A Practical Robustness Measure of Incentive Mechanisms (Extended Abstract)

Yuan Liu Jie Zhang Bo An Sandip Sen†
School of Computer Engineering, Nanyang Technological University, Singapore
†Department of Mathematical and Computer Sciences, The University of Tulsa, USA
yliu3,zhangj,boan@ntu.edu.sg, †sandip-sen@utulsa.edu

ABSTRACT

In this abstract, we propose a general robustness definition for a desired equilibrium and a practical robustness measure against a non-equilibrium strategy, inspired by the studies of evolutionary game theory. We also propose a framework to quantitatively evaluate the robustness, and provide theoretical analysis of the framework.

Categories and Subject Descriptors

I.2.11 [**ARTIFICIAL INTELLIGENCE**]: Distributed Artificial Intelligence-Intelligent agents, Multiagent systems

General Terms

Experimentation, Performance

Keywords

Mechanism Design, Bounded Rational Agents, Robustness

1. INTRODUCTION

Game theory provides a powerful theoretical framework to analyze strategic interactions where the payoff of a player depends on its strategy and those of others. Incentive mechanisms, based on game theoretical analysis, have been designed to promote the desired behavior of rational players [5]. However, in realistic scenarios, players with bounded rationality may deviate from the desired equilibrium strategy, causing other rational players to also deviate, and as a result the incentive mechanisms may fail to achieve the expected performance.

The study of the bounded rationality is increasingly important in the field of mechanism design. Bergemann and Morris [1] introduce the concept of bounded rationality for incentive mechanisms where players may make mistakes unintentionally due to reasons unknown to mechanism designers. Filtering approaches have been proposed to eliminate the impact of naive Byzantine (irrational) players who randomly deviate from equilibrium strategies. Moreover, qualitative analysis of the robustness against various attacks has been proposed, where a mechanism is claimed to be either *resistent* or *vulnerable* with respect to an attacking strategy. One common limitation of existing approaches is that they cannot tell to what extent incentive mechanisms are robust against specific non-equilibrium strategies. This work provides an evaluation measurement for researchers to develop a practical incentive mechanism.

Appears in: *Alessio Lomuscio, Paul Scerri, Ana Bazzan, and Michael Huhns (eds.), Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), May 5-9, 2014, Paris, France.*
Copyright © 2014, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

2. ROBUSTNESS MEASURE

Our robustness measure of incentive mechanisms is inspired by the concept of evolutionary stable strategy (ESS) as studied in evolutionary game theory — the application of game theory to biology [4] — to study the evolution of strategies in the population(s) of individuals who are competing with each other for survival and reproduction. Players' equilibrium strategy profile[1] is called a desired strategy profile (equilibrium) which leads to the desired outcome of the mechanism, assuming that the players are rational. Players who instead adopt a non-equilibrium strategy is regarded as *bounded rational* players.

An incentive mechanism in a game has n players $I = \{1, ..., n\}$ which are modeled by n populations: for each player position, there is a large population of individuals (agents), and each such individual is choosing a fixed action or following a pure strategy. Let $S_i = \{s_1^i, ..., s_{m_i}^i\}$ be the set of pure strategies available to the individuals of population i and $\Delta_i = \{x_i \in \mathbb{R}^{m_i} | \sum_{s_j^i \in S_i} x_i(j) = 1, x_i(j) \geq 0, j = 1, ..., m_i\}$ be the set of possible strategy profiles for population i, i.e., each $x_i(j)$ corresponds to the fraction of individuals in the population i playing strategy $s_j^i \in S_i$. Note that x_i is formally equivalent to a mixed strategy for the player $i \in I$ in the n-player game. The combination of n population profiles is $x \in \Theta$ where $\Theta = \times_{i \in I} \Delta_i$. Suppose $\epsilon \in (0, 1)$ proportion of the population i invade the mechanism by playing a different profile $y_i \in \Delta_i$. As a result, the new population profile for population i becomes $x_i^\epsilon = \epsilon \cdot y_i + (1-\epsilon) x_i$. If we denote the set of payoff functions for all populations as $u = \{u_1, ..., u_n\}$ where the payoff of an agent in population i taking strategy s_j^i is $u_i(s_j^i, x_{-i})$. The average payoff of the rational players in the population i before and after the invasion can be expressed as $u_i(x_i, x_{-i}) = \sum_{j=1}^{m_i} x_i(j) u_i(s_j^i, x_{-i})$ and $u_i(x_i, x_{-i}^\epsilon) = \sum_{j=1}^{m_i} x_i(j) u_i(s_j^i, x_{-i}^\epsilon)$ respectively and the average payoff of the mutant invaders in population i is $u_i(y_i, x_{-i}^\epsilon) = \sum_{j=1}^{m_i} y_i(j) u_i(s_j^i, x_{-i}^\epsilon)$. Now, we can describe a general mechanism as $\mathcal{M} = \{I, \Theta, u\}$ and the robustness of a desired equilibrium can be defined as follows.

Definition 1. *[Robustness of a desired equilibrium] Given an incentive mechanism $\mathcal{M} = \{I, \Theta, u\}$ with a desired equilibrium $x \in \Theta$, the robustness of x is R such that*

$$R = \arg\max_{\epsilon \in [0,1]} u_i(x_i, x_{-i}^\epsilon) > u_i(y_i, x_{-i}^\epsilon), \forall y \in \Theta (y \neq x), \forall i \in I. \quad (1)$$

In other words, the robustness of a desired equilibrium is the maximum proportion of invaders such that the desired equilibrium strategy is still the best strategy for rational players in each population.

However, the computational cost of formally analyzing the robustness measure in Definition 1 makes such analysis infeasible.

[1]For ease of analysis, we consider a single desired equilibrium of a mechanism. It can be extended to handle multiple equilibria.

First, the size of the population corresponding to each player position is assumed to be infinite in the definition so as to model all possible mixed strategies of the player. The payoff of a population is thus difficult to calculate through aggregating all the individuals' utilities. Second, the strategy of the mutants $y \neq x$ is continuous and it may be impossible to verify Equation (1) for all $y \in \Theta$. Moreover, the size of players I in a realistic game could be very large, making the robustness measure computationally prohibitive. Given the difficulty of formally analyzing the robustness of incentive mechanisms, we consider only finite populations to be able to realistically measure the robustness of these incentive mechanisms.

Moreover, the mutants or invaders in a realistic mechanism always attack the mechanism using several mature techniques or strategies, e.g., constant attack and whitewashing attack in reputation systems [2]. It is also observed that these attacks are typically conducted by a single player. Thus, it is practical to measure the robustness of an incentive mechanism against representative types (and their combinations) of mutants in a single population.

We note that the robustness measure in Definition 1 is a static concept. A mechanism is always implemented in real time and players could dynamically adjust their strategies as time goes on, which could be modeled by the well-known replicator dynamics (RD) [4] model for capturing long-range evolutionary dynamics with natural selection forces. If agents using the desired equilibrium strategy could achieve higher payoff (fitness) in a population, then they are more likely to reproduce and produce offsprings which use the same strategy (replicator) and eventually through this process the strategy distribution of the population will converge to a stable equilibrium profile.

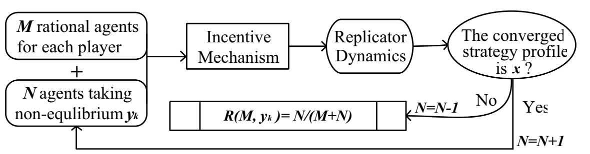

Figure 1: Evaluating Robustness based on Replicator Dynamics

We now propose an evolutionary dynamics based practical robustness measure that evaluates the robustness of an equilibrium against any specific non-equilibrium strategy (Figure 1). In this approach, agents adopting non-equilibrium strategy $y_k \neq x_k$ are gradually added to the original *rational population* (population without mutants) and the resultant populations repeatedly interacts under the mechanism. The strategy distribution of the rational populations evolves through RD until convergence. Successively more mutants are added until the population converges to distribution different from the desired equilibrium.

DEFINITION 2. *[Robustness against a non-equilibrium strategy] The robustness of an incentive mechanism $\mathcal{M}$, with desired equilibrium strategy profile x, against a non-equilibrium strategy y_k, used by bounded-rational invaders, is*

$$R(M, y_k) = \frac{N}{N+M} \tag{2}$$

where M is the finite population size for each player position and N is the maximum number of bounded rational invaders such that the converged strategy profile, under RD, does not deviate from the desired equilibrium.

We assume that each player position has the same population size and the robustness of an incentive mechanism is a function of the population size and the non-equilibrium strategy.

However, there are still two challenges imposed on theoretically evaluating the robustness of a realistic incentive mechanism: (1) the strategy y_k launched by bounded rational agents may be too complex to be theoretically modeled; (2) the settings of incentive mechanisms may be too complex to be theoretically abstracted.

3. EVALUATION FRAMEWORK

Therefore, based on the practical robustness measure defined in Definition 2, we have proposed a simulation based framework in [3] to quantitatively evaluate the robustness of an incentive mechanism against a non-equilibrium strategy. In this framework, we gradually add agents taking a specific non-equilibrium strategy and let the populations evolve over time to observe the converged strategy profile. We stop inserting more agents when the equilibrium strategies are abandoned by the rational populations in the converged population profile after a sufficiently large number of evolutionary dynamics steps with a small probability ε. Regarding the parameter ε, we study its effects in measuring the robustness value in the evaluation framework.

PROPOSITION 1. *There exists $\varepsilon = G(M)$ where $\frac{\mathrm{d}G(M)}{\mathrm{d}M} < 0$ such that the evaluated robustness of a mechanism is independent of the population size M.*

PROPOSITION 2. *When $\varepsilon = \varepsilon_0$ is a constant, the robustness of a mechanism increases with the population size M where $M > \frac{1}{\varepsilon_0} + 1$, and finally converges at a value when the agents of a population taking a non-equilibrium strategy can achieve $\frac{\varepsilon_0}{1-\varepsilon_0}$ times payoff more than that of the desired equilibrium strategy.*

We have shown that the robustness of an incentive mechanism measured in the evaluation framework is an increasing function of the population size and is lower bounded, provided that ε is a constant. As ε is set to be very large (e.g. 1), then the framework would allow the non-equilibrium to achieve $\frac{\varepsilon_0}{1-\varepsilon_0} \to +\infty$ times payoff than the equilibrium strategy when the population is large. To avoid this impractical phenomenon, in the simulations of [3], we thus set $\varepsilon = 0.05$ and gradually increase the population size until the robustness value converges.

4. CONCLUSION AND FUTURE WORK

In this abstract, we have proposed formal definitions of a robustness measure for incentive mechanisms in the presence of bounded rational players. In Future work, the evaluation framework [3] will be extended to search for the worst non-equilibrium strategy, and also study the robustness of incentive mechanisms when rational players form coalitions to take cooperative strategies.

5. ACKNOWLEDGMENT

This work is partially supported by the A*Star SERC grant (1224104047) awarded to Dr. Jie Zhang.

6. REFERENCES

[1] D. Bergemann and S. Morris. *Robust Mechanism Design: The Role of Private Information and Higher Order Beliefs*. World Scientific Publishing and Imperial College Press, 2012.

[2] S. Jiang, J. Zhang, and Y.-S. Ong. An evolutionary model for constructing robust trust networks. In *Procedings of AAMAS*, pages 813–820, 2013.

[3] Y. Liu and J. Zhang. Robustness evaluation of incentive mechanism. In *Proceedings of AAMAS*, pages 1293–1294, 2012.

[4] W. H. Sandholm. *Evolutionary Game Theory*. University of Wisconsin, 2007.

[5] J. Zhang, R. Cohen, and K. Larson. Combining trust modeling and mechanism design for promoting honesty in e-marketplaces. *Computational Intelligence*, 28(4):549–578, 2012.