Private Enforcement against Collusion in Mechanism Design

Chen, Zhijun

School of Economics, Zhejiang University

May 2006

Private Enforcement against Collusion in Mechanism Design

Zhijun Chen[1]
School of Economics, Zhejiang University, China
E-mail: chenzj1219@gmail.com

Version: May 2006 First version: November 2005
(Presented in EEA Conference 2006)

This paper brings a new point of view into the theory of collusion-proof mechanism design, which highlights the principle of divide and conquer. We relax the restriction of publicly enforced grand contract in the framework of Laffont-Martimort-Itoh, which allows us to incorporate the approach of private enforcement into the theory. In a setting of moral hazard with mutually observable actions, we develop a multi-stage mechanism integrated with secret reporting and private transferring and show that the first-best allocation can be implememted in spite of collusion, which implies that preventing collusion entails no cost under new approach.

JEL Classification: C72, D82

Key Words: secret report, private contract, collusion-proof mechanism design

1. INTRODUCTION

The phenomenon of collusion has been an austere threat to organizations since the origin of human society. Fighting collusion thus has been a permanent challenge to the politicians and scholars in political science and sociology. However, collusive phenomena have been by and large ignored by economists, with a few exceptions of Chichago School's approach to regulatory capture in 1970s. Since 1980s, there has been a growing interest in studying collusive behavior, in various environments such as industrial organization, regulation and political economy.

Recent academic research on the theory of Collusion-Proof Mechanism Design has made much progress since the seminal paper Tirole (1986). The basic framework in the environment of multi-agent is developed by Laffont and Martimort (1997, 2000) under adverse selection setting and Itoh (1993) under moral hazard setting based on contract theory, which allows to integrate collusion as part of the general mechanism design under a hierarchy of organizations. As a shortcut of the methodology, it is assumed that a coalition is formed through a side contract, which is designed and enforced by a benevolent third party called side mediator.

The so-called "full-side-contracting" assumption plays a key role in collusion-proof mechanism design. It provides a simple and neat approach of modeling coalition formation and thus eliminates the problem of finding an extensive form for describing the collusive game between agents. The third party paradigm can be seen as a black box for the repeated interaction by which collusion emerges, which characterizes the upper bound that can be achieved by the coalition.

[1]We are grateful to Hideshi Itoh, David Martimort, Patrick Rey and Jean Tirole for their comments, as well as participants of my presentation in EEA conference 2006. Finacial support from China Scholarship Council is acknowledged.

The main problem of Laffont-Martimort-Itoh framework is that, in most cases, the optimal outcomes in the absence of collusion cannot be implemented in the presence of collusion. As shown in Laffont-Martimort (2000) under adverse selection setting, if the agents' types are correlated, preventing collusion entails additional incentive cost which brings further distortion of efficiency due to the trade-off between rent-extraction and efficiency. While under moral hazard with mutually observable actions, Itoh (1993) shows that there exists no collusion-proof grand mechanism which can implement the first-best allocations under collusion, whereas the first-best can be uniquely implemented in the absence of collusion as shown in Ma (1988).

The reason why preventing collusion is costly in Laffont-Martimort-Itoh framework is that the relative situation between the principal and collusive agents is not characterized adequately. In this framework, except for the full-side-contracting assumption, there are two restrictions in this framework: first, the communication between the principal and agents is public; second, all transfers in the grand contract are also public, in other words, the grand contract is enforced publicly. The side contract is enforced privately which makes the coalition a black box for the principal, whereas the grand contract is enforced publicly which makes the grand mechanism a transparent box to the coalition, that is, the relative situation between the principal and collusive agents is asymmetric. Under this framework, the coalition can be regarded as an integrated entity embedded into the grand mechanism, which implies that collusion is perfect and the highest collusive allocation can be achieved. Hence, the Laffont-Martimort-Itoh framework establishes a benchmark of the theory which characterizes the lowest bound that can be attained by a collusion-proof grand mechanism.

We argue that a coalition is not an integrated entity since there exists incongruence of interests among coalition members. Introducing a side mediator allows the coalition to settle down the conflict through side contracting, however, the upper bound that can be achieved by the coalition is determined by the grand mechanism because the side contract is embedded as a subgame into the grand mechanism. Under the publicly enforced grand contract, collusion is perfect and the highest bound can be achieved, whereas it can be made imperfect under privately enforcement. When the bilateral communication and transfer between the principal and each agent are enforced in a private way which are neither observable nor verifiable by other players, coalition formation turns to be a game with incomplete information which entails transaction costs. Collusion thus becomes imperfect and the highest collusive allocation might be unattainable. This highlights the principle of "divide and conquer" in fighting collusion.

This paper is motivated to find out an optimal collusion-proof mechanism which can implement the optimal no-collusion outcomes in the presence of collusion, by appealing to the approach of private enforcement of the grand contract. The basic idea of this approach is quite intuitive. When forming a coalition, agents must share common information about collective actions through communication. Collusion can be prevented by the principal if he can extract the collusive information from the coalition members. However, revealing collusive information through any public reporting is infeasible when the coalition is enforced by a side mediator, which implies that the only effective way of extracting collusive information is to induce secret report by the agents.

To breed secret reporting in the grand mechanism, the following necessary conditions must be met: first, there must be some channels for secret reporting; second, when an informant cannot bring hard evidence, an incentive compatible rewarding scheme must be designed to warrant his truth-telling; third, the identity of a whistle-blower must be hidden information, which implies that the rewards and fines for the whistle-blower must be enforced privately.

In a setting of moral hazard with mutually observable actions, we apply the approach of private enforcement to collusion-proof mechanism design. We propose a multi-stage indirect mechanism integrated with secret reporting and private transferring to extract collusive information. Under this mechanism, each agent is required to submit a secret report to the principal after a given effort pair has been taken. Each agent is allowed to win a bonus lottery if he reports the collusive actions, and his losing peer is levied a heavy fine. When both agents report the collusive actions, they must compete for the bonus which gives each agent only half chance of wining. Moreover, the bonus and fines are enforced through a private way which are neither observable nor verifiable by the side mediator. Since the identity of a whistle-blower is hidden information under private enforcement, each agent has incentives to defect and win the bonus under this mechanism. The collusive game thus turns to be a classical Prisoner's Dilemma in which defecting is a dominant strategy for both agents, hence the collusive outcome cannot be enforced in equilibrium. As a result, the first-best outcome can be implemented in spite of collusion.

The main contribution of this paper is of two aspects: first, in the setting of moral hazard with mutually observable actions, by incorporating the approach of private enforcement into a multi-stage indirect mechanism, it shows that the first-best outcome can be implemented in spite of collusion, which is in contrast with the impossibility result in Itoh (1993).

Second, this paper brings a new point of view into the theory of collusion-proof mechanism design, which highlights the principle of divide and conquer in fighting collusion. This principle is introduced into the theory of anti-collusion by Chen (2006a) at the first time. Under a simple model of tournament where the efficient effort levels are impossible to be implemented through simple mechanisms due to perfect collusion, Chen (2006a) shows that, by manipulating information under a sophisticated and indirect mechanism with a biased promotion rule, the principal can create asymmetric information among agents. This brings a trade-off between rent-extraction and efficiency into the coalition which reduces the efficiency of the coalition and shrinks the stakes of collusion. As a result, the efficient effort levels are possible to be implemented under the sophisticated mechanism. Moreover, Chen (2006a) shows that the indirect mechanism can perform better than direct mechanism in fighting collusion, which implies that restricting attention only to direct revelation mechanisms would be never optimal in collusion-proof mechanism design, a result also demonstrated in this paper.

There are other new approaches motivated to resolve the main problem in the Laffont-Martimort-Itoh framework. In a recent paper by Che and Kim (2005), it is shown that collusion can be prevented at no cost in a large class of circumstances under adverse selection settings with risk-neutral agents, including both uncorrelated types and correlated types. In that paper, the authors develop a general method for collusion-proof mechanism design, which utilizes the idea of "selling the firm to the agents". However, when there are only two agents with correlated types, preventing collusion entails strictly costs under their mechanism. Moreover, their approach is not robust with risk-averse agents. Their approach is quite different from our paper, and cannot be generalized to the case of moral hazard.

The approach of private enforcement are widely employed in fighting collusion. About 300 years ago, the Emperor Yongzheng in China Tsing Dynasty created a monitoring system to fight collusion and corruption. The most famous instrument was the so-called "sealed-box", which was granted by the Emperor to his favored ministers for conveying their secret reports. A sealed box had two keys: one for the reporter and another for the Emperor, and it can be delivered directly to the Emperor through a special and secret channel, whereas the regular reports can be sent only through bureaucratic channels. In a secret report, if the reporter can provide significant evidence on other's criminal activities, such as collusion and corruption, he can win the Emperor's trust as a reward; otherwise he can be punished for false reporting. Since enforcing of this system, it was recorded that, in 10 years, thousands of bureaucrats involved in criminal activities had been thrown into jails or executed, resulting a significant reduction of collusion and corruption at that time.

Nowadays, the policy of secret reporting is extensively used to fight organized crimes with a great success. Most notably, the practice of this policy in the case of Italy against Sicilian Mafia as well as US against drug-dealing and related crimes brings a wide debate on the role of snitch which causes a fundamental shift in the country's anti-drug laws – including federal mandatory minimum sentencing and conspiracy provisions[2].

To economists, perhaps, the most famous case of secret reporting is the Leniency Program in antitrust law enforcement. In 1993, US Antitrust Division of Department of Justice has redesigned its Corporate Leniency Policy which establishes that criminal sanctions can be avoided if a colluding firm reveals information either before an investigation is opened or the Division has not yet been able to prove collusion[3]. The Leniency Program that breeds whistle-blowing is widely regarded as a tremendous success. Since its introduction, an unprecedented number of cartels has been detected and successfully prosecuted, enormous fines (up to US$ 500 millions) has been levied against participants, and several top executives have served jail sentences in the US[4].

The rest is organized as follows. In section 2, the basic model is presented. We then turn to the first-best implementation in the absence of collusion in section 3 as a benchmark. In section 4, we review the methodology the theory and then turn to our main theme in section 5, which shows that first-best can be implemented in spite of collusion by appealing to the approach of private enforcement. In section 6, we extend this approach to the environment of unverifiable outputs, and show that nonverifiability of outputs has no bites to collusion-proofness. We finally summarize in section 7.

2. THE MODEL

We consider a three-person game with one principal and two agents. A risk-neutral principal has two tasks delegated to two identical and risk-averse agents, A_1 and A_2 , whose effort levels are mutually observable but cannot be observed by the principal. Each agent $n, n = 1, 2$ can choose an effort level $e_n \in E = \{0, 1, ..., N\}$ at a cost $C(e_n)$, which is strictly increasing and convex and satisfies $C(0) = 0$.

[2] A useful introduction is at http://www.pbs.org/wgbh/pages/frontline/shows/snitch/

[3] The European Union introduces a similar leniency program in 1996.

[4] See Motta and Polo (2003), Spagnolo (2003), Aubert, Rey and Kovacic (2005) and Chen and Rey (2006) for detailed analysis.

For a given effort pair $e = (e_1, e_2)$, each production process generates randomly one of finitely possible outputs $q^n \in \{q_1, ..., q_I\}, n = 1, 2$ with $q_1 < q_2 < ... < q_I$. For $i, j \in \{1, ..., I\}$, let $q_{ij} = (q_i, q_j)$ denote a pair of joint output, where the first variable stands for A_1's output and the second variable for A_2's output. Let $\pi_{ij}(e)$ denote the joint probability distribution of the output pair q_{ij} given an effort pair e. Assume that different effort pairs are distinguishable statistically, that is, the following condition holds:

Condition $(C.1)$ (pair-wise identifiability of efforts).

$\Pi(e^l) \neq \Pi(e^k), \forall e^l \neq e^k$, where $\Pi(e^l) = (\pi_{ij}(e^l))_{i,j}$.

An allocation $x = (t, e)$ is a set of payments and efforts, where $t = (t^1(q_{ij}), t^2(q_{ij}))$ is a vector of output-contingent transfers from the principal to the agents, and $e = (e_1, e_2)$ is a pair of effort levels.

For an allocation $x = (t, e)$, the principal's utility can be expressed as:

$U(e) = \sum_{i,j} \pi_{ij}(e)[S(q_{ij}) - t^1(q_{ij}) - t^2(q_{ij})]$,

and agent A_n's utility can be expressed by:

$V_n(e) = \sum_{i,j} \pi_{ij}(e)V(t^n(q_{ij})) - C(e_n)$,

where $S(.)$ is the principal's revenue of the outputs and $V(.)$ is agents' von-Neumann-Morgenstern utility function, which is continuous, strictly increasing and concave. For outputs $q_{ij} = (q_i^1, q_j^2)$, denoting by $v_{ij}^n = V(t^n(q_{ij}))$ agent n's von-Neumann-Morgenstern utility level, we can replace the principal's control variables with $t^n(q_{ij}) = h(v_{ij}^n)$, where $h = V^{-1}$.

Following Grossman and Hart (1983) and Mookherjee (1984), the principal's optimization program can be decomposed into two stages: At the first stage, given any effort pair, find the incentive scheme that minimizes expected incentive costs subject to agents' incentive compatibility and participation constraints. At the second stage, the principal chooses an optimal effort pair which maximizes his expected utility. In this paper, to fix idea, we focus only on the principal's program at the first stage, that is, the implementation problem.

3. BENCHMARK: FIRST-BEST ALLOCATION WITHOUT COLLUSION

When effort levels are mutually observable by each other, the first-best incentive contract can be implemented uniquely through a Moore-Repullo Mechanism in the absence of collusion, as shown in Ma (1988). Suppose that the principal wants to implement effort pair $e^* = (e_1^*, e_2^*)$. Since agents are risk-averse, optimal risk sharing rules then implies that agents' payments must be independent of the random outputs under perfect monitoring. Therefore, the first-best utility-payments are $C(e_1^*)$ and $C(e_2^*)$ respectively, that is $v_{ij}^n = C(e_n^*)$, for all i, j, $n = 1, 2$. To fix idea, we develop a multi-stage mechanism to implement the first-best, with slight difference from Ma's mechanism. The timing of the game is illustrated as follows:

[Timing of the game]

Stage 0: The principal proposes a grand mechanism G^0, which is approved if no agent vetoes.

Stage 1: Both agents take efforts simultaneously which are mutually observed.

Stage 2: Each agent submits a public report of effort pair $\widehat{e}^1$and $\widehat{e}^2$ respectively. If $\widehat{e}^1 = \widehat{e}^2 = e^*$, then the first-best allocation is enforced and each agent gets utility payment $C(e_1^*)$ and $C(e_2^*)$ respectively; otherwise, a win-loss game is triggered which entitles the winner a bonus lottery and levies the loser a fine in addition to the first-best payments:

Case (1): If $\widehat{e}^1 = e^*$ and $\widehat{e}^2 \neq e^*$, then agent A_2 wins a bonus lottery $X(\widehat{e}^2) = (x_{ij}(\widehat{e}^2))_{ij}$, which entitles him an output-contingent reward scheme $x_{ij}(\widehat{e}^2)$ based on his report, whereas agent A_1 is levied a fine $\overline{R}$ as the loser, and *vice versa.*

Case (2): If $\widehat{e}^1 \neq e^*$ and $\widehat{e}^2 \neq e^*$, then the winner and loser are determined by tossing an unbiased coin.

Stage 3: Outputs realize; the contract is enforced.

In this mechanism, $\overline{R} > C(e_1^*) + C(e_2^*)$, and lotteries $X(\widehat{e}^n)$ satisfy, for each $\widehat{e}^n, n = 1, 2,$

$$W(\widehat{e}^n | e^*) = \sum_{ij} x_{ij}(\widehat{e}^n)\pi_{ij}(e^*) < 0,$$

$$\text{and } \overline{R} > W(\widehat{e}^n | \widehat{e}^n) = \sum_{ij} x_{ij}(\widehat{e}^n)\pi_{ij}(\widehat{e}^n) > 0. \qquad (1)$$

Note that, condition $(C.1)$ implies the existence of lotteries $X(\widehat{e}^n)$ satisfying (1).

Under Condition (1), each agent has incentives to win the bonus lottery by reporting the truth if any unwanted effort pair has been taken, while no agent has incentives to trigger the win-loss game when the proposed effort pair has been taken. Hence, with mutual monitoring, each agent will take the proposed effort level and report the truth under this mechanism. As a result, the first-best allocation can be implemented through a unique Subgame Perfect Nash Equilibrium, as expressed by the following proposition

Proposition 1. *(Ma (1988)): Assume condition $(C.1)$ hold. If agents' effort levels are mutually observable, then the effort pair e^* can be subgame-perfect implemented through a multi-stage mechanism with first-best utility-payments to agents $C(e_1^*)$ and $C(e_2^*)$.*

Proof. We show that the proposed effort pair e^* can be enforced and the first-best allocation can be implemented uniquely under this mechanism.

Consider the following strategy for each agent: agent A_n takes effort e_n^* at Stage 1 and reports $\widehat{e}^n = e^*$ at Stage 2. Given that effort pair e^* has been taken by agents at Stage 1, then no agent has incentives to report other effort pair and trigger the win-loss game at Stage 2, since by (1) no agent can benefit from winning the lottery in this case. Therefore, reporting truth is optimal for both agents at Stage 2.

Now back to Stage 1, if agent A_1 takes e_1 instead of e_1^*, then agent A_2 has incentives to report $\widehat{e}^2 = (e_1, e_2^*)$ and wins the lottery, which makes agent A_1 strictly worse off. Hence, no one has incentives to deviate from σ^* unilaterally, which implies that σ^* constitutes a SPNE.

On the other hand, given that any other effort pair $e^d \neq e^*$ has been taken, then each agent has incentives to report $\widehat{e}^n = e^d$ and win the bonus lottery. The competitive reporting in the win-loss game makes each agent strictly worse off. Hence, effort pair e^d cannot be implemented in equilibrium. ∎

4. COLLUSION-PROOF MECHANISM DESIGN: ANOTHER POINT OF VIEW

4.1. Coalition Formation

Unfortunately, the grand mechanism G^0 is not robust to collusion. Since the payment schemes are output-independent, agents can benefit from reducing the effort levels and then cheating the principal collectively. To characterize collusive behavior, we assume that a coalition is formed followed by the grand contract and is enforced by a side contract:

Assumption 1. *(full side contracting) A coalition formed by the agents is sustained through a side contract which is designed and enforced by a benevolent side mediator.*

Assumption 1 provides a shortcut in methodology which allows us to deal with coalition formation under a neat and simple framework built on contract theory. Given the grand mechanism, under this assumption, we can characterize the problem of coalition formation by an optimization program of the side mediator, which maximizes the total welfare of the coalition subject to coalition incentive compatibility constraints and coalition participation constraints. According to Myerson (1982), Assumption 1 then ensures that the Revelation Principle holds in side contracting: without loss of generality, the side mediator can restrict his attention only to incentive compatible direct mechanisms.

Under mechanism G^0, the side mediator can choose an effort pair to maximize joint welfare of the coalition rather than the welfare of the principal, and this can be achieved by sending a pair of false messages to the principal. Suppose that a benevolent side mediator proposes a side contract to the agents as follows: Both agents take efforts $\hat{e} = (0, 0)$ and then report e^*. The stakes of collusion under this side contract are $C(e_1^*) + C(e_2^*)$, which can be reallocated inside the coalition without any transaction costs. It is obvious that this side contract can be enforced under grand mechanism G^0, which results in an inefficient effort pair $\hat{e} = (0, 0)$. Hence, under full-side-contracting, the first-best outcome cannot be implemented, as shown by Itoh (1993).

Proposition 2. *(Itoh (1993)) There exists no Collusion-Proof mechanism which implements the first-best allocation.*

Proof. See Itoh (1993). ∎

4.2. A Review of Methodology

The impossibility result of Proposition 2 appeals to rethink the methodology of the theory of collusion-proof mechanism design. Under the framework of Laffont-Martimort-Itoh, except for the basic assumption of full-side-contracting, there are two restrictions on methodology: first, all communication between the principal and agents is public; second, all transfers in the grand contract are also public, in other words, the grand contract is enforced publicly.

The main difficulty the theory of collusion-proof mechanism design confronted is that the Revelation Principle fails under collusion. A coalition has incentives to manipulate information revelation according to its own interests, which flunks the incentive compatibility of the grand contract. To prevent collusion, the principal must revise the original grand contract to make it incentive compatible for collectively truth-telling by the coalition. This sheds light on a so-called Principle of Collusion-Proofness: for any initial contract which is not collusion-proof, without loss of generality, the principal can restrict his attention to optimal collusion-proof grand contract which replicates the outcome of the coalition and thus no collusion arises in equilibrium. The Principle of Collusion-Proofness is indeed a version of Revelation Principle in the presence of collusion.

Under public enforcement of the grand contract, the communication between the principal and agents and reallocation of collusive stakes between agents can be fully controlled and manipulated by the side mediator. Hence, the strategic interaction between the principal and the coalition can be regarded as a game played by the principal and the side mediator. Appealing to the Principle of Collusion-Proofness, the principal can, without loss of generality, design an incentive compatible mechanism for collectively truth-telling, which leaves sufficient rents to the coalition. Therefore, by integrating collusion as a subgame into the grand mechanism and then modeling the grand mechanism as a game played by the principal and coalition, the Revelation Principle is reestablished. Hence, without loss of generality, the principal can restrict his attention only to direct revelation mechanism which is optimal and collusion-proof.

Under this framework, the trade-off between extraction of collusive rents and efficiency leads to further distortion of efficiency in the grand mechanism, which implies that the optimal outcomes in the absence of collusion might be infeasible under collusion. As shown in Laffont and Martimort (2000) under adverse selection setting, if the agents' types are correlated, the first-best outcomes implemented by Cremer-McLean Mechanism cannot be achieved under collusion and thus the collusion-proof grand mechanism entails additional incentive costs to the principal. Under the environment of moral hazard with mutually observable actions, Itoh (1993) shows that it is impossible to implement first-best allocation in the presence of collusion. Moreover, he shows that the principal can implement any effort pair with less costs under agents mutual monitoring and side contracting than no side contract, which implies that the Collusion-Proof Principle does not work under this environment.

The reason why preventing collusion is costly in Laffont-Martimort-Itoh framework is that the relative situation between the principal and collusive agents is not characterized adequately. When restricting to public enforcement, the principal is assigned the weakest power in grand contracting with agents and thus the relative situation between the principal and the coalition is asymmetric. The side contract is enforced privately which makes the coalition a black box for the principal; whereas the grand contract is enforced publicly which makes the grand mechanism a transparent box to the coalition. Due to this asymmetric relative situation, the coalition can be regarded as an integrated entity embedded into the grand mechanism, which implies that collusion is perfect and the highest collusive allocation can be achieved. If we regard the no-collusion environment as the upper benchmark of the theory in which the highest bound of allocation can be achieved by a grand mechanism, then the framework of Laffont-Martimort-Itoh provides a lower benchmark of the theory under perfect collusion, which characterizes the lowest bound that can be attained by a collusion-proof grand mechanism. Under the environment where the lowest bound does not coincide with the highest bound, preventing collusion entails strict costs and thus the optimal no-collusion outcomes are infeasible in the presence of collusion.

Hence, it would be not optimal to restrict attention only to publicly enforced grand contracts in fighting collusion. We argue that a coalition is not an integrated entity since there exists incongruence of interests among coalition members. Introducing a side mediator as a shortcut in methodology allows the coalition to settle down the conflict through side contracting, however, the upper bound that can be achieved by the coalition is determined by the grand mechanism because the side contract is embedded as a subgame into the grand mechanism. Under the publicly enforced grand contract, collusion is perfect and the highest bound can be achieved, whereas it can be made imperfect under privately enforcement of grand mechanism. When the bilateral communication and transfer between the principal and each agent are enforced in a private way which are neither observable nor verifiable by other players, coalition formation turns to be a game with incomplete information which entails transaction costs. Collusion thus becomes imperfect and the highest collusive allocation might be unattainable. This highlights the principle of "divide and conquer" in fighting collusion.

This paper is motivated to find out an optimal collusion-proof mechanism that can prevent collusion at no cost by appealing to the approach of private enforcement of the grand contract. We focus on the setting of moral hazard with mutually observable actions to fix ideas, since we are not ambitious to build a generalized framework for optimal collusion-proof mechanism design.

The basic idea behind the approach of private enforcement is quite intuitive. Agents must communicate to exchange information in order to coordinate their collusive actions when they form a coalition. On the other hand, the common information shared in the coalition can undermine the coalition if it is extracted by the principal. However, revealing collusive information through any public reporting is infeasible when the coalition is enforced by a side mediator, which implies that the only effective way of extracting collusive information is to induce secret report by the agents. To this end, the following necessary conditions must be met: first, there must be some channels for secret reporting; second, when an informant cannot bring hard evidence, an incentive compatible rewarding scheme must be designed to warrant his truth-telling; third, the identity of a whistle-blower must be hidden information, which implies that the reward and fine for the whistle-blower must be enforced through a private way.

To put these ideas into work, we propose a multi-stage mechanism integrated with secret reporting and private transferring to extract collusive information. Under this mechanism, after taking the proposed actions, each agent is required to submit a secret report of the effort pair, which provides a channel for whistle-blowing. If both agents report that the proposed effort pair has been taken, then the first-best payments are enforced. Otherwise, a win-loss game is triggered which allows each agent to win a bonus lottery by reporting the collusive actions. If only one agent reports the collusive effort pair, he wins the lottery automatically which increases his utility payment, whereas his losing peer is levied a heavy fine. If both agent report the collusive actions, then they must share the chance of winning and losing. Moreover, the bonus lotteries and fine are enforced privately which are neither observable nor verifiable by the side mediator. Since the identity of a whistle-blower is hidden information under private enforcement, each agent has incentives to defect and win the bonus lottery. This turns the collusive game to be a classical prisoner's dilemma where defecting is a dominant strategy for both agents, hence the collusive outcomes cannot be enforced in equilibrium. As a result, collusion can be deterred and the first-best can be implemented under this mechanism.

5. FIRST-BEST IMPLEMENTATION UNDER COLLUSION

5.1. Timing and Mechanisms

Suppose the principal wants to induce effort pair e^* through a grand mechanism G in the presence of collusion. The mechanism we propose here is a revised version of mechanism G^0 which allows us to incorporate secret reporting and private transferring of the lotteries and fine into the grand contract, with its timing illustrated as follows:

[Timing of the game]

Stage 0: The principal proposes a grand mechanism G, which is approved if no agent vetoes.

Stage 0.5: The side mediator proposes a side mechanism S; it is ratified if no agent vetoes.

Stage 1: Both agents take efforts simultaneously which are observed by the coalition.

Stage 2: Each agent submits a secret report of effort pair $\widehat{e}^1$and $\widehat{e}^2$ respectively. If $\widehat{e}^1 = \widehat{e}^2 = e^*$, then the first-best allocation is enforced and each agent gets utility payment $C(e_1^*)$ and $C(e_2^*)$ respectively; otherwise, a win-loss game is triggered which entitles the winner a bonus lottery and levies the loser a fine in addition to the first-best payments:

Case (1): If $\widehat{e}^1 = e^*$ and $\widehat{e}^2 \neq e^*$, then agent A_2 wins a bonus lottery $X(\widehat{e}^2) = (x_{ij}(\widehat{e}^2))_{ij}$, which entitles him an output-contingent reward scheme $x_{ij}(\widehat{e}^2)$ based on his report, whereas agent A_1 is levied a fine $\overline{R}$ as the loser, and *vice versa.*

Case (2): If $\widehat{e}^1 \neq e^*$ and $\widehat{e}^2 \neq e^*$, then the winner and loser are determined by tossing an unbiased coin.

Stage 3: Outputs realize; all contracts are enforced.

In this mechanism, the lotteries $X(\widehat{e}^n)$ and fine $\overline{R}$ are the same as in mechanism G^0 and satisfy condition (1).

Under this mechanism, allowing secret reporting at Stage 2 opens a back door for whistle-blowing when collusion emerges. Moreover, to conceal the identity of a whistle-blower, the communication and transfers in the win-loss game must be enforced privately which are neither observable nor verifiable by the side mediator.[5]

Assumption 2 *(Private Enforcement):* The communication between the principal and agents and the transfers of lotteries and fine are enforced privately.

Remark 1. Under this assumption, the total payments of the agents are decomposed into two parts: the first-best payments which are public, and the bonus and fine which are private[6]. One may argue that, under private enforcement, the problem of commitment by the principal will arise since the lottery and fine are transferred in a private way. In practice, this problem can be solved either by a reputation mechanism or by arbitration. In this one-shot game, we assume that the principal can employ an arbitrator to enforce the private transfers. If one insists that private enforcement will entail additional cost to the principal, in this mechanism, the expected arbitration cost is zero since the private transfers are enforced only in the path out of equilibrium which occurs with probability zero.

When designing a side contract, under assumption 1, without loss of generality, the side mediator can restrict his attention to direct side contracts. To fix ideas, we assume that the side mediator can monitor the agents' effort levels[7]. However,

[5] In the practice of law enforcement, all information about secret reporting, including the prize are kept in secret in order to protect the whistle-blower.

[6] One may argue that, while the agent who has received a bonus can keep secret, the agent who has received a fine won't remain silent. We can image that the bonus and fine are transfered through private accounts, which generates no hard evidence. If one agrees with the private enforcement of the side contract, he should also agree with the private enforcement of the grand contract.

[7] This assumption gives the side mediator more power in designing side contract, which improves the efficiency of the coalition. Later we will show that our approach is robust to collusion even under this assumption, hence this assumption has no bite to collusion-proofness under our mechanism.

the subgame at Stage 2 is a black box for the side mediator, which makes it impossible to design a side contract contingent on any information at Stage 2. Given the grand mechanism G, let $S = \{\phi(e), (s^1_{ij}(\phi, m))_{ij}, (s^2_{ij}(\phi, m))_{ij}\}$ denote the side mechanism, where

$\phi(e) = (\phi_1(e), \phi_2(e)) = e^c$ is a collective manipulation of effort pair;

$m = (m_1, m_2) \in E^2$ is a pair of messages sent by the agents about their secret reports;

$s^n_{ij}(\phi, m)$ is the output-contingent side payment that agent A_n receives, which also depends on the collective manipulation and messages.

The side mediator is not a source of money and therefore the coalition's budget is balanced: $h(s^1_{ij}(\phi, m)) + h(s^2_{ij}(\phi, m)) \leq 0$, for all i, j.

Under this side mechanism, agent A_n's expected utility is revised to

$V^c_n(e^c) = C(e^*_n) + \sum_{i,j} \pi_{ij}(e^c) s^n_{ij}(\phi, m)) - C(e^c_n)$.

5.2. The Main Result

Under the grand mechanism G and side mechanism S, we show that only null side contract that enforces the no-collusion outcome can arise in equilibrium, which implies that the first-best allocation can be implemented in spite of collusion.

Proposition 3. *Assume condition (C.1) hold. If agents' effort levels are mutually observable, then the effort pair e^* can be enforced and first-best allocations can be implemented through an optimal collusion-proof mechanism which has unique SPNE.*

Proof. Given the grand mechanism G, suppose the side mediator wants to induce a collusive effort pair $e^c \neq e^*$ through a side contract $S = \{\phi(e^*) = e^c, (s^1_{ij}(\phi, m))_{ij}, (s^2_{ij}(\phi, m))_{ij}\}$, supported by collusive strategy σ^c as follows:

"Agent A_n takes effort e^c_n at Stage 1; then reports $\widehat{e}^n = e^*$ to the principal and sends message $m_n = e^*$ to the side mediator at Stage 2. "

The proposition is shown through three steps.

Claim 1: *Under this mechanism, the message pair m is not informative and can be thus ignored.*

Note that, to implement the collusive strategy, the side mediator must punish the agent who reports $\widehat{e}^n \neq e^*$ and triggers a win-loss game. This implies that an agent who has reported $\widehat{e}^n = e^c$ has no incentives to tell the truth. He can always cheat the side mediator by sending a false message $m_n = e^*$, since the final outcomes of the subgame at Stage 3 are neither observable nor verifiable. Moreover, sending a false message does not change the probability distribution of production outputs, which implies that false reporting is not identifiable statistically in side contracting. Hence no incentive compatible side mechanisms can extract the true information on agents' secret reports when deviation from collusive strategy occurs.

Claim 1 implies that side payments $s^n_{ij}(\phi, m)$ can be reduced to $s^n_{ij}(\phi)$ without loss of generality.

Claim 2: *σ^c cannot constitute a SPNE and thus side contract S cannot be implemented in equilibrium.*

Given that effort pair e^c has been taken at Stage 1, consider agent A_1's optimal strategy at Stage 2. If he follows the collusive strategy, he can get

$V^c_1(\sigma^c) = C(e^*_1) + \sum_{i,j} \pi_{ij}(e^c) s^1_{ij}(\phi) - C(e^c_1)$, given that his peer follows the collusive strategy; and

$V^c_1(\sigma^c_1, \widehat{\sigma}^c_2) = V^c_1(\sigma^c) - \overline{R} < 0$, if his peer deviates by reporting $\widehat{e}^2 = e^c$.

Suppose now agent A_1 deviates by reporting $\widehat{e}^1 = e^c$. If agent A_2 follows the collusive strategy, then agent A_1 wins the lottery $X(e^c)$ which updates his expected utility payment

$V_1^c(\widehat{\sigma}_1^c, \sigma_2^c) = C(e_1^*) + \sum_{i,j} \pi_{ij}(e^c)x_{ij}(e^c) + \sum_{i,j} \pi_{ij}(e^c)s_{ij}^1(\phi) - C(e_1^c) = V_1^c(\sigma^c) + W(e^c|e^c) > V_1^c(\sigma^c)$.

If, however, agent A_2 also deviates from the collusive strategy, then agent A_1 will win the lottery with probability $\frac{1}{2}$, which revises his utility payment to

$$V_1^c(\widehat{\sigma}_1^c, \widehat{\sigma}_2^c) = C(e_1^*) + \sum_{i,j} \pi_{ij}(e^c)s_{ij}^n(\phi) - C(e_1^c) + \frac{1}{2}[\sum_{i,j} \pi_{ij}(e^c)x_{ij}(e^c) - \overline{R}]$$
$$= V_1^c(\sigma^c) + \frac{1}{2}[W(e^c|e^c) - \overline{R}] > V_1^c(\sigma_1^c, \widehat{\sigma}_2^c).$$

Therefore, reporting $\widehat{e}^n = e^*$ is dominated at Stage 2 and thus σ^c cannot be a NE in the subgame.

Claim 3: *only the null side contract* $S_0 = \{\phi(e^*) = e^*, (s_{ij}^1(\phi))_{ij} = 0, (s_{ij}^2(\phi))_{ij} = 0\}$ *can be implemented through a unique SPNE.*

Suppose the side mediator proposes the null side contract S_0 supported by the following strategy σ^*:

"Agent A_n takes effort e_n^* at Stage 1; then reports $\widehat{e}^n = e^*$ at Stage 2."

Given effort pair e^* taken by the agents, suppose that one agent, say agent A_1 deviates from strategy σ^* by reporting $\widehat{e}^n \neq e^*$, then he triggers the win-loss game. By condition (1), he cannot benefit from unilateral deviating even though he wins the lottery. Therefore, reporting $\widehat{e}^n = e^*$ is optimal for both agents at Stage 2. At Stage 1, no agent is allowed to deviate from e^* under the monitoring of side mediator. Hence, no one has incentives to deviate from σ^* unilaterally, which implies that σ^* constitutes a SPNE. The uniqueness of SPNE is obvious. Q.E.D ∎

Under this grand mechanism, any side contract that proposes an collusive effort pair $e^c \neq e^*$ cannot be sustained by any SPNE, because in the subgame of Stage 2, each agent has incentives to report e^c and trigger the win-loss game, which turns this subgame to be a classical Prisoner's Dilemma. Under assumption 2, for the side mediator, the subgame at Stage 2 likes a black box where the reports and outcomes are neither observable nor verifiable, hence it is impossible to design any incentive compatible side contract to extract true information. As a result, only the null side contract that induces the no-collusion effort pair e^* can be sustained by a Subgame Perfect Nash Equilibrium, which implements the first-best outcome.

Proposition 3 is in contrast with Itoh (1993), who shows that there exists no collusion-proof mechanism that implements the first-best incentive contract under full side contracting. As we argued, while restricting to publicly enforced grand contract, Itoh's impossibility result establishes a lower benchmark for the theory. Under private enforcement, the principal is assigned more power in grand contracting to the agents, which allows him to bring incomplete information into the coalition and undermine its efficiency. As a result, the first-best can be implemented through an optimal collusion-proof grand mechanism.

6. EXTENSION: UNVERIFIABLE OUTPUTS

The mechanism we developed above can be also applied to the environment of unverifiable outputs. Without loss of generality, we assume that agents' relative performance is verifiable[8].

[8] For example, there exists a kind of well-established technology to measure and verify the relative performance, which can be utlilized by a courte of law.

Assume that there are three verifiable states of natures, namely, "$\{q^1 > q^2\}$", "$\{q^1 < q^2\}$" and "$\{q^1 = q^2\}$", denoted by α, β and γ respectively. Let $Q = \{q_{ij}\}$ be the set of output pairs, then $\alpha = \{q_{ij}|q_i^1 > q_j^2\}$, $\beta = \{q_{ij}|q_i^1 < q_j^2\}$, $\gamma = \{q_{ij}|q_i^1 = q_j^2\}$. Notice that, the set of states denoted by $\Omega = \{\alpha, \beta, \gamma\}$ is a coarser partition of output space Q.

For any effort pair $e = (e_1, e_2)$, the probability distributions $P(e) = (p_\alpha(e), p_\beta(e), p_\gamma(e))$ can be expressed by:

$p_\alpha(e) = \Pr\{q^1 > q^2|e\} = \sum_{i>j} \pi_{ij}(e)$, $p_\beta(e) = \Pr\{q^1 < q^2|e\} = \sum_{i<j} \pi_{ij}(e)$, $p_\gamma(e) = \Pr\{q^1 = q^2|e\} = \sum_{i=j} \pi_{ij}(e)$.

Assume that different effort pairs are distinguishable statistically, that is, the following condition holds:

Condition $(C.2)$: $\forall e^l \neq e^k$, $P(e^l) \neq P(e^k)$.

Condition $(C.2)$ thus ensures that the optimal collusion-proof mechanism we employed under the environment of verifiable outputs can also be applied under unverifiable outputs. Therefore, nonverifiability has no bites to the optimal collusion-proof implementation, as concluded immediately:

PROPOSITION 4. *Assume Condition $(C.2)$ hold. Assume also that actions are mutually observable and outputs are not verifiable, then the effort pair e^* can be enforced and first-best incentive contract can be implemented through an optimal collusion-proof mechanism which has a unique SPNE.*

Proof. The proof is exactly the same as in the previous section and thus omitted. ∎

Remark 2. Notice that, condition $(C.2)$ implies condition $(C.1)$ and is thus stronger than the later. Nevertheless, it is not restrictive and can be satisfied generically.

The assumption of $|\Omega| = 3$ is necessary to meet condition $(C.2)$. Suppose not, there are only two verifiable states, namely, "win" or "loss". Let $\{w, l\}$ denote the states "agent A_1 wins" and "agent A_1 loses" respectively. Assuming a fair tie-broken rule, then the probability distribution can be expressed by:

$p_w(e) = \Pr\{q^1 > q^2|e\} + \frac{1}{2}\Pr\{q^1 = q^2|e\} = \sum_{i>j} \pi_{ij}(e) + \frac{1}{2} \sum_{i=j} \pi_{ij}(e)$;

$p_l(e) = \Pr\{q^1 < q^2|e\} + \frac{1}{2}\Pr\{q^1 = q^2|e\} = \sum_{i<j} \pi_{ij}(e) + \frac{1}{2} \sum_{i=j} \pi_{ij}(e)$.

The basic assumption of identical production technologies suggests that, for any symmetric effort pairs with $e_1 = e_2$, the induced matrix of probability distribution is also symmetric, that is, $\pi_{ij}(e) = \pi_{ji}(e)$, which implies $p_w(e) = p_l(e) = \frac{1}{2}$. In other words, when agents' production technologies are identical, exerting the same effort levels results in the same probabilities of win or loss, which equals to $\frac{1}{2}$ in two agents cases. In this case, condition $(C.2)$ is violated and the problem of non-identifiablility of symmetric effort pairs arises. As a result, the mechanism we developed is not sufficient to implement the first-best under nonverifiable outputs in the presence of collusion.

7. CONCLUSIONS

This paper brings a new point of view into the theory of collusion-proof mechanism design, which highlights the principle of divide and conquer in fighting collusion. Two restrictions in the framework of Laffont-Martimort-Itoh are relaxed, which allows us to incorporate the approach of private enforcement into the theory. In a setting of moral hazard with mutually observable actions, a multi-stage indirect mechanism integrated with secret reporting and private transferring is developed to show that the first-best allocation can be implemented in spite of collusion, which implies that preventing collusion entails no cost under new approaches.

This paper also appeals to rethink the methodology of the theory of collusion-proof mechanism design, in order to find out more powerful mechanisms and approaches for preventing collusion. As pointed out by Tirole (1986), in general, we should investigate what kind of allocation can be implemented by the principal when all types of bilateral coalitions are allowed, including coalitions between the principal and agents. When a bilateral coalition between the principal and agent can be enforced by a contract in a private way, collusion between agents then turns to be a game with incomplete information which brings transaction costs to the coalition. This highlights the principle of divide and conquer. However, there are difficulties in characterizing the contract equilibria when sequentially forming of coalitions is allowed, since in each round, a coalition formation can impose negative externalities on other players, which in turn induces another round of coalition formation. Hence, in this endeavor, a generalized theory on multi-contracting with externalities should be developed on which the new theory of collusion-proof mechanism design can be built, and this deserves further devotions by the theorists in the future.

8. REFERENCES

Aubert, C, P. Rey and Kovacic, W (2005) "The Impact of Leniency and Whistle-blowing Program on Cartels" *International Journal of Industrial Organization,* forthcoming

Che, Y and Kim, J (2005) "Robustly Collusion-Proof Implementation" *Forthcoming in Econometrica.*

Chen, Z (2006a) “Divide and Conquer” *working paper* in GREMAQ

Chen, Z (2006b) "Secret Report and Optimal Collusion-Proof Mechanism Design: the Case of Moral Hazard" *working paper* in GREMAQ

Chen, Z and P. Rey (2006) "Optimal Design of Leniency Mechanism" *working paper* in GREMAQ

Grossman, S and Hart, O (1983) "An Analysis of the Principal-Agent Problem" *Econometrica* 51: 7-45

Itoh, H (1993) “ Coalitions, Incentives and Risk Sharing” *Journal of Economic Theory* 60: 410-427

Laffont, J-J and Martimort, D (1997) “Collusion under Asymmetric Information” *Econometrica*, 65: 875-911

Laffont, J-J and Martimort, D (2000) “Mechanism Design with Collusion and Correlation” *Econometrica*, 68: 309-342

Ma, C (1988) "Unique Implementation of Incentive Contracts with Many Agents" *Review of Economic Studies*, 55: 555-572

Mookherjee, D (1984) "Optimal Incentive Schemes with Many Agents" *Review of Economic Studies,* 51:433-446

Moore, J and Repullo, R (1988) " Subgame Perfect Implementation" *Econometrica*, 46: 1191-1220

Motta, M and M. Polo (2003) "Leniency Programs and Cartel Prosecution" *International Journal of Industrial Organization* 21: 347-379

Myerson, R (1982) "Optimal Coordination Mechanism in Generalized Principal-Agent Problems" *Journal of Mathematical Economics* 10: 67-81

Spagnolo, G (2003) "Optimal Deterrence Mechanisms Against Cartels and Organized Crime" *Working Paper*

Tirole, J (1986) “Hierarchies and Bureaucracies: On the Role of Collusion in Organizations” *Journal of Law, Economics and Organization*, 2: 181-214

Tirole, J (1992) “Collusion and Theory of Organizations” In *Advances in Economic Theory*, J-J Laffont (eds.) Sixth World Congress, Cambridge University Press, 71-155