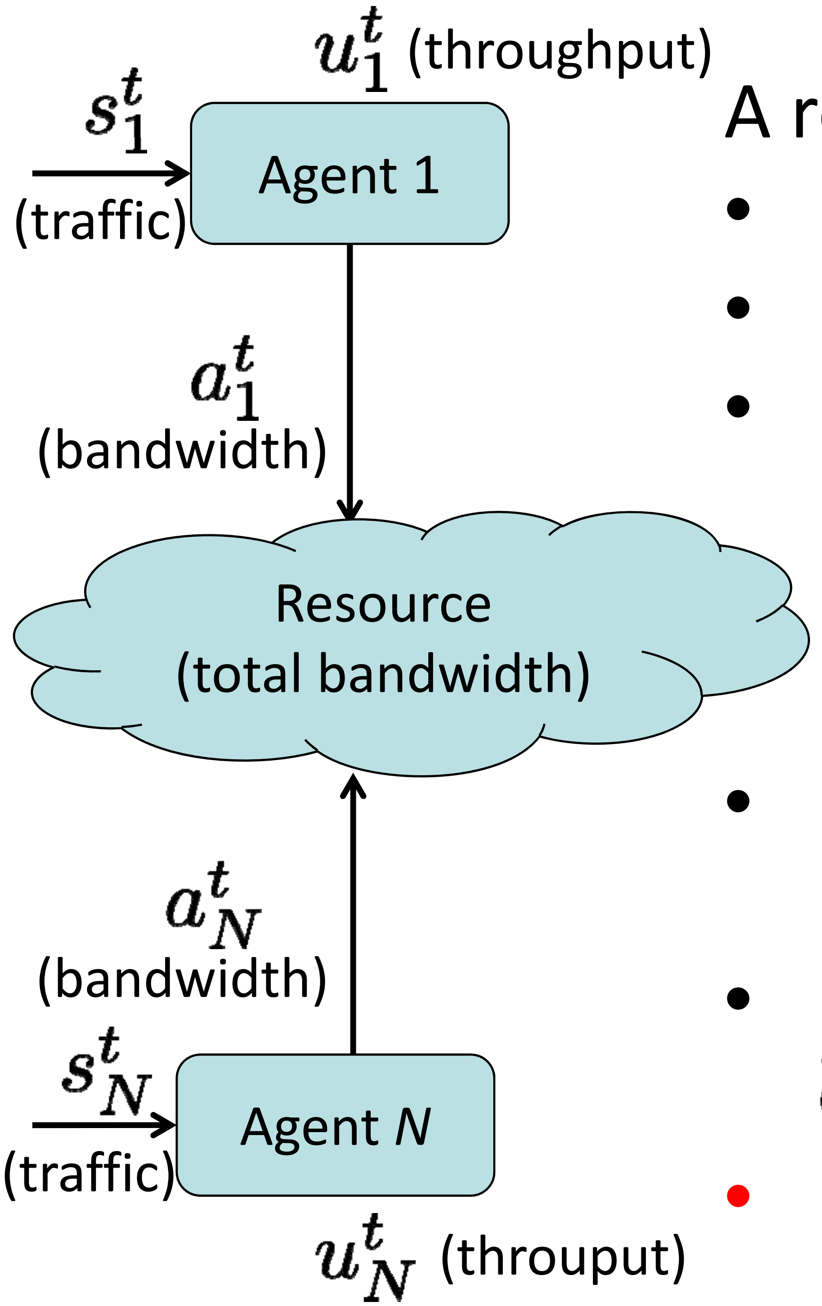
A resource exchange scenario:

- Anonymous agents 1, …, *N*
- Time is slotted *t* = 0, 1, 2, …
- At each time slot *t*:
  1. Random matching into pairs
  2. Server chooses “serve” or “not”
  3. Client monitors with errors
- Anonymity, random matching → rating mechanisms
- Propose the first rating mechanism that achieves social optimum under monitoring errors
- Nonstationary

Resource sharing with dynamic private states

- Interaction
  - **everybody interacts with everybody**
  - agents interact in pairs
- Externality
  - one's action affects the others' payoffs directly and negatively
  - one's action affects the others' payoffs directly and positively
  - **one's action does not affect the others' payoffs, but is couple d with the others' actions through constraints**
- Monitoring
  - perfect / **imperfect**
- State
  - none (the system stays the same) / public / **private**
- Deviation-proof
  - **no** / yes

A resource sharing problem



A resource sharing scenario:

- A resource shared by agents 1, …, *N*
- Time is slotted *t* = 0, 1, 2, …
- At each time slot *t*:
  1. Agent *i* observes state s_i^t
  2. Agent *i* chooses action a_i^t
  3. Receives payoff $u_i^t = u_i(s_i^t, a_i^t)$
- Strategy:

$$\pi_i : s_i^t \mapsto a_i^t$$

- Long-term payoff:

$$U_i(\pi_i, \boldsymbol{\pi}_{-i}) = \mathbb{E}\{(1-\delta)\sum_{t=0}^{\infty} \delta^t u_i^t\}$$

- Optimal Multi-user MDP

Final conclusions

- Three classes of resource sharing/exchange problems
- Optimal policies are often nonstationary → new tools
- Future works
  - Different interactions
  - Network topologies
  - Different state transition dynamics
  - Learning
  - Many other dimensions

Thank you!

Backup Slides

Engineering literature - II

Distributed optimization/consensus
(A. Ozdaglar, A. Nedich, etc.)

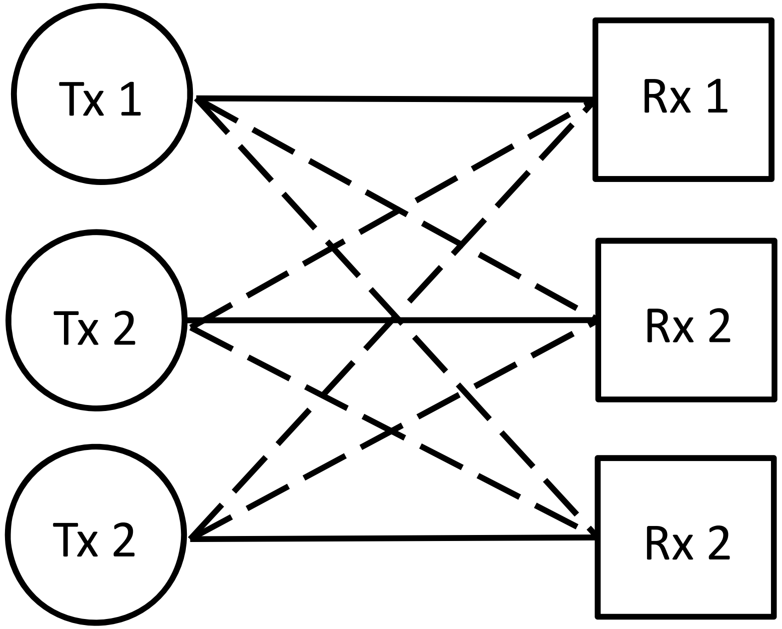
- Jointly concave payoff
  (not suitable for resource sharing)
- Myopic optimization (find the optimal action)

Our work

- Not jointly concave in general
- Foresighted optimization (find the optimal policy)

Illustration – Stationary policies

A simple network with three homogeneous users:



Direct channel gains: 1
Cross channel gains: 0.25

Noise power at both users' receivers: 5 mW

Both users discount throughput and energy consumption by $\delta = 0.6$.

Min. average throughput requirement: 1.5 bits/s/Hz

Channel gains are fixed. No PU.

State: channel conditions (fixed), PU activity (always idle)

Action: transmit power levels

Stationary policy : Both users transmit at fixed power levels simultaneously

Instantaneous power levels: (186, 186, 186) mW
Average energy consumption: (186, 186, 186) mW

Illustration – Simple nonstationary policies

A simple nonstationary policy: round-robin TDMA (cycle = 3)

Transmit schedule: 123 123 123 …

(Actions are time dependent)

Instantaneous power levels: (33, 144, 1432) mW

Power levels increase with the delay (the position in the cycle)

Average energy consumption: (17, 44, 263) mW

User 1: $33 \cdot \frac{1}{1+\delta+\delta^2} = 17$

User 2: $144 \cdot \frac{\delta}{1+\delta+\delta^2} = 44$

User 3: $1432 \cdot \frac{\delta^2}{1+\delta+\delta^2} = 263$

Better….

Illustration – Simple nonstationary policies

Performance improvement by increasing the cycle length

Round-robin (cycle = 4):

Optimal transmit schedule: 1233 1233 1233…

Instantaneous power levels: (43, 212, 249) mW

Power levels increase with the delay (the position in the cycle)

But the difference between user 2 and user 3 is small (user 3 has two slots)

Average energy consumption: (20, 58, 66) mW

User 1: $43 \cdot \frac{1}{1+\delta+\delta^2+\delta^3} = 20$

User 2: $212 \cdot \frac{\delta}{1+\delta+\delta^2+\delta^3} = 58$

User 3: $249 \cdot \frac{\delta^2+\delta^3}{1+\delta+\delta^2+\delta^3} = 66$

Illustration – Optimal nonstationary policies

The optimal policy is NOT cyclic

Transmit schedule: 123323213231…

Instantaneous power levels: (108, 108, 108) mW

Performance gains (total average energy consumption reduction):

80% compared to stationary policy;

67% compared to round-robin TDMA of cycle 3;

25% compared to round-robin TDMA of cycle 4.

Longer cycles to approach the optimal nonstationary policy?

Step 1 – Recursive decomposition

- *Recursive* decomposition:
  - continuation payoffs $[\gamma_1(y), \ldots, \gamma_N(y)]^T$ can be decomposed,
  - $\gamma_i(y) = (1-\delta) \cdot u_i(\boldsymbol{a}) + \delta \cdot \left[\sum_{y'=0}^{1} \rho(y'|\boldsymbol{a})\gamma_i'(y')\right] \ \forall\ y = 0, 1$
    $\gamma_i(y) \geq (1-\delta) \cdot u_i(a_i', \boldsymbol{a}_{-i}) + \delta \cdot \left[\sum_{y'=0}^{1} \rho(y'|a_i', \boldsymbol{a}_{-i})\gamma_i'(y')\right], \ \forall a_i'$

Different continuation payoff function γ_i'
-> different decomposition
-> ***Nonstationary*** policy!

Self-generating set: a set of payoff vectors in which every payoff vector can be decomposed by an action profile, ***and the continuation payoff vector lies in the set***

⇒ **All payoffs in the self-generating set are equilibrium payoffs!**

Publications

5 journal papers accepted as the first author

- Y. Xiao and M. van der Schaar, "Optimal foresighted multi-user wireless video," Accepted subject to minor revision by *JSTSP, special issue on Visual Signal Processing for Wireless Networks.*
- Y. Xiao and M. van der Schaar, "Energy-efficient nonstationary spectrum sharing," Accepted by *IEEE Trans. Commun..* Available at arXiv.
- Y. Xiao and M. van der Schaar, "Dynamic Spectrum Sharing Among Repeatedly Interacting Selfish Users With Imperfect Monitoring," *JSAC special issue on Cognitive Radio Systems*, Nov. 2012.
- Y. Xiao, J. Park, and M. van der Schaar, "Repeated Games With Intervention: Theory and Applications in Communications," *IEEE Trans. Commun.*, Oct. 2012.
- Y. Xiao, J. Park, and M. van der Schaar, "Intervention in Power Control Games with Selsh Users," *IEEE JSTSP, Special issue on Game Theory In Signal Processing*, Apr. 2012.

Publications

3 journal papers submitted as the first author

- Y. Xiao and M. van der Schaar, “Foresighted Demand Side Management," Submitted. Available at: http://arxiv.org/abs/1401.2185
- Y. Xiao and M. van der Schaar, “Socially-Optimal Design of Service Exchange Platforms with Imperfect Monitoring,” Submitted. Available at: http://arxiv.org/abs/1310.2323
- Y. Xiao, W. Zame, and M. van der Schaar, “Technology Choices and Pricing Policies in Public and Private Wireless Networks,” Submitted. Available at: http://arxiv.org/abs/1011.3580

Publications

Other journal papers as the 2nd or 3rd author

- M. Alizadeh, Y. Xiao, A. Scaglione, and M. van der Schaar, “Dynamic Incentive Design for Participation in Direct Load Scheduling Programs,” Submitted. Available at: http://arxiv.org/abs/1310.0402
- L. Song, Y. Xiao, and M. van der Schaar, “A Repeated Game Framework

  For Demand Side Management in Smart Grids,” Submitted. Available: http://arxiv.org/abs/1311.1887
- M. van der Schaar, Y. Xiao, and W. Zame, “Designing Ecient Resource Sharing For Impatient Players Using Limited Monitoring,” Submitted. Available at: http://arxiv.org/abs/1309.0262
- J. Xu, Y. Andreopoulos, Y. Xiao and M. van der Schaar, “Non-stationary Resource Allocation Policies for Delay-constrained Video Streaming: Application to Video over Internet-of-Things-enabled Networks,” Accepted by *IEEE JSAC, Special Issue on Adaptive Media Streaming*.
- L. Canzian, Y. Xiao, W. Zame, M. Zorzi, M. van der Schaar, “Intervention with Private Information, Imperfect Monitoring and Costly Communication: Design Framework," *IEEE Trans. Commun.*, Aug. 2013.
- L. Canzian, Y. Xiao, W. Zame, M. Zorzi, M. van der Schaar, “Intervention with Complete and Incomplete Information: Application to Flow Control,” *IEEE Trans. Commun.*, Aug. 2013.