The Korean Economic Review
Volume 39, Number 1, Winter 2023, 241-256.
DOI : 10.22841/kerdoi.2023.39.1.008

Strong Collusion-Proof Implementation*

Jinwoo Kim**

The robustly collusion-proof (RCP) mechanism proposed by Che and Kim (2006) enables a principal to attain any payoff that could be achieved without any collusion, even when agents collude. Although the RCP mechanism is robust to various collusive arrangements that agents may devise, it relies on agents not to form certain extreme beliefs following a rejection of a collusive side contract. This paper strengthens the collusion-proofness notion to be robust to such beliefs, as well as any other aspects of coalition formation and its behavior, and proposes a mechanism that implements virtually any non-collusive payoff for the principal in this considerably strong collusion-proof sense. The key issue is to guarantee the participation of agents in an RCP mechanism. The proposed mechanism achieves this situation by adding an option that each agent can exercise to protect himself against possible hold-up by his collusive partners.

JEL Classification: D47, D82
Keywords: Robustly Collusion-proof Implementation, Strongly Collusion-proof Implementation, Virtual Strongly Collusion-proof Implementation

I. Introduction

Che and Kim (2006) (CK hereafter) have proposed a method to deal with collusion among agents who are privately informed of their types. They particularly show that, even in the face of collusive agents, a principal can attain any surplus level that can be achieved without collusion, by shifting the entire payoff risks to the coalition members. This idea of "selling the firm to the coalition" requires

Received: Oct. 8, 2021. Revised: June 6, 2022. Accepted: July 15, 2022.
* I am grateful to the comments of Yeon-Koo Che. This research was supported by the Ministry of Education of the Republic of Korea and National Research Foundation of Korea (NRF-2019S1A5A2A01049533). Financial support from the Center for National Competitiveness in the Institute of Economic Research of Seoul National University is gratefully acknowledged.
** Professor, Department of Economics and SIER, Seoul National University, Republic of Korea; jikim72@snu.ac

minimally in the way of restricting agents' private information (e.g., multidimensional types, interdependence and correlation) or their preferences and/or technologies. However, it does require collusive side contracts to be *interim incentive compatible* and *interim individually rational*. The former involves no restriction because coalition members do not observe each other's private information. Meanwhile, the latter is not as well justified. Once an agent signs on the principal's contract, the former is no longer guaranteed his reservation payoff. In fact, his continuation payoff may fall below that level against some, possibly non-equilibrium, play by his opponents. With such an outcome being used as a threat, a collusive agreement may credibly force some (type of) agent to sustain a below-reservation payoff.[1] Anticipating such a scenario, agents may occasionally refuse to participate a robustly collusion-proof (RCP) mechanism, which could undermine the implementation of the desired outcome.

To illustrate this problem, suppose a buyer wishes to procure a good valued at \$1 from one of two suppliers, S1 and S2. S1 has a known cost of 2/3 and S2 has privately known cost distributed uniformly over [0, 23]. The second-best outcome for the buyer, if collusion were not a problem, would be to procure from S2 for payment of 1/3 if his cost is less than 1/3, but otherwise from S1 for payment of 2/3 even though she is less efficient. Thus, the buyer can procure the good at the expected cost of 1/2. Such a mechanism would be susceptible to collusion by S1 and S2. The latter can always announce that his cost is higher than 1/3. Thereafter, S1 will win the contract for a payment of 2/3 from the buyer, and he secretly subcontracts from the more efficient S1. Suppliers are jointly better off in this way; and the buyer is worse off because her cost of procurement is 2/3 instead of 1/2.

CK's RCP mechanism solves this problem. In the RCP mechanism, the buyer selects S1 as a prime contractor, requires him to deliver the good for a fixed fee of 1/2, with the instruction that S2 should supply the good for a fee of 1/3 paid by S1 when the former's cost is less than 1/3; otherwise, S1 should supply the good himself. Given this contract, which supplier produces the good is of no concern to the buyer; as long as both suppliers accept the buyer's contract, the buyer gets the good at the optimal price of 1/2. Indeed, it is reasonable to expect both suppliers accept the contract. Clearly, S2 can never lose from participating in the mechanism because he retains an option not to produce. A similar reasoning appears to work for S1. By charging the price of 1/3 to S2, S1 breaks even on average.[2] Hence, S1 will participate.

---

[1] Collusive agreement could serve a predatory purpose against a minority group of agents or against only some types to attain better joint payoffs on average.

[2] With probability 1/2, S2's cost is less than 1/3. Thus, the latter will accept to deliver the good to S1 for a fee of 1/3, so S1's cost of delivering the good to the buyer is 1/3 in this case. With the remaining probability 1/2, S2 will reject the offer. Hence, S1 will have to produce the good by himself, costing him 2/3. The expected cost of procurement is $1/2=(1/2)(1/3)+(1/2)(2/3)$.

Although the preceding argument is reasonable, it involves restriction on agents' beliefs on how the collusion game will proceed. To illustrate, suppose both parties participate in the aforementioned mechanism but (some type of) S2 makes a collusive proposal to S1 that requires S2 to supply the good for a fee of $P = 2/3$ (to be paid by S1). S1 will never accept such a proposal if he forms a "passive belief." That is, if he believes that when he refuses such a proposal, then S2 will act in a non-cooperative fashion, accepting to produce the good if and only if his cost is less than 1/3 (In this case, the second-best outcome is once again implemented). Suppose instead S1 believes that, following his refusal of S2's offer, S2 will report 2/3 regardless of his type. Accordingly, S1 will be forced to supply the good at the cost of 2/3 for a payment only of 1/2 from the buyer. Hence, given such a (out-of-equilibrium) belief, S1 will accept the fee of $P = 2/3$. Clearly, this situation entails a net loss (of 1/6=2/3–1/2) for S1. Hence, if S1 anticipates this course of behavior from participating in the buyer's mechanism, then he will not participate. The second best outcome is not implemented in this case.[3] Although the belief that leads to the failure of the RCP mechanism appears unreasonable, it reflects one difficulty with this mechanism: in practice, the principal may have difficulty controlling agents' relative bargaining power in their collusive negotiation.

This paper suggests a method to significantly strengthen CK's notion of collusion-proof to solve the preceding problem. In fact, the current notion will require no restriction on collusive behavior, except that collusion occurs after the contract acceptance decision, which is a key assumption of CK and Laffont and Martimort (1997, 2000). First, we do not require agents to collude constantly: collusion may involve only some agents and may occur only occasionally when "the condition is right."[4] Second, our collusion-proof contract works regardless of the number of coalitions, whom each coalition comprises, how they operate, and who proposes what type of side contracts. In particular, the mechanism can deal with only subsets of agents forming subcoalitions. A side contract can be proposed by a third party or by one of the informed agents, or can even be a consequence of negotiations among several agents. Third, the contract design requires no knowledge of any of these matters. Lastly, we do not restrict the agents' out-of-equilibrium beliefs. Improvements gained on these four dimensions are substantial.

The proposed mechanism builds on the RCP mechanism. Hence, rather than attempting to prevent agents from engaging in collusion, the mechanism shifts payoff risks to agents, so that the principal is guaranteed the desired payoff level. This construction effectively solves the collusion problem as long as all agents participate in the mechanism. Therefore, the key issue is to ensure agents'

[3] Given non-participation by S1, the buyer may either cancel the procurement or buy the good from S2 for a fee. Either way, the second best outcome is not implemented.

[4] Therefore, we relax the assumption often made implicitly in the literature, including CK, in which the collusive proposal needs to be accepted by all types of agents.

participation. To do so, the mechanism augments the message space of the RCP mechanism. The added message acts as an option that agents can exercise to protect them from possible hold-up by their coalitional partners, thereby creating a strict incentive for participation.

The augmented mechanism involves an "integer game," which is often used in the classic implementation literature to rule out undesirable equilibria. Although such games may not be appealing, there is often no compelling rationale for limiting message spaces. At minimum, it is useful to understand whether or not an outcome is implementable, given no restriction on the message space. Our use of the integer game is in similar spirit; it helps us clarify the theoretical upper bound on the extent to which collusion problems can be solved.

II. Strong Collusion-Proofness

We begin with a brief summary of the CK model. In particular, there are one principal and a set $N := \{1,...,n\}$ of $n \ge 2$ agents. Each agent i has type θ_i drawn from an arbitrary set, Θ_i. Let us denote $\theta := (\theta_1,...,\theta n) \in \times_{i=1}^{n} \Theta_i =: \Theta$ and assume that θ is distributed according to some prior distribution $\mu \in \Delta\Theta$. For an agent $i \in N$, let $\mu(\cdot \mid \theta_i) \in \Delta\Theta_{-i}$ denote the distribution of other agents' type profile, conditional on agent i's type being θ_i.

Given a type profile $\theta \in \Theta$, if the principal chooses allocation $q \in Q$ and pays $t_i \in \mathbb{R}$ to each agent $i \in N$, then the latter receives as follows:

$$s_i(q,\theta) + t_i,$$

while the former receives:

$$v(q) - \sum_{i \in N} t_i .$$

Allocation space Q is arbitrary but includes a null allocation $\emptyset$, such that $s_i(\emptyset,\cdot) = 0$. Note that each agent's payoff function can depend on other agents' types, allowing for the "interdependent values" case. Agents are said to have "private values" if each agent's payoff is independent of others' types. That is, for each $i \in N$, $s_i(q,\theta) = s_i(q,\theta_i)$, $\forall q \in Q$, $\forall \theta \in \Theta$.

A direct mechanism $M = (q,t)$ consists of an allocation rule $q : \Theta \to Q$ and transfer rule $t = (t_1,\ldots,t_n) : \Theta \to \mathbb{R}^n$. We say M is interim individually rational if:

$$U_i^M(\theta_i) := \mathbb{E}_{\tilde{\theta}_{-i}}[s_i(q(\theta_i,\tilde{\theta}_{-i}),\theta) + t_i(\theta_i,\tilde{\theta}_{-i}) \mid \theta_i] \ge 0, \quad \forall i, \theta_i,$$

where the reservation utility level is normalized to zero. It is *interim incentive compatible* if:

$$U_i^M(\theta_i) \geq \mathbb{E}_{\tilde{\theta}_{-i}}[s_i(q(\theta_i', \tilde{\theta}_{-i}), \theta) + t_i(\theta_i', \tilde{\theta}_{-i}) \mid \theta_i], \quad \forall i, \theta_i, \theta_i'.$$

Let $\mathcal{M}$ denote the set of all direct mechanisms that are interim individually rational and interim incentive compatible. The set $\mathcal{M}$ includes all possible equilibrium outcomes that may arise from some contracts. These outcomes need not have all agents participate in some contract. For example, an outcome in which an agent accepts a contract only occasionally or not at all is consistent with interim incentive compatibility and individual rationality. A mechanism $M = (q, t) \in \mathcal{M}$ is said to *implement* an (expected) payoff of $V \in \mathbb{R}$ if:

$$V = \mathbb{E}\left[v(q(\tilde{\theta})) - \sum_{i \in N} t_i(\tilde{\theta}) \right].$$

Let $\mathcal{V}$ denote the set of all implementable payoffs. We particularly focus on its subset $\mathcal{V}_0$, such that every $V \in \mathcal{V}_0$ is implementable by $M \in \mathcal{M}$ that gives strictly positive payoffs to μ-a.e. type of each agent. There is minimal loss in restricting attention to $\mathcal{V}_0$. Often, $\mathcal{V}_0 = \mathcal{V}$,[5] and even if $\mathcal{V} \not\subset \mathcal{V}_0$ (as with the full extraction outcome of Crémer and McLean (1985, 1988)), any $V \in \mathcal{V}$ can be implemented arbitrarily closely by a mechanism that gives strictly positive payoffs to all agents. Hence, $cl(\mathcal{V}_0) = \mathcal{V}$.

The basic extensive form is described as follows. First, each agent $i \in N$ draws his type $\theta_i \in \Theta_i$. Second, the principal offers a contract $C = (B, \xi, \tau)$, which consists of *message spaces* $B = (B_1, \ldots, B_n)$ and *outcome function* $(\xi, \tau): B \to Q \times \mathbb{R}^n$ that maps each message profile to a decision $q \in Q$ and transfers $t \in \mathbb{R}^n$. Third, the agents decide, independently and simultaneously, whether to accept or reject the contract. If some agent rejects C, then an accepting agent receives a small transfer $\varepsilon > 0$ and is assigned a null allocation. If all agents accept the contract, then contract C takes effect. Once the acceptance decision is made, coalitions of agents may be formed, and their collusive proposals may be proposed and agreed upon. Lastly, agents are asked to send messages from B, and the outcome $(q, t) \in Q \times \mathbb{R}^n$ is decided based on the outcome function. What messages are submitted and what precise outcome arises depend on the existing (or non-existing) collusive agreement. To be as inclusive as possible in modeling collusion, the precise extensive form governing collusion is left unspecified. The

[5] For example, in a standard problem satisfying a single crossing property and uncorrelated types, all but the worst type of agent enjoy strictly positive payoff even for the second-best payoff, sup $\mathcal{V}$.

only important assumption, which is retained from CK, is that coalitions are formed only after agents make the acceptance decision non-cooperatively.[6]

Instead of directly modeling collusion, we study the outcomes that may arise as a result of collusive manipulation. Formally, let $\sigma = (\sigma_1, \ldots, \sigma_n) : \Theta \mapsto B$ denote the strategy profile that agents may adopt to submit their messages, *possibly as a result of collusion*.[7] Notice that we allow the strategy σ_i by each agent i to depend on possibly the entire profile of types $\theta \in \Theta$, and not just his own. The reason is that agents may coordinate their reports after some round of communication of their types within coalitions they belong to. We are interested in the outcome that arises following the principal's offering of C. We say that a direct mechanism $\tilde{M} = (\tilde{q}, \tilde{t}) : \Theta \mapsto Q \times \mathbb{R}^n$ is *a feasible outcome of* C, if there exists $\sigma : \Theta \mapsto B$, such that, for μ-a.e. $\theta \in \Theta$,

$$(RE) \quad v(\tilde{q}(\theta)) = v(\xi(\sigma(\theta)))$$
$$(BB) \quad \sum_{i \in N} \tilde{t}_i(\theta) = \sum_{i \in N} \tau_i(\sigma(\theta)),$$

and

$$(IC) \quad \tilde{M} \;\; \textit{is interim incentive compatible}.$$

The first two conditions specify the (minimal) technological constraints for collusive manipulation. Condition (*RE*) reflects possible reallocation ability of the coalition. Formally, it means that collusive agents can reallocate $\tilde{q}$ as long as it is does not affect the principal's payoff. For example, in an auction of a single item, coalition members can reallocate the good once a member wins the item. However, they cannot get it allocated to some agent if no agent wins the auction. Condition (*BB*) means that side transfers must be budget balanced among the entire set of agents, which is necessary for side transfers within each coalition to be budget balanced. Lastly, condition (*IC*) reflects the informational asymmetry facing coalition members. Note that the three conditions are necessary requirements of any Bayes Nash equilibrium, irrespective of the specific structure of coalitions (e.g., whether or not there is active coalition, how many there are, who proposes the proposal).

We focus on the outcomes implemented by the Bayes Nash equilibria following an (arbitrary) contract C. Let $\mathcal{M}(C) \subset \mathcal{M}$ be the set of all such outcomes. No presumption is made on agents' participation in C for a given equilibrium

[6] As noted in CK, there are important extensions that consider collusion at the participation stage (see Che and Kim, 2007; Pavlov, 2006).

[7] Possible randomization over messages can be introduced, just as in CK, without affecting our results. We do not consider randomization to economize on space.

outcome; an outcome in $\mathcal{M}(C)$ may involve partial or non-participation by some agents. Clearly, every element of $\mathcal{M}(C)$ must be a feasible outcome of C (in the sense of satisfying the preceding three conditions).

Definition 1. *A payoff $V \in \mathcal{V}$ is* ***strongly collusion-proof (SCP) implementable*** *if there exists a contract C, such that (i) there exists a (weak) perfect Bayesian equilibrium of the game following C and that (ii) every $\tilde{M} \in \mathcal{M}(C)$ implements exactly V.*

Definition 2. *A payoff $V \in \mathcal{V}$ is* ***virtually strongly collusion-proof (VSCP) implementable*** *if for each $\varepsilon > 0$ there exists a contract C, such that (i) there exists a (weak) perfect Bayesian equilibrium of the game following C and that (ii) every $\tilde{M} \in \mathcal{M}(C)$ implements $V' > V - \varepsilon$.*

This notion of collusion-proofness is inclusive of all existing notions. If a payoff V is SCP implementable, then it is RCP implementable in the sense of CK and also SCP implementable in the sense of Laffont and Martimort (1997, 2000).[8] This notion is considerably stronger than any existing notion because no assumption is made on agents' collusive behavior (except for its timing) or the principal's knowledge of it.

III. SCP Implementation

In this section, we propose a method for SCP implementing any payoff that a principal may attain in the absence of collusion. Fix any $V \in \mathcal{V}_0$ and consider a mechanism $M = (q,t) \in \mathcal{M}$ that implements V with strict positive payoff accruing to μ-a.e. type of each agent.

In the case of independent types, CK's RCP implementation of M is based on the following mechanism: $\hat{M} = (\hat{q},\hat{t})$, such that $\hat{q}(\cdot) := q(\cdot)$ and that for each $\theta \in \Theta$:

$$\hat{t}_i(\theta) = \kappa_i v(q(\theta)) + \mathbb{E}_{\tilde{\theta}_{-i}}[t_i(\theta_i,\tilde{\theta}_{-i}) - \kappa_i v(q(\theta_i,\tilde{\theta}_{-i}))]$$
$$- \frac{1}{n-1}\sum_{j \neq i} \mathbb{E}_{\tilde{\theta}_{-j}}[t_j(\theta_j,\tilde{\theta}_{-j}) - \kappa_j v(q(\theta_j,\tilde{\theta}_{-j}))] - \rho_i,$$

where

[8] To be precise, strong collusion-proofness of Laffont and Martimort requires the null side contract to be the unique equilibrium outcome, which is not the requirement of our SCP implementation. Nevertheless, unique implementation of the principal's *payoff* is required in both notions.

$$\rho_i := \frac{1}{n-1}\left[(1-\kappa_i)v(q(\tilde{\theta})) - \sum_{j\neq i} t_j(\tilde{\theta})\right] \text{ and } \sum_{i\in N}\kappa_i = 1.$$

It is straightforward to verify:

$$\mathbb{E}_{\tilde{\theta}_{-i}}[\hat{t}_i(\theta_i,\tilde{\theta}_{-i}) \mid \theta_i] = \mathbb{E}_{\tilde{\theta}_{-i}}[t_i(\theta_i.\tilde{\theta}_{-i}) \mid \theta_i], \ \ \forall \theta_i \in \Theta_i, \ \ \forall i \in N \tag{1}$$

$$v(\hat{q}(\theta)) - \sum_{i\in N}\hat{t}_i(\theta) = V, \ \ \forall \theta \in \Theta. \tag{2}$$

Condition (1) ensures that $\hat{M}$ is interim-payoff equivalent to M, and condition (2) guarantees the payoff of V to the principal for any manipulation satisfying (*RE*) and (*BB*), *provided that all agents participate with probability one*. CK also demonstrate the (generic) existence of an RCP mechanism $\hat{M}$ satisfying (1) and (2) for any $V \in \mathcal{V}$ when types are correlated for $n > 3$ agents or for $n = 3$ agents, one of whom has more than two possible types (see Lemmas 1 and 2 of CK). For our purpose, V *is RCP implementable* if (1) and (2) are met.[9]

We provide two results on the SCP implementation. For the independent private value case, we show that any $V \in \mathcal{V}$ is SCP implementable (Theorem 1). For other cases (e.g., allowing for correlated or interdependent types), we establish that any $V \in \mathcal{V}$ is VSCP implementable (Theorem 2).

3.1. Exact Implementation for the Independent Private Value Case

Assume that types are private and independently distributed. That is, for each $i \in N$, $\mu(\cdot \mid \theta_i) = \mu_i \in \Delta\Theta_i, \forall \theta_i$. The main idea of the SCP implementation is to augment the RCP mechanism $\hat{M}$, so that all agents participate with probability one in every equilibrium, and that they deliver a definite surplus of V to the principal whenever they participate. To this end, define first an (auxiliary) mechanism $M^i = (q^i, t^i) : \Theta \to Q \times \mathbb{R}^n$ for each $i \in N$ as follows:[10] if agent i reports θ_i, then the principal *randomly generates other agents' types* using the distribution μ_{-i}, where $\mu_{-i} = \times_{j\neq i}\mu_j$. With $\tilde{\theta}_{-i}$ thus generated,

$$q^i(\theta_i,\theta_{-i}) := \hat{q}(\theta_i,\tilde{\theta}_{-i}) \ \text{ and } \ t^i_j(\theta_i,\theta_{-i}) := \hat{t}_j(\theta_i,\tilde{\theta}_{-i}), \forall \theta \in \Theta, \forall j \in N. \tag{3}$$

Note that M^i does not depend on the reports of agents other than i. For agent

[9] Although the RCP implementation of CK is not defined with respect to the particular mechanism satisfying (1) and (2), their proofs use the mechanism. Hence, there is no difference with the current definition of the RCP implementation.

[10] This auxiliary mechanism is similar to the one adopted by Palfrey and Srivastava (1991) for unique Bayesian implementation, although there was no issue of *collusion* in that paper.

i, M^i replicates the same interim payoff as $\hat{M}$ or M. Hence, M^i is interim incentive compatible for agent i because the types of other agents are randomly generated using μ_{-i}. Thus, it is as if other agents are truth-telling. Furthermore, M^i ensures the payoff of V for the principal because for any reported type profile $\theta \in \Theta$ and randomly generated types $\tilde{\theta}_{-i}$,

$$v(q^i(\theta_i, \theta_i)) - \sum_{j \in N} t^i_j(\theta_i, \theta_{-i}) = v(\hat{q}(\theta_i, \tilde{\theta}_{-i})) - \sum_{j \in N} \hat{t}_j(\theta_i, \tilde{\theta}_{-i}) = V. \tag{4}$$

Contract $C_{\hat{M}} = (B, \xi, \tau)$ that we will use for SCP implementation is defined as follows. First, the message space for each agent $i \in N$ is $B_i = \Theta_i \times \mathbb{Z}_+$, where $\mathbb{Z}_+$ is the set of nonnegative integers. That is, upon participating, each agent is asked to announce his type and a nonnegative integer. Recall that the contract takes effect only when all agents participate. Second, the outcome function of $C_{\hat{M}}$ is defined as follows:

$$(\xi(\theta, \mathbf{z}),, \tau(\theta, \mathbf{z})) = \begin{cases} (\hat{q}(\theta), \hat{t}(\theta)) & \text{if } \mathbf{z} = 0 \\ (q^{i^*(\mathbf{z})}(\theta), t^{i^*(\mathbf{z})}(\theta)) & \text{if } \mathbf{z} \neq 0 \end{cases},$$

where $i^*(\mathbf{z})$ is a random selection from a set $\{j \mid z_j \geq z_{j'}, \forall j' \in N \text{ and } z_j \neq 0\}$ and $(q^{i^*(\mathbf{z})}(\theta), t^{i^*(\mathbf{z})}(\theta))$ is given as defined in (3). That is, the allocation and transfer rules follow $\hat{M}$ if no agent announces a positive integer; otherwise, $M^{i^*(\mathbf{z})}$ with $i^*(\mathbf{z})$ defined as an agent who announces the highest positive integer.

Note that the contract $C_{\hat{M}}$ augments two features to $\hat{M}$. First, when all other agents participate, this contract gives each agent i an option to trigger M^i (by announcing a high enough integer) and to realize the same interim payoff as he would under M. This feature, along with the $\varepsilon > 0$ payoff when others do not participate, gives an incentive to participate in every Bayesian Nash equilibrium. Second, whether M^i or $\hat{M}$ (or its manipulation) is implemented, the principal will be guaranteed to receive at least V as long as all agents participate. The two design features lead to the SCP implementation of each $V \in \mathcal{V}_0$.

Theorem 1. *Suppose that types are independently distributed and private. Then, every $V \in \mathcal{V}_0$ is SCP implementable if it is RCP implementable.*

Proof. We prove that $C_{\hat{M}}$ previously defined SCP implements V. The proof consists of the following three steps.

Step 1: *In every Bayesian Nash equilibrium occurring after $C_{\hat{M}}$ is offered, all agents accept $C_{\hat{M}}$ with probability 1.*

Fix any equilibrium following the offering of contract $C_{\hat{M}}$. By rejecting this contract, an agent i with type θ_i receives his reservation payoff of zero. Suppose he accepts the contract. If some other agent rejects the contract, then he will receive the payoff of $\varepsilon > 0$. If all other agents also accept the contract, then he must receive at least the expected payoff of $U_i^M(\theta_i)$ in equilibrium, or else, there is a profitable deviation of receiving (arbitrarily close to) that payoff by reporting (θ_i, n) for sufficiently large n (which will trigger M^i). Given that by assumption $U_i^M(\theta_i) > 0$ for μ-a.e. θ_i, the agent must be strictly better off by accepting the contract with probability one. Applying the argument to all agents, we conclude that all agents must accept $C_{\hat{M}}$ with probability one.

Step 2: *Whenever all agents accept $C_{\hat{M}}$, the principal receives the payoff of V (regardless of possible collusion of any kind).*

If all agents participate, then either $\hat{M}$ or M^i is played possibly with some collusive manipulations. In any case, the principal must receive V since both $\hat{M}$ and M^i ensure the payoff of V for the principal, as presented in (2) and (4). Thus, any manipulation satisfying (*RE*) and (*BB*) also gives V to the principal.

Step 3: *Contract $C_{\hat{M}}$ admits a (weak) perfect Bayesian equilibrium.*

Consider a strategy profile: *All agents participate. No collusive side contract is proposed in any coalition, and each agent i with type θ_i reports (θ_i, 0). If a side contract is proposed in a coalition, then all agents refuse the contract.*

We show that this strategy profile forms a weak perfect Bayesian equilibrium. If no deviation occurs from the suggested profile, then an agent i with type θ_i enjoys the (noncollusive) interim payoff of $U_i^M(\theta_i)$. Given that $\hat{M}$ is interim incentive compatible and interim individually rational, no agent can do strictly better by deviating from the suggested strategy as long as no collusion is proposed. In particular, given no collusion proposal is made in a coalition, no agent i of type θ_i in such a coalition has an incentive to trigger M^i. The reason is that it only gives him $U^M(\theta_i)$, which is the same as his equilibrium payoff. If a collusive side contract is proposed in a coalition, then each agent i with type θ_i turns it down and reports (θ_i, z_i) with $z_i > 2$. The latter action is supported by the off-the-equilibrium belief of agent i that whenever the collusive side contract is proposed, each agent $j \neq i$ will reject it and report $z_j > 0$, such that $\max_{j\neq i} z_j < z_i$. Reporting $z_i > \max_{j\neq i} z_j$ is agent i's optimal (continuation) strategy in this off-the-equilibrium path. The reason is that by reporting $z_i \leq \max_{j\neq i} z_j$, some mechanism M^k will be implemented and yield each agent j (including agent

i) a payoff no greater than $U_j(\theta_j)$.[11] That is, no type of any agent will be better off from a collusive deviation; the same is true for the third party maximizing the payoffs of agents in the coalition. Hence, there exists a weak perfect Bayesian equilibrium, in which no collusion is proposed and agents play $\hat{M}$. ▌

Remark 1. *Step 3 of the preceding theorem establishes a non-collusive equilibrium, one in which no agent proposes a collusive side contract. However, its true purpose is to show existence of "an" equilibrium, not necessarily of the noncollusive kind. Steps 1 and 2 demonstrate that the principal receives the desired level of payoff in "every" Bayes Nash equilibrium.*

An important feature of the SCP implementation is to ensure that each agent participates in the mechanism without fearing that they might suffer from collusive manipulations. By triggering an auxiliary mechanism M^i, agent i can guarantee payoff that is sufficiently high to justify his participation in the original mechanism. Once every agent participates, the "selling to the coalition" feature of RCP guarantees the principal the desired payoff level.

3.2. Virtual Implementation for the General Case

The proposed SCP mechanism relies on the independent private types for agents. Recall that auxiliary mechanism M^i constructs the allocation and transfers based on agent i's report of his type θ_i and random variables θ_{-i} that are generated according to the prior distribution of the other agents' types. This feature makes the proposed mechanism inapplicable for the general case, in which agents' types are correlated or interdependent. If the types are correlated, then one could still generate θ_{-i} using the right distribution of the other agents' types conditional on agent i's report. Given M^i, agent i could secure the noncollusive payoff, $U_i^{\hat{M}}(\theta_i)$, (by reporting his true type) but may be able to do better by misreporting.[12] This case will induce agent i to trigger M^i when he should not. If the types are interdependent, then mechanism M^i entails a different problem. When a collusive proposal is rejected, agent i's posterior belief on other agents' types may

---

[11] To see this, note that with M^k, agent k obtains $U_k^{\hat{M}}(\theta_k)=U_k^{M}(\theta_k)$, while each agent $j\neq k$ obtains as follows:

$$
\begin{aligned}
&\mathbb{E}_{\tilde{\theta}_k}[\mathbb{E}_{\tilde{\theta}_{-k}}[s_j(\hat{q}(\tilde{\theta}),\theta_j)+\hat{t}_j(\tilde{\theta})]\\
\leq\ &\mathbb{E}_{\tilde{\theta}_{-j}}[s_j(\hat{q}(\theta_j,\tilde{\theta}_{-j}),\theta_j)+(\hat{t}_j(\theta_j,\tilde{\theta}_{-j})]=U_j^{\hat{M}}(\theta_j)=U_j^{M}(\theta_j),
\end{aligned}
$$

where the inequality follows from the incentive compatibility of mechanism $\hat{M}=(\hat{q},\hat{t})$.

[12] For instance, mechanism $\hat{M}$ that achieves the full extraction of rents from the agents with correlated types can be made incentive compatible only by utilizing all agents' reports.

differ from his prior belief. Thus, his payoff from triggering M^i may not be the same as in $\hat{M}$ and could be lower, which may offer insufficient payoff protection from participation.

Hence, our approach for the general case will be slightly different. We still augment the RCP mechanism with an auxiliary mechanism that offers a strict incentive for each agent to participate. However, the incentives are created solely by transfers with a null allocation $\emptyset$. This approach imposes an arbitrarily small payoff loss for the principal. Hence, strong collusion proof implementation is obtained in the "virtual" sense.

Theorem 2. *$V \in \mathcal{V}$ is VSCP implementable if V is RCP implementable.*

Proof. Fix $\varepsilon > 0$ and consider a mechanism $\hat{M}_\varepsilon = (\hat{q}, \hat{t} + \varepsilon)$. Clearly, this mechanism RCP implements $V - \varepsilon$. Let us augment $\hat{M}_\varepsilon$ in the same way as mechanism $\hat{M} = (q^i, t^i)$ was augmented in Theorem 1, except that M^i is defined as follows: for all $\theta \in \Theta$,

$$q^i(\theta) := \emptyset \quad \text{and} \quad t^i_j(\theta) := \begin{cases} \varepsilon / 2 & \text{if } j = i \\ -\max\{\frac{V - v(\emptyset)}{n-1}, 0\} & \text{if } j \neq i \end{cases}.$$

Note that the principal's payoff from M^i being implemented is at least $V - \varepsilon / 2$ for the following reason:

$$\begin{aligned} & v(q^i(\theta)) - \sum_{j \in N} t^i_j(\theta) = v(\emptyset) - \varepsilon / 2 + \sum_{j \neq i} \max\left\{\frac{V - v(\emptyset)}{n-1}, 0\right\} \\ \geq \quad & v(\emptyset) - \varepsilon / 2 + (n-1)\frac{V - v(\emptyset)}{n-1} = V - \varepsilon / 2 . \end{aligned} \tag{5}$$

Let us call the augmented mechanism $C_{\hat{M}_\varepsilon}$ and prove that SCP implements $V - \varepsilon$.

Step 1: *In every Bayesian Nash equilibrium occurring after $C_{\hat{M}_\varepsilon}$ is offered, all agents accept $C_{\hat{M}_\varepsilon}$ with probability 1.*

The proof is analogous to Step 1 for the proof of Theorem 1. The only difference is that given $C_{\hat{M}_\varepsilon}$, each agent i must receive at least $\varepsilon / 2 > 0$ as his equilibrium payoff; otherwise, he can trigger M^i and receive (arbitrarily close to) $\varepsilon / 2$. That is, all agents must accept $C_{\hat{M}_\varepsilon}$ with probability one.

Step 2: *Whenever all agents accept $C_{\hat{M}_\varepsilon}$, the principal receives the payoff of $V - \varepsilon$ at* *least (regardless of possible collusion of any kind).*

If all agents participate, then either $\hat{M}_\varepsilon$ or M^i for some $i \in N$ is implemented. In either case, the principal receives $V-\varepsilon$ because $\hat{M}_\varepsilon$ RCP implements $V-\varepsilon$ and M^i yields at least $V-\varepsilon/2$ for the principal, as shown in (5).

Step 3: *Contract* $C_{\hat{M}_\varepsilon}$ *admits a (weak) perfect Bayesian equilibrium.*

Construct an equilibrium strategy profile in the same way as we did in Step 3 for the proof of Theorem 1. To prove that there is no profitable deviation, note first that no agent i of type θ_i has an incentive to trigger M^i because he only receives $\varepsilon/2$ by doing. By contrast, playing $\hat{M}_\varepsilon$ truthfully will enable him to receive $U_i^M(\theta_i)+\varepsilon \geq \varepsilon$. Moreover, no one has an incentive to propose a collusive side contract. The reason is that the side contract will be rejected by every agent, given the belief that collusion will trigger M^k for some member k. From M^k, agent k will receive $\varepsilon/2$, which is clearly less than the payoff from playing $\hat{M}_\varepsilon$ truthfully. In addition, agent $j \neq k$ in the same coalition will receive $s_j(q^k(\theta),\theta)-t_j^k(\theta) = -\max\{\frac{V-v(\emptyset)}{n-1},0\} \leq 0$, which is also less than the payoff from $\hat{M}_\varepsilon$. ▌

Using the results of CK, Theorem 2 implies the following result.

Corollary 1. *Any* $V \in \mathcal{V}$ *is VSCP implementable either if* μ *is uncorrelated across agents or if* μ *is a generic distribution on finite type space and there are* $n \geq 3$ *agents (one of whom must have three or more types in case* $n=3$ *).*

Proof. CK proves that the stated conditions are sufficient for any V to be RCP implementable. ▌

Remark 2. *The "generic" distribution in the Corollary refers to a pairwise identifiability condition (PI') defined in CK. The significance of the correlated type case is the full rent extraction result by Crémer and McLean. Theorem 2 suggests that the (almost) full extraction can be made SCP as long as there are more than two agents, even when only two of them are collusive. This case was left untreated in CK's RCP implementation.*

IV. Conclusion

The motivating example should be revisited. Suppose again that the buyer signs a contract with S1 and S2, in which the buyer pays 1/2 to S1 for delivery of the good, and the good is to be produced by S2 when his cost is less than 1/3 for a fee of 1/3 paid by S1 but is otherwise produced by S1 himself. Within this contract, our SCP contract will add an option that will allow S1 to force, with probability 1/2, S2 to deliver the good for a fee of 1/3 paid by S1.[13] This option secures S1 a zero payoff, even after signing on the principal's contract. Hence, it allows S1 to reject any demand by S2 of fee $P > 1/3$. Accordingly, both agents participate with probability one, and the seller can be assured of delivery of the good for the price of 1/2, just as would be required for the second-best outcome.

[13] The SCP contract will have a similar option for S2 to protect his reservation payoff. However, this situation is less of a concern in the context of the previous discussion.

References

Che, Y.-K. and J. Kim (2006), "Robustly Collusion-Proof Implementation," *Econometrica*, 74, 1063–1107.

__________________ (2007), "Optimal Collusion-Proof Auctions," *Journal of Economic Theory*, 144, 565–603.

Crémer, J. and R. P. McLean (1985), "Optimal Selling Strategies under Uncertainty for a Discriminating Monopolist when Demands are Interdependent," *Econometrica*, 53, 345–362.

_________________________ (1988), "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, 56, 1247–1257.

Laffont, J.-J. and D. Martimort (1997), "Collusion under Asymmetric Information," *Econometrica*, 65, 875–911.

___________________________ (2000), "Mechanism Design with Collusion and Correlation," Econometrica, 68, 309–342.

Palfrey, T. and S. Strivastava (1991), "Efficient Trading Mechanisms with Preplay Communication," *Journal of Economic Theory*, 55, 17–40.

Pavlov, G. (2008), "Auction Design in the Presence of Collusion," *Theoretical Economics*, 3, 383–429.

담합차단을 위한 메커니즘 디자인*

김 진 우**

초 록 Che and Kim(2006)은 대리인들이 담합을 시도하는 경우에도 메커니즘 디자이너로 하여금 손실을 입지 않게끔 하는 메커니즘을 설계하는 방법을 제시하였다. 이 메커니즘은 다양한 형태의 담합을 차단할 수 있으나, 담합에 가담한 대리인 들이 서로에 대해 가지는 믿음—특히 담합이 실패할 경우 서로가 구사할 전략에 대한 믿음—이 극단적으로 주어지는 경우에는 제대로 작동하지 않을 수 있다. 본 연구는 대리인들의 담합방식 및 서로에 대해 가지는 믿음에 상관없이 담합을 차단할 수 있는 메커니즘을 제시한다. 이를 가능케하는 메커니즘의 핵심적인 요소는, 각 대리인에게 다른 대리인들의 위협적인 담합전략에 대응하여 자신의 보수를 보장할 수 있게끔 하는 추가적인 전략을 부여하는 것이다.

핵심 주제어: 담합차단 메커니즘 디자인, 담합에 강건성을 가진 메커니즘
경제학문헌목록 주제분류: D47, D82

투고 일자: 2021. 10. 8. 심사 및 수정 일자: 2022. 6. 6. 게재 확정 일자: 2022. 7. 15.

* 이 논문은 2019년 대한민국교육부와 한국연구재단의 인문사회분야 중견연구자지원사업의 지원을 받아 수행된 연구이다(NRF-2019S1A5A2A01049533). 아울러 서울대학교 경제연구소 국가경쟁력센터로부터의 연구비지원에 대해 감사를 표한다.

** 서울대학교 경제학부 교수, e-mail: jikim72@snu.ac.