Reasoning about Incentive Compatibility

Suguman Bansal
Rice University

Incentive Compatibility

- A multi-agent system is incentive compatible if selfish agents interact *truthfully*.
- Incentive compatible systems are more trusted by agents.

Example: Auction Systems

- Bidders are agents
  - **T**: Bidder's actual valuation of the item
  - **B** : Value bidder bids at
- Highest bidder wins the bid
- Bidder utility
  - **T - B**, if bidder wins
  - **0**, otherwise
- Bidder bids **truthfully** if **T == B**
- Auction system is not incentive compatible

Bidder 1 T	Bidder 1 B	Bidder 2 B	Bidder 1 Utility
10	8	7	2
10	10	7	0
10	6	7	0

Problem

All agents in a multi-agent system M are selfish.
Is multi-agent system M incentive compatible?
Generally, what properties are true on multi-agent systems under selfishness assumptions?

Prior Work

Analysis of selfish agents extensively studied
- No unifying framework for analysis of systems of selfish agents
- No automated approaches to analysis of selfish agents

Our Contribution

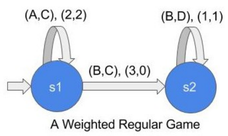
Weighted Regular Games A unifying framework for modeling systems of selfish agents
- Quantitative abstraction of these systems

ComputeNash: An algorithm to compute all Nash equilibria in weighted regular games

Framework: Weighted Regular Games

- Quantitative abstraction of systems of selfish agents
  - Quantitative utilities indicate agent motives
- Finite state model
- Executions in model correspond to collective behavior of agents
  - Action-tuples correspond to collective action of agents
  - Weight-tuples correspond to collective utility of agents
    - i-th weight sequence corresponds to rewards of i-th agent
- Utility of agent with weight sequence A

$$\mathsf{ut}(A, d) = \Sigma_{i=0}^{\infty} \frac{a_i}{d^i} \text{ for } \textit{discount factor } d > 1.$$



A Weighted Regular Game

Nash Equilibria: Analysis in WRG

- An execution is in Nash equilibria if no agent can receive greater utility by unilaterally deviating from the execution
- A notion of collective selfish behavior of all agents
  - Many other notions exist

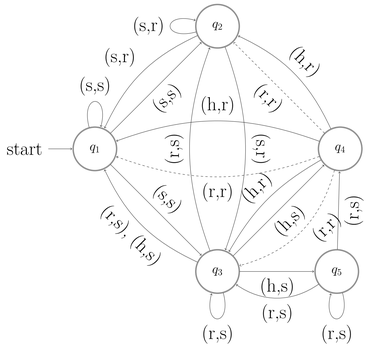
Agnt 2 / Agnt 1	C	D
A	2,2	0,3
B	3,0	1,1

Profile (B,D) is in Nash Equilibria

Case Study: Bitcoin Protocol is not Incentive Compatible

Bitcoin Protocol Model

- A 2-agent game: Dishonest agent A_1, honest agent A_2.
- Agnet A_1 and A_2 Actions:
  - Action s: Searching for a bitcoin.
  - Action r: Releases bitcoin finding one.
- Dishonest agent A_1 Action h: May hide bitcoin after finding it.



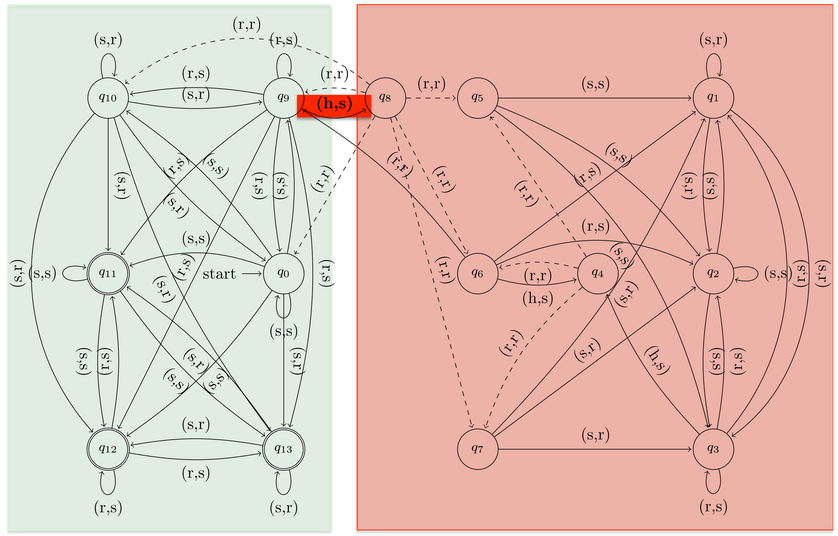
From q1, q2, q3 and q5

Action	r	s,h
Action Utility	1	0

From q4

Action Profile	(h,r)	(r,r)
Action Utility	(0,1)	(2,0) or (0,1)

Figure : Bitcoin Protocol Game, and its weight tuples



- Protocol is not incentive compatible if A_1 acts dishonestly (Action h) in Nash equilibrium
- Incentive Compatibility $IC := \Box \neg (h, \cdot)$
- $IC \not\models$ Nash Equilibria of Bitcoin protocol

Analysis of WRG

Comparing utilities of sequences

Given: Weight sequences A, B, and discount factor $d > 1$,

$$\mathsf{ut}(A, d) > \mathsf{ut}(B, d) \ ?$$

Core Insights
- $\mathsf{ut}(A, d)$ is a number in Base d.

$$\mathsf{ut}(A, d) = a_0 + \frac{a_1}{d} + \frac{a_2}{d^2} + \cdots = (a_0.a_1a_2\ldots)_d$$

- $\mathsf{ut}(A, d) > \mathsf{ut}(B, d)$ iff there exists C s.t.

$$\mathsf{ut}(A, d) = \mathsf{ut}(B, d) + \mathsf{ut}(C, d) \text{ and } \mathsf{ut}(C, d) > 0$$

- C can be guessed non-deterministically using properties of arithmetic in Base d.

ComputeNash

- Computes all Nash equilibria in a WRG
- **Complexity**: Exponential in size of input WRG
- **Observation**: All Nash equilibria forms a regular language

Conclusion

- Presented a unifying framework for analyzing systems of selfish agents
- Analyzed these systems under Nash equilibria

Future Work

- Extension of framework to
  - Probabilistic systems
  - Infinite state models
- Extension to similar results under other notions of rationality
- Adaption of such algorithms to more real world applications.

Collaborators

- Swarat Chaudhuri (Rice)
- Kuldeep S. Meel (Rice)