Coalitional Games in Open Anonymous Environments

Makoto Yokoo†, Vincent Conitzer‡, Tuomas Sandholm‡, Naoki Ohta†, Atsushi Iwasaki†

†Faculty of Information Science
and Electrical Engineering
Kyushu University
6-10-1 Hakozaki, Higashi-ku,
Fukuoka, 812-8581 Japan
yokoo/oota/iwasaki@is.kyushu-u.ac.jp

‡Computer Science Department
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213 USA
conitzer/sandholm@cs.cmu.edu

1 Introduction

Coalition formation is a key capability in automated negotiation among self-interested agents. A coalition of agents can sometimes accomplish things that the individual agents cannot, or can do things more efficiently. To make coalition formation successful, a key question that must be answered is how the gains from cooperation are to be distributed. *Coalitional game theory* provides a number of solution concepts for this. Some of these solution concepts have already been adopted in the multi-agent systems literature [Zlotkin and Rosenschein, 1994; Shehory and Kraus, 1998].

However, existing solution concepts have limitations when applied to open anonymous environments such as the Internet. In such environments, a single agent can use multiple identifiers (or *false names*), pretending to be multiple agents, and distribute its ability (skills) among these identifiers. Alternatively, multiple agents can collude and pretend to be a single agent that combines all of their skills. Furthermore, an agent might try to hide some of its skills.

These manipulations are virtually impossible to detect in open anonymous environments, and have thus become an issue in such environments specifically. That is also the reason why the gamut of these manipulations has not received much research attention previously. In this paper, we develop a new solution concept for coalitional games called the *anonymity-proof core*, which is robust to the manipulations described above. We show that the anonymity-proof core is characterized by certain axiomatic conditions (including that an agent does not have an incentive to use the basic manipulations mentioned above).

2 Model

Traditionally, value division in coalition formation is studied in *characteristic function games*, where each potential coalition (i.e., each subset X of the agents) has a value $w(X)$ that it can obtain. This assumes that utility is transferable (e.g., utility can be transferred using side payments), and that a coalition's value is independent of what non-members of the coalition do. The characteristic function by itself does not give us sufficient information to assess what manipulations may be performed by agents in an open anonymous environment. Thus, instead of defining the characteristic function over agents, we define it over *skills* that the agents possess.

Definition 1 (skills and agents) *Assume the set of all possible skills is T. Each agent i has one or multiple skills $S_i \subset T$. For simplicity, we assume each skill is unique, that is, $\forall i \neq j, S_i \cap S_j = \emptyset$ holds.*

Definition 2 (characteristic function defined over skills) *A characteristic function $v : 2^T \rightarrow \Re$ assigns a value to each set of skills.*

We will denote by w the characteristic function defined over agents, and by v the characteristic function defined over skills. For a given set of agents X, let $S_X = \bigcup_{i \in X} S_i$. Then, we have $w(X) = v(S_X)$. The characteristic function over skills is a more fine-grained representation than the characteristic function over agents.

We assume the following three types of manipulation (or any combination of them) are possible for agents.

hiding skills: If agent i has a set of skills S_i, for any $S'_i \subseteq S_i$, it can declare that it has only S'_i.

false-name: Agent i can use multiple identifiers and declare that each identifier has a subset of S_i.

collusion Multiple agents can collude and pretend to be a single agent, who has the union of their skills.

3 Manipulability of Traditional Solution Concepts

In this section, we show that a well-known (perhaps the best known) solution concept called the *core* [Gillies, 1953; von Neumann and Morgenstein, 1947] is vulnerable against these manipulations. This is true for other solution concepts such as *Shapley value* [Shapley, 1953], *least-core*, or *nucleolus*.

Definition 3 (core) *Given a set of agents W, an outcome, that is, a value division $c^W = (c_1^W, c_2^W, \ldots)$ among agents is in the core if the following two conditions hold:*

1. $\forall X \subset W, \sum_{i \in X} c_i^W \geq w(X)$,
2. $\sum_{i \in W} c_i^W = w(W)$.

The first condition is called the *non-blocking* condition: if this condition does not hold for some set of agents X, then the agents in X have an incentive to collectively deviate from the mechanism and to divide $w(X)$ among themselves. The second condition is called the *feasibility* condition: if $\sum_{i \in W} c_i^W > w(W)$, this outcome is infeasible.

Due to the space limitation, we only show an example where the core is manipulable by a collusion.

Example 1 *Let there be three skills a, b, and c. Let the characteristic function over skills be as follows.*

- $v(\{a,b\}) = v(\{a,c\}) = v(\{a,b,c\}) = 1$,
- $v(\{a\}) = v(\{b\}) = v(\{c\}) = v(\{b,c\}) = 0$.

Let there be three agents $1, 2$, and 3 with skills a, b, and c, respectively. Then, the characteristic function over agents is as follows.

- $w(\{1,2\}) = w(\{1,3\}) = w(\{1,2,3\}) = 1$,
- $w(\{1\}) = w(\{2\}) = w(\{3\}) = w(\{2,3\}) = 0$.

In this example, there is only one outcome in the core, namely outcome $(1, 0, 0)$. This is because if agent 2 (or 3) obtains any value, then the non-blocking condition is violated because agent 1 and agent 3 (or 2) have an incentive to deviate from the mechanism and form their own coalition. We can see that since the skill b and c are completely substitutable, the agents who have these skills have no bargaining power. The least core and the nucleolus also give this outcome.

Now, let us assume that agent 2 and 3 collude and pretend to be a single agent $2'$, who has b and c.

Then, the characteristic function over agents is as follows.

- $w(\{1\}) = w(\{2'\}) = 0$,
- $w(\{1,2'\}) = 1$.

Now, agent 1 and $2'$ are symmetric and have the same bargaining power. Then, if we use the least core or the nucleolus, each agent receives $1/2$. Thus, agent 2 and 3 can increase their value division using a collusion.

4 Anonymity-Proof Core

We develop a new solution concept for our setting which we call *anonymity-proof core*. As we will show, the anonymity-proof core can be characterized by certain axiomatic conditions.

We require that the outcome function π is *anonymous*, that is, the payoff to an agent does not depend on the identifiers of the agents; it depends only on the skills of the agent and the distribution of skills over other agents.

More specifically, given an agent i and a set of other agents Y, let S_i be the set of skills that agent i has, and let $SS_Y = \{S_j \mid j \in Y\}$, where S_j is the set of skills that agent j has. Then, the outcome function $\pi(S_i, SS_Y)$ takes S_i and SS_Y as arguments, and returns the payoff to agent i, when agent i declares its skills as S_i and the other agents declare their skills as SS_Y.

Let the set of agents who joined the mechanism be W, and let the profile of the skills that the agents declared be $k = (k_1, k_2, \ldots)$, where k_i is the set of skills that agent i declared. Let $S_X = \bigcup_{i \in X} k_i$, that is, S_X is the union of the skills declared by a set of agents X; let $S = S_W$; and let $SS_X = \{k_i \mid i \in X\}$. Also, let $SS_{\sim i} = \{k_1, \ldots, k_{i-1}, k_{i+1}, \ldots\}$, that is, a set, each of whose elements is the set of skills corresponding to agent j ($j \neq i$).

We now give six axiomatic conditions that the outcome function π should satisfy.

1. The outcome function π is anonymous.
2. π is never blocked by any coalition, that is, $\forall k, \forall X \subseteq W$, $\sum_{i \in X} \pi(k_i, SS_{\sim i}) \geq v(S_X)$ holds.
3. π is always feasible and always distributes all of the value, that is, $\forall k, \sum_{i \in W} \pi(k_i, SS_{\sim i}) = v(S)$ holds.
4. π is robust against hiding skills, that is, $\forall S', S''$, where $S'' \subseteq S'$, $\forall SS, \pi(S'', SS) \leq \pi(S', SS)$ holds.
5. π is robust against false-name manipulations, that is, $\forall k, \forall X \subseteq W$, $Y = W \setminus X$, $\sum_{i \in X} \pi(k_i, SS_{\sim i}) \leq \pi(S_X, SS_Y)$ holds.
6. π is robust against collusions, that is, $\forall k, \forall X \subseteq W$, $Y = W \setminus X$, $\sum_{i \in X} \pi(k_i, SS_{\sim i}) \geq \pi(S_X, SS_Y)$ holds.

In order to define the anonymity-proof core, we first formally define the core for skills. For a set of skills $S = \{s_1, s_2, \ldots\}$, we define a set of core outcomes for skills $Core(S)$ as follows.

Definition 4 (core for skills) $c^S = (c^S_{s_1}, c^S_{s_2}, \ldots)$ *is in* $Core(S)$ *if it satisfies the following two conditions.*

- $\forall S' \subset S, \sum_{s_j \in S'} c^S_{s_j} \geq v(S')$,
- $\sum_{s_j \in S} c^S_{s_j} = v(S)$.

Now we are ready to define the anonymity-proof core.

Definition 5 (anonymity-proof core) *We say the outcome function π_{ap} is in the anonymity-proof core if π_{ap} satisfies the following two conditions.*

1. *For any set of skills $S \subseteq T$, there exists a core outcome for S, that is, some $c^S = (c^S_{s_1}, c^S_{s_2}, \ldots) \in Core(S)$, such that for any skill profile $k = (k_1, k_2, \ldots,)$ with $\bigcup_i k_i = S$, $\pi_{ap}(k_i, SS_{\sim i}) = \sum_{s_j \in k_i} c^S_{s_j}$ holds.*
2. *$\forall S', S''$, where $S'' \subseteq S'$, $\forall SS$, $\pi_{ap}(S'', SS) \leq \pi_{ap}(S', SS)$ holds.*

The first condition says that for any set of skills reported by the agents, some outcome in the core for that set of skills should be used to distribute the value. The second condition says that an agent has no incentive to hide (some of) its skills.

The following theorems show that the anonymity-proof core is characterized by the six axiomatic conditions. We omit the proof due to space constraint.

Theorem 1 *Any outcome function π_{ap} in the anonymity-proof core satisfies the six axioms.*

Theorem 2 *Any outcome function π that satisfies the six axioms is in the anonymity-proof core.*

References

[Gillies, 1953] Donald Gillies. *Some theorems on n-person games*. PhD thesis, Princeton University, Dept of Mathematics, 1953.

[Shapley, 1953] Lloyd S. Shapley. A value for n-person games. In H Kuhn and A Tucker, eds, *Contributions to the Theory of Games*, vol 2 of *Annals of Mathematics Studies, 28*, 307–317. Princeton U Press.

[Shehory and Kraus, 1998] Onn Shehory and Sarit Kraus. Methods for task allocation via agent coalition formation. *Artificial Intelligence*, 101(1–2):165–200, 1998.

[von Neumann and Morgenstern, 1947] John von Neumann and Oskar Morgenstern. *Theory of games and economic behavior*. Princeton University Press, 1947.

[Zlotkin and Rosenschein, 1994] Gilad Zlotkin and Jeffrey S Rosenschein. Coalition, cryptography and stability: Mechanisms for coalition formation in task oriented domains. In *AAAI*, pages 432–437, 1994.