

## Defending Multiple Terrorist Targets

Sanghoon K. Lee

(Assistant Professor, College of International Studies, Hallym University)

### 대테러전략의 게임이론적 분석

이 상 훈

(한림대학교 국제학부 조교수)

\* 이상훈: (e-mail) slee@hallym.ac.kr, (address) College of International Studies, Hallym University, 1, Hallymdaehak-gil, Chuncheon, Gangwon-do, 200-702, Korea.

- Key Word: **Terrorism**(테러리즘), **Counterterrorism**(대테러), **Contract**(계약)
- JEL Code: D74, H56
- Received: 2013. 4. 30      • Referee Process Started: 2013. 5. 9
- Referee Reports Completed: 2013. 6. 26

*KDI Journal of Economic Policy*, vol. 35, no. 3, 2013

© Korea Development Institute 2013

## ABSTRACT

This paper analyzes a situation where multiple targets are exposed to a potential terrorist attack. The probability of an attack is determined endogenously in a game where a terrorist chooses the target while the targets independently determine the level of deterrence. As each target tries to divert an attack to others, the symmetric equilibrium becomes suboptimal and exhibits dispersion in the level of deterrence. The analysis shows that the first best deterrence level may be achieved when the targets can write a binding risk-sharing contract. Such a contract has limited applicability however as it suffers from a potential verification problem.

-----

본 논문은 테러리스트와 복수의 테러 대상자 사이의 전략적 행동을 게임이론을 통해 분석한다. 기본모형은 테러리스트가 공격 여부와 그 대상을 선택하고 테러 대상자들은 각자 방어 수위를 결정하는 게임으로서 테러의 발생확률이 내생적으로 결정되는 구조를 가지고 있다. 균형에서 테러 대상자들은 테러 공격을 서로에게 전가시키려 테러 방어에 과도한 투자를 하게 되며, 동일한 테러 대상자들도 서로 다른 수준의 방어 수위를 유지하게 된다. 균형의 결과는 테러 대상자들의 총합적 후생을 극대화하는 최적 선택에 비해 비효율적임을 보일 수 있는데, 이러한 문제는 원칙적으로 테러 대상자들이 상호 간에 위험을 분담하는 계약을 맺음으로써 해결할 수 있다. 하지만 그러한 계약들도 계약조건의 검증이 쉽지 않아 그 실효성에 제약이 있게 된다.

## I. Introduction

A series of recent events has made terrorism a subject of increased academic interest. The dominant theme in the literature has been the analysis of various strategic interactions among the players in the scene. This article aims to provide an analytical framework that incorporates two of such interactions: the one between a terrorist and the targets and the one among the targets themselves.

The model considers a situation in which many identical targets are exposed to a potential terrorist threat. The game starts when a terrorist observes the cost of operation. Comparing its cost and the benefit, the terrorist decides whether to initiate an operation or not. Each target invests in deterrence without knowing the terrorist's decision. An operation is completed when the terrorist launches an attack against one of the targets. Two questions emerge naturally: i) is the outcome of the game desirable from the targets' perspective?, and ii) if not, is there any way in which the targets can improve their fate?

The answer to the first question turns out to be negative. The targets' problem comes from two different sources. First, the targets do not internalize the negative externalities when they make their investment decisions. In equilibrium, each target tries to outinvest others knowing that the weakest target will be the victim of an attack. The possibility of this attack-diversion has been recognized in the literature (Sandler, 2003). When every target tries to undercut others, however, no target can actually succeed in the attempt. Consequently, they all end up playing a mixed strategy in the symmetric equilibrium of the game. Moreover, an attack occurs with a lower probability compared to the hypothetical case where the targets coordinate their actions. In other words, there is too much deterrence in equilibrium.

The second problem facing the targets results from the fact that they cannot credibly commit to their actions. The idea, which dates back to Schelling (1960), is certainly not new. Given that the terrorist's decision depends on his belief about the targets' choice, the targets would do better if they could convince the terrorist that an attack is not likely to succeed. But this requires that the targets maintain an excessive level of investment even when the risk of an attack has been kept low. Such a strategy, however, will not be credible and hence cannot succeed in equilibrium.

The targets can mitigate or sometimes even completely resolve the problem by writing a binding contract among themselves. Assuming that neither the targets' investment nor the level of deterrence is verifiable, the only feasible contracts have

the form of a risk-sharing scheme. When the victim's identity can be verified regardless of the outcome of an attack, such a contract can restore the efficiency as long as the first best does not require a complete deterrence. If the identity of a failed attack cannot be verified, however, the targets may achieve the coordination outcome but not the first best. It is argued that this verification problem becomes more severe when the targets involved are sovereign states rather than private parties.

There is a growing literature on game-theoretic analysis of terrorism (Arce and Sandler, 2005; Heal and Kunreuther, 2005; Keohane and Zeckhauser, 2003; Konrad, 2004; Rosendorff and Sandler, 2004; Siqueira, 2005). For a survey of earlier studies, see Sandler and Enders (2004). Most of existing literature focus on a single aspect of the strategic interactions such as the one among the targets, among the terrorist groups, or the one between a target and a terrorist group. A notable exception is Bier *et al.* (2007), which considers a model with a terrorist and two potential targets. In their model, the terrorist is assumed to pick a specific target at the same time when he makes the operational decision. This difference in modeling strategy is discussed further in the next section. By comparing the outcomes of a simultaneous and sequential move games, they show that the targets can enjoy a strategic advantage by moving first. But neither the commitment issue nor its potential remedies is explicitly examined in their article.<sup>1</sup>

The idea that the government may correct the attack-diverting externalities by subsidizing terrorism insurance is explored in Lakdawalla and Zanjani (2005). But the strategic interaction between the terrorist and the targets is not analyzed explicitly in their model. Consequently, a budget-balancing subsidy scheme is shown to implement the first best, which is not the case in this article where the targets' choice is constrained by the commitment problem. The techniques used to characterize the equilibrium distribution function are similar to those in Varian (1980). Due to the differences in the context, however, there are significant changes in the details including the construction of the support of the distribution function.<sup>2</sup>

The rest of the article is organized as follows. Section 2 introduces the basic

---

**1** One of the games considered in Bier *et al.* (2007) has the same structure as the second- and third stage of the game analyzed in this article. In their set-up, the terrorist's preference is assumed to be a private information, which opens up the possibility that equilibrium may exist in pure strategies. They claim indeed that equilibrium must be pure, and characterize such an equilibrium. But their proof, which resorts to the convexity of the targets' cost function, is incomplete at best. This is because the convexity cannot rule out a mixed equilibrium with a connected support although it does rule out an equilibrium with a "gap" in the support.

**2** More specifically, the fixed numbers of "informed" and "uninformed" consumers in Varian (1980) provide natural boundaries for the equilibrium support, which is not the case in this paper where the boundaries need to be determined endogenously from the targets' best response.

model. The symmetric equilibrium of the game is characterized in section 3. The welfare property of the equilibrium as well as the feasibility of contracting solution are analyzed in section 4. Concluding remarks follow in section 5.

## II. Model

There is a terrorist group (= terrorist) and  $n \geq 2$  potential terrorist targets (= targets). The game proceeds as follows. In the first stage, the terrorist draws a type  $\theta \in [\underline{\theta}, \bar{\theta}]$  from a known distribution. The type determines the cost of carrying out a terrorist operation. After observing the type, the terrorist decides whether to start an operation or not. Neither the terrorist's type nor his action in the first stage is observed by the targets. In the second stage, each target makes investment to enhance its capability to deter a future attack. In the third stage, the terrorist chooses a specific target and launches an attack.

A target incurs a loss of  $D > 0$  if an operation succeeds and zero if it fails. An attack against target  $i$  becomes a success with probability  $p_i$ . Failure may occur either during the preparation stage of an operation or at the time of an attack. The first case arises because an operation may be detected before its completion. The second case includes a situation in which an attack is defeated by the target as well as the one in which failure occurs from a natural cause (e.g., a bomb does not go off due to malfunction).

By investing in deterrence efforts, each target can reduce its success probability  $p_i$ . The cost of investment is given by a twice continuously differentiable function  $c(p_i)$ , which is strictly decreasing and convex. Three additional assumptions are made: i)  $\lim_{p \rightarrow 0} c'(p) + D < 0$ , ii)  $D < \lim_{p \rightarrow 0} c(p)$ , and iii) there exists a  $p_{\max} \in (0, 1]$  such that  $c(p) = 0$  for  $p \in [p_{\max}, 1]$  and  $c'(p_{\max}) = 0$ . The first (second) condition is satisfied when the marginal (total) cost gets sufficiently large near  $p = 0$ . The third condition requires that both total and marginal cost become zero as  $p$  approaches 1.

The terrorist maximizes the targets' expected loss net of his cost of operation  $\theta$ . The terrorist's type  $\theta \in [\underline{\theta}, \bar{\theta}]$  is his private information but its distribution  $G(\theta)$  is known to the targets.  $G(\theta)$  is strictly increasing and twice continuously differentiable with  $2G'(\theta) + G''(\theta)\theta > 0$ . The condition, which guarantees that the first-best solution is well-defined, is satisfied if the distribution function is not too concave. Two additional assumptions are made: i)  $p_{\min} \equiv \frac{\underline{\theta}}{D} < p_{\max}$  and ii)

$p_{\max} < \frac{\bar{\theta}}{D}$ . The first (second) condition implies that an operation must be worthwhile (not worthwhile) for the terrorist when the cost is sufficiently small (large). It rules out rather trivial cases where either an attack never occurs or it occurs always in equilibrium.

In the first stage, the terrorist decides whether to start an operation ( $s = s_1$ ) or not ( $s = s_0$ ). Given that an operation has been initiated, the terrorist picks a specific target for attack in the third stage. In order to maximize the targets' expected loss, the terrorist should attack the most vulnerable target. In other words, the terrorist will choose a target  $k$  if and only if

$$p_k = \max p_j.$$

When more than one targets have the highest probability of success, the terrorist is assumed to pick one of the targets randomly.

In any Perfect Equilibrium of the game, the terrorist's optimal decision in the third stage must be taken into account by the targets. Given that an attack occurs only when there is an ongoing operation, a target's payoff depends on whether the terrorist initiated an operation or not. Let  $L_i$  be target  $i$ 's total loss, i.e., the sum of the cost of investment and the expected loss from an attack. Then  $L_i$  is given by

$$L_i(\mathbf{p}, s) = c(p_i) + I(s)\rho_i(\mathbf{p})p_iD$$

where

$$\mathbf{p} = (p_1, \dots, p_n)$$

$$I(s) = \begin{cases} 1 & \text{if } s = s_1 \\ 0 & \text{if } s = s_0 \end{cases}$$

$$\rho_i(\mathbf{p}) = \begin{cases} 1 & \text{if } p_i > \max_{j \neq i} p_j \\ \frac{1}{m} & \text{if } p_i = \max_{j \neq i} p_j, m = \text{number of tied maximum} \\ 0 & \text{if } p_i < \max_{j \neq i} p_j \end{cases}$$

By setting the probability of success at  $p_i$ , target  $i$  incurs a cost of  $c(p_i)$ . The second term in the expression,  $I(s)\rho_i(p_1, \dots, p_n)p_iD$ , follows from the fact that loss from an attack materializes only when i) there is an ongoing operation ( $s = s_1$ ), ii) target  $i$  is the weakest among  $n$  potential targets, and iii) the attack turns out to be successful. The joint loss of the targets is obtained by adding up the individual target's loss:

$$L(\mathbf{p}, s) = \sum_i L_i(\mathbf{p}, s) = \sum_i c(p_i) + I(s) \max\{p_1, \dots, p_n\} D.$$

Targets are allowed to play mixed strategies. The strategy of target  $i$  then becomes a distribution function  $F_i(p)$  with a support in  $[0,1]$ .

In the first stage, the terrorist makes his operational decision. For a given cost of operation  $\theta$ , the terrorist should initiate an operation if

$$E^{\mathbf{p}}[L(\mathbf{p}, s_1)] - \theta \geq E^{\mathbf{p}}[L(\mathbf{p}, s_0)]$$

or

$$\theta \leq E^{\mathbf{p}}[L(\mathbf{p}, s_1)] - E^{\mathbf{p}}[L(\mathbf{p}, s_0)] = E^{\mathbf{p}}[\max\{p_1, \dots, p_n\} D]$$

where the expectation is taken with respect to  $\mathbf{p} = (p_1, \dots, p_n)$ . The optimal decision follows a cutoff rule: an operation is initiated if and only if the cost of operation does not exceed the targets' expected loss from an attack.

Since the targets do not observe the terrorist's choice in the first stage, there is asymmetric information between the terrorist and the targets. The outcome of the game hence depends on how the targets form their beliefs about the state of the game in the second stage. The relevant equilibrium concept is weak Perfect Bayesian Equilibrium. It requires two conditions: *sequential rationality* and *belief consistency*. The first is satisfied if each player makes an optimal decision in every information set. The second condition requires that beliefs should be consistent with equilibrium strategies. Let  $\mu$  be the targets' belief that there is an ongoing operation, i.e.,  $s = s_1$ . The equilibrium condition is then summarized as follows.

**Definition** *The weak Perfect Bayesian Equilibrium of the game is defined as follows:*

- i)  $\mu^* = \Pr(\theta \leq \theta^*)$
- ii)  $s = s_1$  if and only if  $\theta \leq \theta^*$
- iii)  $E^s \left[ \int_{[0,1]^n} L_i(\mathbf{p}, s) d(F_i^*(p_i) F_{-i}^*(\mathbf{p}_{-i})) \right] \leq E^s \left[ \int_{[0,1]^n} L_i(\mathbf{p}, s) d(F_i(p_i) F_{-i}^*(\mathbf{p}_{-i})) \right]$   
for all  $i$  and  $F_i(\cdot)$

where  $\theta^* \equiv \int_{[0,1]^n} \max\{p_1, \dots, p_n\} D dF^*(\mathbf{p})$ ,  $F^*(\mathbf{p}) \equiv F_1^*(p_1) \times \dots \times F_n^*(p_n)$ , and

$$F_{-i}^*(\mathbf{p}_{-i}) \equiv F_1^*(p_1) \times \dots \times F_{i-1}^*(p_{i-1}) F_{i+1}^*(p_{i+1}) \times \dots \times F_n^*(p_n).$$

The first condition guarantees that the targets' belief is consistent with the terrorist's equilibrium strategy in the first stage. In equilibrium, the terrorist uses a cutoff rule to make his first-stage decision. For consistency, therefore, the targets'

belief ( $= \mu^*$ ) must coincide with the actual probability of attack, which occurs when the terrorist's cost is lower than the equilibrium cutoff ( $= \theta^*$ ). The second condition shows the optimality of the terrorist's choice while the third condition gives that of the targets. The optimal cutoff is found where the terrorist becomes indifferent between the two choices  $s_0$  and  $s_1$ . Each target should minimize its expected loss by choosing a distribution over  $p$ . Notice that the terrorist's optimal decision in the third stage is subsumed in the second-stage game and hence omitted in the equilibrium definition.

Before closing this section, it seems appropriate to add a brief remark on the modeling strategy. A distinguishing feature of the game examined in this article is that the terrorist makes his decision in two steps. In the first stage, the terrorist decides whether to participate in terrorist activities. Participation requires a sunk cost, which can be interpreted as the opportunity cost of doing the terrorism business. The decision in the third stage, on the other hand, is assumed to be tactical in nature which can be made flexible depending on the observed strength of the targets.

The advantage of this approach is that it allows one to capture the two sources of inefficiency—attack-diverting externalities and commitment problem—at the same time. One can easily confirm this by examining the consequences of adopting alternative strategies. Consider first the case in which the third stage is merged into the first stage. This implies that the terrorist picks a target without knowing each target's strength. Given that its strength cannot be observed by the terrorist, there will be no incentive for a target to out-invest others hence attack-diversion effect will disappear. If the first stage is merged into the third stage instead, the targets become the first-mover in the game and hence have the commitment power. By construction, the commitment problem does not arise in this case.

### III. Equilibrium

The first step for characterizing the equilibrium is to show that an attack occurs with a positive probability, i.e.,  $\mu^* > 0$ . This can be easily shown by examining the implication of assuming the contrary. If the targets believe that an attack never occurs, there is no point of investing in deterrence. Therefore, only the success probabilities higher than  $p_{\max}$  must be played in equilibrium. But this implies

$$\theta^* = \int_{[0,1]^n} \max\{p_1, \dots, p_n\} DdF^*(\mathbf{p}) \geq p_{\max} D.$$

Given the assumption  $p_{\max}D > \underline{\theta}$ , however, the terrorist should attack with a positive probability because  $\Pr(\theta \leq \theta^*) \geq \Pr(\theta \leq p_{\max}D) > 0$ . This shows that the belief  $\mu^* = 0$  cannot be consistent with the terrorist's strategy because  $\Pr(\theta \leq \theta^*) = \mu^*$  in equilibrium.

One of the distinctive features of the model is that a target can divert a potential attack to others. Consequently, each target has an incentives to slightly 'undercut' other targets' probabilities. With every target trying to out-invest others, one may expect that an equilibrium in pure strategies would be difficult to sustain. It turns out that this is the case indeed.

**Lemma 1** *There is no equilibrium in which the targets use pure strategies.*

Therefore an equilibrium, if exists, must be in mixed strategies. In a mixed-strategy equilibrium, each target uses a distribution function to randomize over success probabilities. Although the model assumes that targets are identical, one cannot exclude in general the possibility that they use different distribution functions in equilibrium. To keep the analysis tractable, however, the following discussion will focus only on symmetric equilibria of the game where the targets use the same distribution function. Let  $F^*(p)$  be the equilibrium distribution function.

For an equilibrium to be sequentially rational, the game needs to be solved backwards starting from the last stage. Once the outcome of the third stage is already embedded in the targets' payoff, the second stage becomes effectively the last stage of the game. The following lemma gives the first result regarding the targets' equilibrium choice.

**Lemma 2** *There is no jump in the equilibrium distribution function.*

Similar to the case with pure strategies, the intuition is again that a jump in a distribution function will provide an opportunity to undercut others' probabilities. Without a jump, the distribution function becomes continuous and the probability of a tie becomes negligible. This makes it possible to express a target's equilibrium loss in a much simpler form. For a given belief  $\mu > 0$ , it follows that

$$\begin{aligned} E^S \left[ \int_{[0,1]^n} L_i(\mathbf{p}, s) dF^*(\mathbf{p}) \right] \\ &= \int_{[0,1]^n} \{c(p_i) + \mu \rho_i(p_1, \dots, p_n) p_i D\} dF^*(\mathbf{p}) \\ &= \int_{[0,1]} \{c(p_i) + \mu F^*(p_i)^{n-1} p_i D\} dF^*(p_i). \end{aligned}$$

The first equality uses the definition  $E^s[I(s)] = \mu$ . The second equality follows from the fact that target  $i$  becomes the weakest among all targets with probability  $F^*(p_i)^{n-1}$ .

Let  $L_{\max}(p) \equiv c(p) + \mu p D$ .  $L_{\max}(p)$  can be interpreted as the maximum possible loss of a target when it sets a success probability  $p$  with certainty. It is the sum of the investment cost  $c(p)$  and  $\mu p D$ , which equals the expected loss from an attack when a target expects to be the weakest for sure. Notice that  $L_{\max}(p)$  is strictly convex given the assumption  $c''(\cdot) > 0$ . Associated with this function are the following two quantities  $\bar{p}$  and  $\underline{p}$ , which play a crucial role in the analysis:

$$\bar{p} \equiv \operatorname{argmin} L_{\max}(p)$$

$$c(\underline{p}) \equiv c(\bar{p}) + \mu \bar{p} D.$$

**Lemma 3** *For a given  $\mu$ ,  $\bar{p}$  and  $\underline{p}$  are uniquely determined. Both  $\bar{p}(\mu)$  and  $\underline{p}(\mu)$  are continuously differentiable with respect to  $\mu$ .*

It turns out that the interval formed by the two values  $[\underline{p}, \bar{p}]$  becomes the support of the equilibrium distribution. A formal proof, which is given in the appendix, requires several steps of reasoning. The main intuition, however, comes from the basic property of a mixed-strategy equilibrium: the targets must be indifferent among all pure strategies played with a positive probability. In particular, this implies that a target should be indifferent between playing only the minimum and the maximum  $p$  in the support of the equilibrium distribution. Let  $p_m$  ( $p_M$ ) be the minimum (maximum) respectively. Since a target becomes the strongest (weakest) by setting  $p = p_m$  ( $p = p_M$ ), it follows that  $c(p_m) = c(p_M) + \mu p_M D$ .

Assuming that there is no ‘gap’ between  $p_m$  and  $p_M$ , the support  $[\underline{p}, \bar{p}]$  can be verified by showing that  $p_m = \underline{p}$  and  $p_M = \bar{p}$ . Given that  $L_{\max}(p)$  is strictly convex and reaches its minimum at  $\bar{p}$ , it is strictly decreasing (increasing) for  $p < \bar{p}$  ( $p > \bar{p}$ ). If the maximum  $p_M$  does not coincide with  $\bar{p}$ , then it must be that  $c(p_M) + \mu p_M D > c(\bar{p}) + \mu \bar{p} D$ . If a target sets  $p = \bar{p}$ , however, the expected loss will be at most  $c(\bar{p}) + \mu \bar{p} D$  contradicting the assumption that  $p = p_M$  is played in equilibrium. A similar argument can establish the minimum  $\underline{p}$ . Suppose to the contrary that the minimum  $p_m$  is not equal to  $\underline{p}$ . By construction, the expected loss at the maximum  $p_M = \bar{p}$  is equal to the cost of investment at  $\underline{p}$ , i.e.,  $c(\underline{p}) = c(\bar{p}) + \mu \bar{p} D$ . Given that  $c(\cdot)$  is strictly decreasing, however, this implies

$c(p_m) \neq c(\bar{p}) + \mu\bar{p}D$ . This is again a contradiction because it violates the indifference condition  $c(p_m) = c(p_M) + \mu p_M D$ . The last step is to check whether there exists a gap in the support. The proof in the appendix shows that such a gap cannot arise in equilibrium given the convexity of  $c(\cdot)$ .

Once  $\bar{p}$  is known to be the maximum in the support, it is straightforward to solve for the equilibrium distribution function. Given that the expected losses must be equalized across all  $p$ 's in the support, it follows that

$$c(p) + \mu F^*(p)^{n-1} p D = c(\bar{p}) + \mu \bar{p} D \text{ for } p \in [\underline{p}, \bar{p}]$$

or

$$F^*(p; \mu) = \left\{ \frac{c(\bar{p}) + \mu \bar{p} D - c(p)}{\mu p D} \right\}^{\frac{1}{n-1}} \text{ for } p \in [\underline{p}, \bar{p}].$$

For a given belief  $\mu$ , therefore, the targets' behavior is completely characterized by the distribution function  $F^*(p; \mu)$ .

In the first stage, the terrorist makes his decision expecting the targets' choice in the next stage. An operation is initiated if and only if the cost of operation does not exceed the targets' expected loss from an attack. The optimal cutoff is given by

$$\theta^* = \int_{[0,1]^n} \max\{p_1, \dots, p_n\} D dF^*(\mathbf{p}).$$

The expression involves an order statistic  $\max\{p_1, \dots, p_n\}$ , which has a distribution function  $\{F^*(p; \mu)\}^n$ . Then the optimality condition for the terrorist can be written in a simpler form as

$$\theta \leq \int_{[0,1]} p D d(F^*(p; \mu))^n = \theta^*(\mu).$$

Notice that the dependency of the optimal cutoff on the targets' belief is made explicit in the expression  $\theta^*(\mu)$ .

The equilibrium of the game is found where the targets' belief becomes consistent with the terrorist's optimal strategy. This requires that the equilibrium belief  $\mu^*$  should satisfy

$$\Pr(\theta \leq \theta^*(\mu^*)) = G\left(\int_{[0,1]} p D d(F^*(p; \mu^*))^n\right) = \mu^*.$$

Given that  $G(\cdot)$  is strictly increasing, the condition can be also written as

$$\int_{[0,1]} p D d(F^*(p; \mu^*))^n = G^{-1}(\mu^*)$$

where  $G^{-1}(\cdot)$  is the inverse of  $G(\cdot)$ .<sup>3</sup>

**Proposition 1** *The symmetric equilibrium of the game is characterized as follows:*

i)  $\mu^* \in (0,1)$  is uniquely determined by  $G(\theta^*(\mu^*)) = \mu^*$

ii)  $s = s_1$  if and only if  $\theta \leq \theta^* = \int_{[0,1]} pDd(F^*(p; \mu^*))^n$

iii)  $F^*(p; \mu^*)$  is given by

$$F^*(p; \mu^*) = \begin{cases} 0 & p < \underline{p}^* \\ \left\{ \frac{c(\bar{p}^*) + \mu^* \bar{p}^* D - c(p)}{\mu^* p D} \right\}^{\frac{1}{n-1}} & p \in [\underline{p}^*, \bar{p}^*] \\ 1 & \bar{p}^* < p \end{cases}$$

where  $\underline{p}^*$  is implicitly defined by  $c(\underline{p}^*) \equiv c(\bar{p}^*) + \mu^* \bar{p}^* D$  and  $\bar{p}^* \equiv \operatorname{argmin} c(p) + \mu^* p D$ .

## IV. Analysis

From the targets' perspective, the outcome of the game is not satisfactory for two reasons. In equilibrium, each target tries to divert an attack to others, which results in a dispersion in the level of investment. The joint expected loss may be reduced further if the targets can coordinate on their efforts. In addition to this coordination problem, the targets suffer from a commitment problem that keeps them from achieving the first-best outcome. Given that the probability of an attack depends on the terrorist's belief about the targets' investment, the targets could lower their expected loss if they could commit to a high level of investment. But such a scheme requires that the targets should maintain the investment high even when the risk of an attack remains low. The targets' strategy will fail, therefore, unless they have a proper mechanism to support the commitment.

---

**3** Although the symmetric equilibrium of the game is completely characterized by the proposition, technically it is not unique in the usual sense. Given that the targets' strategy involves mixing, any distribution function that differs from  $F^*(p; \mu^*)$  on a set of measure zero will also be an equilibrium strategy. Hence, the uniqueness in this case should be the one among these equivalent classes of functions.

To identify each of these issues separately, two hypothetical models are considered in the following analysis. The first model assumes that the targets can coordinate their investments but cannot commit to a certain investment level. The second model allows for both coordination and commitment. The analysis then proceeds to examine whether contracts can be designed to eliminate the two types of inefficiencies.

### 1. Equilibrium with Coordination

The terrorist and the targets play the same game as before except that the targets minimize the joint expected loss by coordinating their actions. The targets' strategy is to choose a joint distribution function  $F(\mathbf{p})$  defined on an  $n$ -dimensional unit cube  $[0,1]^n$ . Notice that this allows for the possibility that two or more targets' choices are correlated with each other. There is no change in the terrorist's problem in the first- and the third stage of the game.

When coordination is allowed, the targets must set the same investment level across all targets. Given that only the weakest target will be attacked by the terrorist, investing more in one target than in others would be simply meaningless. More formally,

**Lemma 4** *In any equilibrium with coordination, the targets choose  $p_1 = p_2 = \dots = p_n$  with probability 1.*

Without loss of generality, therefore, one may restrict attention to a one-dimensional distribution function to examine the targets' problem. Then the joint expected loss can be written as

$$\int_{[0,1]} \{nc(p) + \mu pD\} dF(p).$$

The integrand is strictly convex hence has a unique minimum. Let  $p^c(\mu) \equiv \operatorname{argmin} nc(p) + \mu pD$ . This implies that the joint expected loss will be minimized if  $p^c(\mu)$  is set across all targets.

The terrorist's choice in the first stage follows a cutoff rule. The optimal cutoff has a simple form in this case:

$$\theta^c(\mu) = \int_{[0,1]^n} \max\{p_1, \dots, p_n\} D dF^c(\mathbf{p}) = p^c(\mu)D.$$

The equilibrium of the game is determined by finding a consistent belief  $\mu^c$  that satisfies

$$\Pr(\theta \leq \theta^c(\mu^c)) = G(\theta^c(\mu^c)) = \mu^c.$$

**Proposition 2**     *The equilibrium with coordination is characterized as follows:*

- i)  $\mu^c \in (\mu^*, 1)$  is uniquely determined by  $G(\theta^c(\mu^c)) = \mu^c$
  - ii)  $s = s_1$  if and only if  $\theta \leq \theta^c(\mu^c) = p^c D$
  - iii) the targets set  $p_1 = p_2 = \dots = p_n = p^c \in (p_{\min}, p_{\max})$  with probability 1
- where  $p^c \equiv \operatorname{argmin}_p nc(p) + \mu^c p D$ .

The difference with the equilibrium in the basic model is that now the targets internalize the attack-diverting externalities by coordinating their actions. As a consequence, the investment becomes equalized across all targets and hence the dispersion in deterrence level is no longer observed in equilibrium. Another interesting feature of the equilibrium is that an attack becomes more likely when coordination is allowed, i.e.,  $\mu^c > \mu^*$ . The lack of coordination leads to an over-investment in deterrence efforts.

## 2. Equilibrium with Coordination and Commitment

Although coordination will certainly help, it does not exhaust all the possibilities for improvement. The targets' problem with coordination is given by

$$\min_p nc(p) + \mu p D.$$

The targets minimize the joint expected loss taking their belief  $\mu$  as given. But the probability of an attack, which must coincide with the targets' belief in equilibrium, depends on the terrorist's belief about the targets' action. This suggests that the expected loss could be reduced further if the targets were able to control the terrorist's belief directly.

Consider a modified game where the targets can commit to their choice before the terrorist makes his first move. The rest of the game is the same as before and the targets are assumed to coordinate their decisions. In the new game, therefore, there is an additional stage in which the targets are allowed to make their commitment.

The outcome of the modified game depends on whether such a commitment can be made credible by the targets. A commitment will have no effect in equilibrium, for instance, if the targets can costlessly withdraw from whatever claim they made in

stage zero. This implies that the targets will discard any previous commitment that requires a suboptimal action in the second stage. The only credible commitments in this case must be the ones that would have been chosen in the second stage even without making them in the first place.

If there is a way to make a credible but non-trivial commitment, however, the targets can have a strategic advantage over the terrorist. Suppose that such a commitment has been made by the targets in stage zero. Since the investment must be equalized across all targets with coordination, one may restrict attention to a common success probability  $p$ . Given that the commitment is credible, the terrorist will initiate an operation if and only if the cost of operation  $\theta$  does not exceed its benefit  $pD$ . Then an attack will occur probability  $\Pr(\theta \leq pD) = G(pD)$ . But this must be expected by the targets in the commitment stage. The targets' problem then becomes

$$\min_p nc(p) + G(pD)pD.$$

The objective function correctly reflects the fact that the targets' choice changes the probability of an attack in equilibrium. Committing to a lower  $p$  will influence the terrorist's belief and hence decrease the probability of an attack. In equilibrium, therefore, the targets will invest more in deterrence compared to the case where only coordination is allowed.

**Proposition 3** *The equilibrium with coordination and commitment is characterized as follows:*

- i)  $\mu^o \in [0, \mu^c)$  is uniquely determined by  $G(\theta^o) = \mu^o$
  - ii)  $s = s_1$  if and only if  $\theta \leq \theta^o = p^o D$
  - iii) the targets set  $p_1 = p_2 = \dots = p_n = p^o \in [p_{\min}, p^c)$  with probability 1
- where  $p^o \equiv \operatorname{argmin} nc(p) + G(pD)pD$ .

Compared to the case where only coordination is allowed, the targets invest more in deterrence when they can also commit to their actions. An attack becomes less likely as a result and it may even become totally prevented ( $\mu^o = 0$ ) in equilibrium. In either of the two previous equilibria, on the other hand, an attack always occurs with a positive probability. As will be shown shortly, however, a complete deterrence turns out to be quite difficult to achieve in practice.

### 3. Contracting

The analysis so far has shown that the targets may improve their fate if they can successfully coordinate and commit to their actions. Given that each target has an incentive to pursue its own interest at the expense of others, what remains to be shown is exactly how they may achieve the coordination and commitment.

The answer will be rather straightforward if the targets can write a binding contract on their investment. If a contract requires that each target should invest to set  $p_i = p^o$ , the first-best will be attained as long as a large enough penalty is imposed for a breach. To implement the contract, however, the court should be able to verify the investment made by the targets. The problem is that the investment, which may include the amount of “efforts” as well as monetary spending by the targets, will be difficult to measure and hence to verify. Alternatively, a contract could be written directly on  $p_i$ 's. But verifying them in court will not be an easy task either.

A more realistic approach would be to write a contract based on the outcome of an attack. A contract then must specify a transfer payments between the victim and the rest of the targets. The payments will depend on the three contingencies of the game: i) an attack occurs and succeeds, ii) an attack occurs but fails, and iii) no attack occurs. Let  $B(b)$  be the payment made by the victim (other targets) after a successful attack and  $B'(b')$  be the payment made by the intended victim (other targets) after a failed attack. If an attack does not occur, there is no victim to be treated differently. Define  $a$  to be the payment imposed on all targets in this case. A payment can be negative, which will be the case if a target receives a transfer from others.

To make such a contract enforceable, the identity of the victim needs to be verified. It will be easy to identify the victim once damage has been made by an attack. It may not be so, however, when an attack has been attempted but failed. There are two cases to consider. If the attack was defeated by the target on site, the identification will not be much of a problem. If the operation has been detected before completion, on the other hand, verifying the incident let alone the intended target could be problematic. In the following analysis, the optimal contract will be characterized first under the assumption that all three contingencies can be verified. The case when the second and the third contingency cannot be differentiated will be discussed later.

Without any restrictions on the structure, a contract may specify transfer amounts that are not consistent with each other. For instance, it may happen that a target is entitled to a positive transfer that exceeds the amount of payments made by the rest

of the targets. This motivates a concept of budget-balancing contracts. More formally, a contract is called *budget balancing* if the payments of all targets sum up to zero in each contingency. This implies that a budget-balancing contract must satisfy three conditions: i)  $B + (n - 1)b = 0$ , ii)  $B' + (n - 1)b' = 0$ , and iii)  $na = 0$ . An immediate consequence of this restriction is that  $a$  must be equal to zero, i.e., there should be no transfer among the targets in case an attack has not occurred.

Consider a situation where the targets play the same game as in the basic model except that now they can write a binding contract before the terrorist makes his first move. The question is whether there exists a budget-balancing contract that can help the targets achieve their first-best. It turns out that such a contract does not exist.

**Proposition 4** *Suppose that  $\mu^0 > 0$ . Then the first best can be implemented by a budget-balancing contract. In such a contract,  $b' > 0$ .*

The fact that an optimal contract should set  $b' > 0$  has an interesting implication. Given the budget-balancing condition, this means that  $B' = -(n - 1)b' < 0$ , i.e., a contract should specify a positive transfer to the ‘victim’ of a failed attack. To enforce the clause, however, the court must be able to verify whether there has been such an attempt, and if so, the identity of its intended target. As is pointed out earlier, however, verifying a failed attack may not be an easy task especially when the operation failed because it was detected before completion.

To examine the consequences of this problem, suppose now that the victim’s identity cannot be verified in a failed attack. Then a contract must specify the same transfer payment for all targets in case there has been an unsuccessful attempt by the terrorist, i.e.,  $B' = b'$ . But this implies that  $B' = b' = 0$  because the payments must satisfy the budget-balancing condition  $B' = (n - 1)b'$ .

**Corollary** *Suppose that  $\mu^0 > 0$ . If the victims cannot be verified in failed attacks, there is no budget-balancing contract that implements the first best.*

## V. Concluding Remarks

The analysis has shown that the commitment problem faced by the targets makes it difficult to eliminate terrorism risk completely. Absent an imminent threat, targets

cannot be expected to make sufficient amount of deterrence efforts, which makes them vulnerable to a potential terrorist attack. Although complete deterrence may not be optimal for the targets, this finding does exemplify the limitations of defensive counterterrorism policy.

Contracting turns out to be an effective instrument that can be used to mitigate the targets' problem. An implicit assumption, however, is that such a contract will be signed by all of the targets exposed to a common terrorist threat. Even though the contracting is to achieve a collective good, reaching an agreement may not be an easy task especially when many parties are involved in the process. Moreover, for a contract to function properly, it must contain a clause that penalizes a party who withdraws later. Otherwise, every target will find it irresistible to withdraw from the contract and slightly undercut the other targets. Such a deviation will be profitable because, with only a minor increase in the deterrence investment, the deviant will never incur the loss from an attack nor the transfer payment to other targets. If the targets involved are subject to a common authority, the problem may be resolved by making the participation mandatory. Examples would be private parties in a single country and local/state governments under a federal government. If the targets involved are different sovereign states, however, participation must be voluntary. Without sufficient contractual penalties, therefore, cooperation will be difficult to maintain in this case.

Another issue regarding the implementation of an optimal contract is that it requires the verification of the victim's identity in both successful and failed attacks. The victim needs to be rewarded in a failed attack because it will provide additional incentive for deterrence, which is necessary to achieve the first best. Verifying the victim could be problematic, however, especially when a failure is caused by detection instead of defeat. Given that verification matters to the extent that it affects the targets' incentives, the problem will become most severe when the targets' deterrence depends heavily on detection rather than defeat. But detecting terrorist operations, which requires sophisticated intelligence capabilities as well as the authority to apply them, is mainly the responsibility of the government. This suggests that cooperation might be more difficult to achieve among different countries than among private parties in a single country.

Terrorism is a complicated subject which involves multidimensional strategic interactions. This article examines two of such interactions in a unified framework. Although the analysis is carried out in a stylized model with a single terrorist group and multiple identical targets, the main finding—resolving the coordination and the commitment problem at the same time is in general difficult—seems to be robust. It would be of interest to see whether a more sophisticated mechanism can be found

which may improve upon the simple contracts considered here.

From a counterterrorism perspective, this article deals only with *defensive* measures as opposed to *proactive* measures (Rosendorff and Sandler, 2004). A more complete picture will be obtained if both types of measures are incorporated into the model. Such analysis will be best performed in a dynamic setting where terrorists and targets interact over time. The approach adopted in Keohane and Zeckhauser (2003) seems particularly promising in this respect. These are some of the questions left for future research.

## References

- Arce M., D. G. and T. Sandler, "Counterterrorism: A Game Theoretic Analysis," *Journal of Conflict Resolution* 49, 2005, pp.183~200.
- Aumann, R. J., "Acceptable Points in General Cooperative  $n$ -person Games," in A. W. Tucker and R. D. Luce (eds.), *Contributions to the Theory of Games*, Princeton: Princeton University Press, 1959.
- Bier, V., S. Oliveros, and L. Samuelson, "Choosing What to Protect: Strategic Defensive Allocation Against an Unknown Attacker," *Journal of Public Economic Theory* 9, 2007, pp.563~587.
- Chung, K. L., *A Course in Probability Theory*, New York: Academic Press, 1974.
- Heal, G. and H. Kunreuther, "IDS Models of Airline Security," *Journal of Conflict Resolution* 49, 2005, pp.201~217.
- Keohane, N. O. and R. J. Zeckhauser, "The Ecology of Terror Defense," *Journal of Risk and Uncertainty* 26, 2003, pp.201~229.
- Konrad, K. A., "The Investment Problem in Terrorism," *Economica* 71, 2004, pp.449~459.
- Lakdawalla, D. and G. Zanjani, "Insurance, Self-protection, and the Economics of Terrorism," *Journal of Public Economics* 89, 2005, pp.1891~1905.
- Rosendorff, P. and T. Sandler, "Too Much of a Good Thing?: The Proactive Response Dilemma," *Journal of Conflict Resolution* 48, 2004, pp.657~671.
- Sandler, T., "Collective Action and Transnational Terrorism," *World Economy* 26, 2003, pp.779~802.
- Sandler, T. and W. Enders, "An Economic Perspective on Transnational Terrorism," *European Journal of Political Economy* 20, 2004, pp.301~316.
- Schelling, T., *The Strategy of Conflict*, Cambridge: Harvard University Press, 1960.
- Siqueira, K., "Political and Militant Wings within Dissident Movements and Organizations," *Journal of Conflict Resolution* 49, 2005, pp.218~236.
- Varian, H., "A Model of Sales," *American Economic Review* 70, 1980, pp.651~659.

## Appendix

### Proof of Lemma 1

Suppose not. Then there exists a  $\mathbf{p}^* = (p_1^*, \dots, p_n^*)$  such that target  $i$  sets  $p_i^*$  with probability 1 in equilibrium. It suffices to consider the case  $\mu^* > 0$ . Notice first that  $\max_j p_j^* \leq p_{\max}$  in this case. If  $\max_j p_j^* > p_{\max}$ , the target with the highest probability may reduce its expected loss by choosing  $\max_j p_j^* - \varepsilon$ . This is because, for a small enough  $\varepsilon > 0$ , the expected loss from an attack will decrease while the cost of investment will remain constant. But this in turn implies that  $p_i^* = p^*$  for all  $i$ . To confirm, suppose to the contrary that  $p_i^* < \max_j p_j^*$  for some  $i$ . Then target  $i$ 's cost of investment must be strictly decreasing for  $p_i \in (p_i^*, \max_j p_j^*)$  because  $p_i^* < \max_j p_j^* \leq p_{\max}$ . This means that target  $i$  can lower its expected loss by setting  $p_i' = p_i^* + \varepsilon$  instead of  $p_i^*$ . As long as  $\varepsilon > 0$  is small enough to satisfy  $p_i^* + \varepsilon < \max_j p_j^*$ , target  $i$  can reduce its cost of investment by  $c(p_i^*) - c(p_i^* + \varepsilon)$  without changing the probability of being attacked. The assumption  $p_i^* < \max_j p_j^*$  hence leads to a contradiction. Hence, the only remaining possibility becomes  $p_i^* = p^* \leq p_{\max}$  for all  $i$ . But this is not consistent with equilibrium, either. Suppose that such an equilibrium exists. Then the expected loss of each target must be  $c(p^*) + \frac{1}{n}\mu^*p^*D$ . If a target undercuts  $p^*$  by  $\varepsilon > 0$ , however, it can divert a potential attack to others with probability 1. The total expected loss will become  $c(p^* - \varepsilon)$ , which will be smaller than the equilibrium loss  $c(p^*) + \frac{1}{n}\mu^*p^*D$  for a small enough  $\varepsilon > 0$ .  $\square$

### Proof of Lemma 2

The proof is essentially the same as the one in Varian (1980). Suppose that there is a jump at  $p$ . Since there can be only countably many jumps in any probability distribution, it is possible to find a small  $\varepsilon > 0$  such that a jump does not occur at  $p - \varepsilon$ . Consider a deviation strategy where a target  $i$  sets  $p - \varepsilon$  with the

probability with which he used to set  $p$ , and sets  $p$  with probability zero. The net gain from such a deviation becomes

$$\begin{aligned} & \Pr(p_j > p \text{ for some } j \neq i)c(p) - \Pr(p_j > p \text{ for some } j \neq i)c(p - \varepsilon) \\ & \quad + \Pr(p_j < p \text{ for all } j \neq i)\{c(p) + \mu p D\} \\ & \quad - \Pr(p_j < p - \varepsilon \text{ for all } j \neq i)\{c(p - \varepsilon) + \mu(p - \varepsilon)D\} \\ & + \sum_{k=1}^{n-1} \Pr(p_j \leq p \text{ for all } j \neq i, k \text{ other targets set } p_j = p \text{ for } j \neq i)\left\{c(p) + \mu \frac{p}{k+1} D\right\} \\ & - \sum_{k=1}^{n-1} \Pr(p_j \leq p \text{ for all } j \neq i, k \text{ other targets set } p - \varepsilon < p_j \text{ for } j \neq i)c(p - \varepsilon). \end{aligned}$$

As  $\varepsilon$  converges to zero, the sum of the first four terms converges to zero but the sum of the last two terms remains positive. This leads to a contradiction because the deviation strategy leads to a lower expected loss than the assumed equilibrium strategy.  $\square$

### Proof of Lemma 3

Given the assumptions  $\lim_{p \rightarrow 0} c'(p) + D < 0$  and  $c'(p_{\max}) = 0$ , the first-order condition for the minimization problem is satisfied as an equality in  $(0, p_{\max}]$ . In other words, there exists a  $\bar{p} \in (0, p_{\max}]$  such that

$$c'(\bar{p}) + \mu D = 0.$$

This follows from the fact that the function  $c'(\cdot) + \mu D$  is continuous with  $\lim_{p \rightarrow 0} c'(p) + \mu D \leq \lim_{p \rightarrow 0} c'(p) + D < 0$  and  $0 \leq c'(p_{\max}) + \mu D = \mu D$ . Since the objective function is strictly convex, such a  $\bar{p}$  must be the unique solution to the minimization problem. Given that  $c'(\cdot)$  is strictly increasing, the first-order condition can be rewritten as

$$\bar{p} = c'^{-1}(-\mu D)$$

where  $c'^{-1}(\cdot)$  is the inverse of  $c'(\cdot)$ .  $\bar{p}(\mu)$  must be continuously differentiable with respect to  $\mu$  given that  $c(\cdot)$  is twice continuously differentiable.

For a given  $\bar{p}$ , it is straightforward to show that  $\underline{p} \in (0, \bar{p}]$  is also uniquely determined. The assumption  $D < \lim_{p \rightarrow 0} c(p)$  implies

$$\lim_{p \rightarrow 0} c(p) > D \geq c(p_{\max}) + \mu p_{\max} D \geq c(\bar{p}) + \mu \bar{p} D$$

The second inequality follows from  $c(p_{\max}) = 0$  and  $0 \leq \mu, p_{\max} \leq 1$  while the third equality uses the fact  $\bar{p} = \operatorname{argmin} L_{\max}(p)$ . Since  $c(\cdot)$  is strictly decreasing with  $c(\underline{p}) \leq c(\bar{p}) + \mu\bar{p}D$ , there must be a unique  $\underline{p} \in (0, \bar{p}]$  such that  $c(\underline{p}) = c(\bar{p}) + \mu\bar{p}D$ . Given that  $\bar{p}(\mu)$  is continuously differentiable, so is  $\underline{p}(\mu)$ , which is a continuously differentiable function of  $\bar{p}$ .  $\square$

### Proof of Proposition 1

The following claim 1 through claim 5 establish the first part of the proposition. Define  $M(p) \equiv c(p) + F^*(p)^{n-1}\mu pD$ . Let  $\operatorname{supp}F(\cdot)$  be the support of the distribution function  $F(\cdot)$ , i.e.,  $\operatorname{supp}F(x) \equiv \{x: F(x + \varepsilon) - F(x - \varepsilon) > 0 \text{ for any } \varepsilon > 0\}$ .

**Claim 1** For all  $p', p'' \in \operatorname{supp}F^*(p)$ ,  $M(p') = M(p'')$ .

*Proof* Suppose not. Then there exists  $\hat{p} \in \operatorname{supp}F^*(p)$  such that

$$M(\hat{p}) > \min_{p \in \operatorname{supp}F^*(p)} M(p) \equiv \underline{M}.$$

Notice that the support of a distribution function is closed (see Chung, 1974). Given that  $\operatorname{supp}F^*(p) \subset [0, 1]$  is also bounded, the minimum on the right-hand side is well-defined. Let  $\underline{P} = \{p: p = \operatorname{argmin} M(p)\}$ . Given that  $c(p)$  and  $F^*(p)$  are both continuous,  $M(p)$  is also continuous. This means that there exists an interval  $(\hat{p} - \varepsilon, \hat{p} + \varepsilon)$  such that for all  $p \in (\hat{p} - \varepsilon, \hat{p} + \varepsilon)$ ,  $M(p) > \underline{M}$ . Notice that  $\hat{p} \in \operatorname{supp}F^*(p)$  implies  $F^*(\hat{p} + \varepsilon) - F^*(\hat{p} - \varepsilon) > 0$ . Consider a deviation strategy which assigns probability 1 to a  $p \in \underline{P}$ . The net gain from such a deviation is given by

$$\int_{[0,1]} M(p) dF^*(p) - \underline{M} = \int_{[0,1]} \{M(p) - \underline{M}\} dF^*(p) \geq \int_{(\hat{p}-\varepsilon, \hat{p}+\varepsilon)} \{M(p) - \underline{M}\} dF^*(p) > 0$$

which contradicts the assumptions that  $F^*(p)$  is the equilibrium strategy.  $\square$

**Claim 2**  $F^*(\bar{p}) = 1$  and  $F^*(\underline{p}) = 0$ .

*Proof* Define  $p_m \equiv \min p \in \operatorname{supp}F^*(p)$  and  $p_M \equiv \max p \in \operatorname{supp}F^*(p)$ . Suppose to the contrary that  $F^*(\bar{p}) < 1$ . This implies  $p_M > \bar{p}$ . Given that  $p \in \operatorname{supp}F^*(p)$ , the equilibrium expected loss is given by

$$M(p_M) = c(p_M) + F^*(p_M)^{n-1}\mu p_M D = c(p_M) + \mu p_M D.$$

Recall that  $L_{\max}(p) = c(p) + \mu p D$  is strictly increasing in  $p$  for  $p > \bar{p}$ . If a target deviates by playing  $p = \bar{p}$  with probability one, the net gain becomes

$$c(p_M) + \mu p_M D - M(\bar{p}) \geq c(p_M) + \mu p_M D - \{c(\bar{p}) + \mu \bar{p} D\} > 0$$

which contradicts the assumptions that  $F(p)$  is the equilibrium strategy. A similar argument establishes  $F^*(\underline{p}) = 0$ . Assuming  $F^*(\underline{p}) > 0$  implies  $p_m < \underline{p}$ . Then the equilibrium expected loss must be equal to

$$M(p_m) = c(p_m) + F^*(p_m)^{n-1} \mu p_m D = c(p_m).$$

But a target incurs at most  $c(\bar{p}) + \mu \bar{p} D = c(\underline{p})$  by playing  $p = \bar{p}$  with probability one. Therefore, net gain from such a deviation becomes at least

$$c(p_m) - c(\underline{p}) > 0$$

given that  $c(\cdot)$  is strictly decreasing. Again, this contradicts the assumption that  $F^*(p)$  is the equilibrium strategy.  $\square$

**Claim 3**  $F^*(\bar{p} - \varepsilon) < 1$  for any  $\varepsilon > 0$ .

*Proof* Suppose not. Then it follows that  $p_M < \bar{p}$ . Then the equilibrium expected loss must be equal to  $M(p_M) = c(p_M) + \mu p_M D$ . Consider a deviation strategy that assigns probability one to  $p = \bar{p}$ . Notice that  $p_M < \bar{p}$  implies  $M(\bar{p}) = c(\bar{p}) + \mu \bar{p} D$ . Given that  $L_{\max}(p) = c(p) + \mu p D$  is strictly decreasing in  $p$  for  $p < \bar{p}$ , the net gain from such a deviation becomes

$$M(p_M) - M(\bar{p}) = c(p_M) + \mu p_M D - \{c(\bar{p}) + \mu \bar{p} D\} > 0$$

which leads to contradiction.

Given that  $\bar{p} \in \text{supp } F^*(p)$ , it follows that  $M(p) = M(\bar{p})$  for all  $p \in \text{supp } F^*(p)$ . The equation  $M(p) = M(\bar{p})$  implicitly determines the equilibrium distribution function  $F^*(p)$ :

$$M(p) = c(p) + F^*(p)^{n-1} \mu p D = c(\bar{p}) + \mu \bar{p} D = M(\bar{p})$$

$$F^*(p) = \left\{ \frac{c(\bar{p}) + \mu \bar{p} D - c(p)}{\mu p D} \right\}^{\frac{1}{n-1}}.$$

To confirm that  $F^*(p)$  satisfies the requirements for a distribution function. Notice first that  $F^*(\underline{p}) = 0$ ,  $F^*(\bar{p}) = 1$ . Also, for  $p \in (\underline{p}, \bar{p})$

$$\begin{aligned} \frac{\partial F^*(p)}{\partial p} &= -\frac{1}{n-1} F^*(p)^{\frac{2-n}{n-1}} \frac{c'(p)p + c(\bar{p}) + \mu\bar{p}D - c(p)}{\mu p^2 D} \\ &= -\frac{1}{n-1} F^*(p)^{\frac{2-n}{n-1}} \frac{\{c'(p) + \mu D\}p + c(\bar{p}) + \mu\bar{p}D - \{c(p) + \mu p D\}}{\mu p^2 D} > 0 \end{aligned}$$

where the inequality follows from  $c(\bar{p}) + \mu\bar{p}D < c(p) + \mu p D$  and  $c'(p) + \mu D < 0$  for  $p \in (\underline{p}, \bar{p})$ .  $\square$

**Claim 4**  $F^*(\underline{p} + \varepsilon) > 0$  for any  $\varepsilon > 0$ .

*Proof* Suppose not. This implies  $p_m > \underline{p}$ . Then the equilibrium expected loss must be equal to  $M(p_m) = c(p_m)$ . Given that  $\bar{p} = p_M \in \text{supp } F^*(p)$ , however, the equilibrium expected loss must be equal to  $c(\bar{p}) + \mu\bar{p}D$ . But this leads to a contradiction because

$$c(p_m) < c(\underline{p}) = c(\bar{p}) + \mu\bar{p}D$$

given that  $c(\cdot)$  is strictly decreasing.  $\square$

**Claim 5**  $F^*(p'') - F^*(p') > 0$  for all  $\underline{p} < p' < p'' < \bar{p}$ .

*Proof* Suppose not, i.e., there exist  $p', p'' \in (\underline{p}, \bar{p})$  such that  $F^*(p'') - F^*(p') = 0$  and  $p' < p''$ . Let  $P_F \equiv \{p \in (\underline{p}, \bar{p}) : F^*(p) = F^*(p') = F^*(p'')\}$  be the preimage of  $F^*(p')$ . Given that the function  $F^*(\cdot)$  is continuous, the set  $P_F$  must be closed. Then  $p_L \equiv \min p \in P_F$  and  $p_H \equiv \max p \in P_F$  are well-defined. By construction, it must be that  $p_L \leq p' < p'' \leq p_H$  and  $p_L, p_H \in \text{supp } F^*(p)$ . Then the equilibrium expected loss is equal to

$$c(p_L) + F^*(p')^{n-1} \mu p_L D = c(p_H) + F^*(p')^{n-1} \mu p_H D.$$

Let  $M_F(p) \equiv c(p) + F^*(p')^{n-1} \mu p D$ . Given that  $c(p)'' > 0$ , the function  $M_F(p)$  is strictly convex with  $M_F(p_L) = M_F(p_H)$ . This implies  $M_F(p) < M_F(p_L)$  for all  $p \in (p_L, p_H)$ . If a target sets any  $p \in (p_L, p_H)$  with probability one, however, the expected loss becomes  $c(p) + F^*(p')^{n-1} \mu p D = M_F(p)$ , which is

smaller than the equilibrium expected loss  $M_F(p_L)$ . This contradicts the assumption that  $F^*(p)$  is the equilibrium distribution function.

For the second part of the claim, let  $\Delta^*(\mu) \equiv G(\theta^*(\mu)) - \mu$ . Then  $\Delta^*(\mu^*) = 0$ . Notice that

$$\theta^*(\mu) = \int_{\underline{p}}^{\bar{p}} p D d(F^*(p; \mu))^n = \bar{p}D - D \int_{\underline{p}}^{\bar{p}} (F^*(p; \mu))^n dp$$

where the last equality follows from integrating by parts. Given that  $\bar{p}(\mu)$ ,  $\underline{p}(\mu)$ , and  $(F^*(p; \mu))^n$  are all continuous in  $\mu$ ,  $\Delta^*(\mu)$  is also continuous. Given the continuity of  $\Delta^*(\cdot)$ , the existence of  $\mu^*$  is proved if one can show that  $\Delta^*(1) < 0 < \lim_{\mu \rightarrow 0} \Delta^*(\mu)$ . To see this, notice first that the equilibrium distribution  $(F^*(p; \mu))^n$  becomes degenerate at  $p = p_{\max}$  as  $\mu \rightarrow 0$ . This implies

$$\lim_{\mu \rightarrow 0} \Delta^*(\mu) = \lim_{\mu \rightarrow 0} G\left(\int_{[0,1]} p D d(F^*(p; \mu))^n\right) = G(p_{\max}D) > 0$$

where the inequality follows from the assumption  $p_{\max}D > \underline{\theta}$ . Moreover,  $p \leq p_{\max}$  implies

$$\theta^*(1) = \int_{[0,1]} p D d(F^*(p; 1))^n \leq p_{\max}D$$

therefore,

$$\Delta^*(1) = G(\theta^*(1)) - 1 \leq G(p_{\max}D) - 1 < G(\bar{\theta}) - 1 = 0$$

where the last inequality follows from the assumption  $p_{\max}D < \bar{\theta}$ .

To prove the uniqueness of  $\mu^*$ , it suffices to show that  $\Delta^*(\mu)$  is strictly decreasing in  $\mu$ . Using the envelope theorem to evaluate  $\frac{\partial c(\bar{p}) + \mu \bar{p}D}{\partial \mu} = \bar{p}D$ , one can show that

$$\frac{\partial (F^*(p; \mu))^n}{\partial \mu} = -\frac{n}{n-1} \left\{ \frac{c(\bar{p}) + \mu \bar{p}D - c(p)}{\mu p D} \right\}^{\frac{1}{n-1}} \frac{c(\bar{p}) - c(p)}{\mu^2 p D} > 0$$

Given that  $\frac{\partial (F^*(p; \mu))^n}{\partial \mu}$ ,  $\frac{\partial \bar{p}(\mu)}{\partial \mu}$ , and  $\frac{\partial \underline{p}(\mu)}{\partial \mu}$  are all continuous,  $\frac{\partial \theta^*(\mu)}{\partial \mu}$  can be evaluated using Leibniz's integral rule. This leads to

$$\frac{\partial \theta^*(\mu)}{\partial \mu} = \frac{\partial \bar{p}(\mu)}{\partial \mu} D - (F^*(\bar{p}; \mu))^n \frac{\partial \bar{p}(\mu)}{\partial \mu} D + (F^*(\underline{p}; \mu))^n \frac{\partial \underline{p}(\mu)}{\partial \mu} D - D \int_{\underline{p}}^{\bar{p}} \frac{\partial (F^*(p; \mu))^n}{\partial \mu} dp$$

$$= -D \int_{\underline{p}}^{\bar{p}} \frac{\partial (F^*(p; \mu))^n}{\partial \mu} dp < 0$$

given that  $F^*(\bar{p}; \mu) = 1$  and  $F^*(\underline{p}; \mu) = 0$ . Since  $\frac{\partial G}{\partial \theta} > 0$ , it follows that  $\frac{\partial \Delta^*(\mu)}{\partial \mu} = \frac{\partial G}{\partial \theta} \frac{\partial \theta^*}{\partial \mu} - 1 < 0$  as claimed.  $\square$

#### Proof of Lemma 4

Suppose not, i.e., there exists a set  $\tilde{P} = \{\mathbf{p}: p_i \neq p_j \text{ for some } i, j\}$  and  $\Pr(\mathbf{p} \in \tilde{P}) > 0$ . For each  $\mathbf{p} = (p_1, \dots, p_n)$ , define  $\mathbf{p}' = \max\{p_1, \dots, p_n\}(1, \dots, 1)$ . Consider an alternative strategy where the targets set  $\mathbf{p}'$  with the probability with which they used to set  $\mathbf{p}$ . The gains from such a strategy is given by

$$\begin{aligned} & \int_{[0,1]^n} [\sum_i c(p_i) + \mu \max\{p_1, \dots, p_n\} D] dF(\mathbf{p}) \\ & - \int_{[0,1]^n} [\sum_i c(\max\{p_1, \dots, p_n\}) + \mu \max\{p_1, \dots, p_n\} D] dF(\mathbf{p}) \\ & = \int_{\tilde{P}} [\sum_i \{c(p_i) - c(\max\{p_1, \dots, p_n\})\}] dF(\mathbf{p}) > 0 \end{aligned}$$

since  $\sum_{i=1}^n \{c(p_i) - c(\max\{p_1, \dots, p_n\})\} > 0$  for  $\mathbf{p} \in \tilde{P}$  and  $\Pr(\mathbf{p} \in \tilde{P}) > 0$ . This implies that a strategy where  $\Pr(\mathbf{p} \in \tilde{P}) > 0$  is never optimal, which leads to a contradiction.  $\square$

#### Proof of Proposition 2

First, it is shown that there exists a unique  $\mu^c \in (0, 1)$  satisfying the belief-consistency condition. Notice that the first-order condition for the targets' minimization problem holds as an equality, i.e.,  $nc'(p^c) + \mu D = 0$  given that  $\lim_{p \rightarrow 0} nc'(p) + \mu D < \lim_{p \rightarrow 0} c'(p) + D < 0$  and  $nc'(p_{\max}) + \mu D \geq 0$ . Since  $c'(\cdot)$  is strictly increasing, this implies  $p^c(\mu) = c'^{-1}\left(-\frac{\mu D}{n}\right)$ , where  $c'^{-1}(\cdot)$  is the inverse of  $c'(\cdot)$ .  $p^c(\cdot)$  must be continuous and strictly decreasing given that  $c'^{-1}(\cdot)$  is continuous and strictly increasing. Let  $\Delta^c(\mu) \equiv G(\theta^c(\mu)) - \mu$ . Then the equilibrium belief  $\mu^c$  must satisfy  $\Delta^c(\mu^c) = 0$ . Notice that  $\Delta^c(\cdot)$  is continuous and strictly decreasing given that  $\theta^c(\cdot) = p^c(\cdot)D$  is continuous and strictly decreasing while  $G(\cdot)$  is continuous and strictly increasing. This implies that the solution to the equation  $\Delta^c(\mu) = 0$ , if exists, must be unique. Given the assumption  $\underline{\theta} < p_{\max}D < \bar{\theta}$ , the solution does exist because  $\Delta^c(0) = G(p_{\max}D) > 0$  and  $\Delta^c(1) = G(p^c(1)D) - 1 \leq G(p_{\max}D) - 1 < 0$ . If  $\mu^c > 0$ , it must be that

$nc'(p_{\max}) + \mu^c D = \mu^c D > 0$ . This means  $p^c < p_{\max}$ . Also  $G(p_{\min}D) = G(\underline{\theta}) = 0$  implies  $p_{\min} < p^c$ .

To see that  $\mu^* < \mu^c$ , it is helpful to define  $\bar{\mu}^*$  such that  $G(\bar{p}(\bar{\mu}^*)D) = \bar{\mu}^*$ . Since  $\bar{p}(\cdot)$  and  $G(\cdot)$  are both continuous with  $G(\bar{p}(0)D) = G(p_{\max}D) > 0$  and  $G(\bar{p}(1)D) - 1 \leq G(p_{\max}D) - 1 < 0$ , such a  $\bar{\mu}^* \in (p, 1)$  is uniquely determined. Then it must be that  $\mu^* < \bar{\mu}^*$ . To confirm, suppose to the contrary that  $\mu^* \geq \bar{\mu}^*$ . Given that  $G\left(\int_{[0,1]} pDd(F^*(p; \mu^*))^n\right) = \mu^*$ , it follows

$$G\left(\int_{[0,1]} pDd(F^*(p; \mu^*))^n\right) - G(\bar{p}(\bar{\mu}^*)D) = \mu^* - \bar{\mu}^* \geq 0.$$

Since  $G(\cdot)$  is strictly increasing, this implies  $\int_{[0,1]} pd(F^*(p; \mu^*))^n \geq \bar{p}(\bar{\mu}^*)$ . Notice that the distribution function  $(F^*(p; \mu))^n$  has a support  $[\underline{p}(\mu), \bar{p}(\mu)]$  and  $\bar{p}(\mu)$  is strictly decreasing in  $\mu$ . This implies  $\int_{[0,1]} pd(F^*(p; \mu^*))^n < \bar{p}(\mu^*) \leq \bar{p}(\bar{\mu}^*)$ , which is a contradiction.

A similar argument can be used to establish  $\bar{\mu}^* < \mu^c$ . To see this, suppose to the contrary that  $\bar{\mu}^* \geq \mu^c$ . Given that  $G(p^c(\mu^c)D) = \mu^c$ , it follows

$$G(\bar{p}(\bar{\mu}^*)D) - G(p^c(\mu^c)D) = \bar{\mu}^* - \mu^c \geq 0.$$

Since  $G(\cdot)$  is strictly increasing, it must be that  $\bar{p}(\bar{\mu}^*) \geq p^c(\mu^c)$ . Consider  $\bar{p}(\mu^c)$  and  $p^c(\mu^c)$ . From the first-order conditions,

$$nc'(p^c(\mu^c)) + \mu^c D = 0$$

$$c'(\bar{p}(\mu^c)) + \mu^c D = 0.$$

which implies  $nc'(p^c(\mu^c)) = c'(\bar{p}(\mu^c))$ . But this leads to  $\bar{p}(\mu^c) < p^c(\mu^c)$  because  $c'(\cdot) < 0$  is strictly increasing. Given that  $\bar{p}(\cdot)$  is strictly decreasing and  $\bar{p}(\bar{\mu}^*) \geq p^c(\mu^c)$ , it follows that  $\bar{\mu}^* < \mu^c$ , which is a contradiction.  $\square$

### Proof of Proposition 3

Characterizing the equilibrium boils down to showing that there exists a unique  $p^o \in [p_{\min}, p^c]$ . Then  $\mu^o \in [0, \mu^c]$  follows immediately because  $p_{\min} \leq p^o < p^c$  implies  $0 = G(p_{\min}D) \leq G(p^oD) < G(p^cD) = \mu^c$ . To see that  $p^o$  is uniquely determined in  $[p_{\min}, p^c]$ , notice first that it is never optimal to set  $p^o < p_{\min}$ . The probability of an attack becomes zero once  $p$  reaches  $p_{\min}$ . If the targets set  $p$

below  $p_{\min}$ , therefore, they will increase the cost of investment without lowering the probability of an attack. Similarly, any  $p^o > p_{\max}$  cannot be optimal, either. Otherwise, the targets will be able to reduce the risk of an attack without any extra costs of investment. Without loss of generality, therefore, the domain for the targets' minimization problem can be restricted to  $p \in [p_{\min}, p_{\max}]$ .

Given the assumption  $2G'(\theta) + G''(\theta)\theta > 0$ , the targets' objective function is strictly convex for  $p \geq p_{\min}$  because

$$\begin{aligned} \frac{\partial^2(nc(p)+G(pD)pD)}{\partial p^2} &= nc''(p) + 2G'(pD)D^2 + G''(pD)pD^3 \\ &= nc''(p) + \{2G'(pD) + G''(pD)pD\}D^2 > 0. \end{aligned}$$

This implies that the solution to the minimization problem must be unique. In contrast to the equilibrium with coordination only, however, the first-order condition may not hold as an equality. Differentiating the objective function with respect to  $p$  gives

$$nc'(p) + G'(pD)pD^2 + G(pD)D \equiv \Gamma(p).$$

The second derivative is shown to be positive, which means that  $\Gamma(\cdot)$  is strictly increasing. There are two cases to consider: i)  $\Gamma(p_{\min}) \geq 0$  and ii)  $\Gamma(p_{\min}) < 0$ . If  $\Gamma(p_{\min}) \geq 0$ , the objective function is increasing for all  $p \in [p_{\min}, p_{\max}]$ . This implies that the minimum is attained at  $p = p_{\min}$ . If  $\Gamma(p_{\min}) < 0$ , the first-order condition holds as an equality, i.e., there exists a  $p^o \in (p_{\min}, p^c)$  such that  $\Gamma(p^o) = 0$ . To confirm, notice that

$$\Gamma(p^c) = nc'(p^c) + G'(p^cD)p^cD^2 + G(p^cD)D > 0$$

because  $nc'(p^c) + G(p^cD)D = 0$ . Given that  $\Gamma(\cdot)$  is continuous with  $\Gamma(p_{\min}) < 0$  and  $\Gamma(p^c) > 0$ , such a  $p^o \in (p_{\min}, p^c)$  must exist.  $\square$

#### Proof of Proposition 4

For a given contract  $(b, B, b', B')$ , a target's loss is given by

$$\begin{aligned} L_i(\mathbf{p}, s) &= c(p_i) \\ &+ I(s) \left[ \rho_i(\mathbf{p}) \{p_i(D + B) + (1 - p_i)B'\} + \{1 - \rho_i(\mathbf{p})\} \left\{ \max_{j \neq i} p_j b + \left(1 - \max_{j \neq i} p_j\right) b' \right\} \right] \end{aligned}$$

The difference with the basic model is that a target incurs a loss (or a gain, if the payment is negative) even when it is not the victim of an attack. The loss of a target

will depend on whether it is the victim of the attack and also whether the attack was successful or not. The second term on the right-hand side then gives a target's expected loss from an attack. The joint loss of the targets can be obtained by adding all individual losses:

$$\begin{aligned} L(\mathbf{p}, s) &= \sum_i c(p_i) + I(s) \left[ \max_i p_i \{D + B + (n - 1)b\} + (1 - \max_i p_i) \{B' + (n - 1)b'\} \right] \\ &= \sum_i c(p_i) + I(s) \max_i p_i D \end{aligned}$$

Due to the budget-balancing conditions, the transfer payments cancel each other out once they are summed up across all targets. Consequently, the expression for the targets' joint loss is exactly the same as the one in the basic model. Given that the terrorist cares only about the targets' joint loss and not its distribution, this implies that contracting does not change the terrorist's objective function, either.

Given a contract, the equilibrium of the game is found where the terrorist and the targets play their optimal strategies and the targets' belief becomes consistent with the terrorist's equilibrium strategy. To differentiate it with the equilibrium without contracting,  $\mu^{**}$ ,  $\theta^{**}$ , and  $F_i^{**}(\cdot)$  will be used to denote the equilibrium belief and the strategies.

**Lemma 5** *In any equilibrium with contracting,  $\mu^{**} > 0$ .*

*Proof* Suppose to the contrary that  $\mu^{**} = 0$ , i.e., the targets believe that an attack never occurs in equilibrium. Then it is never optimal for a target to choose a  $p < p_{\max}$  with a positive probability. This implies

$$\theta^{**} = \int_{[0,1]^n} \max\{p_1, \dots, p_n\} D dF^{**}(\mathbf{p}) \geq p_{\max} D.$$

Notice that the belief consistency requires  $G(\theta^{**}(\mu^{**})) = \mu^{**}$ . But this leads to a contradiction because

$$0 = \mu^{**} = G(\theta^{**}) \geq G(p_{\max} D) > G(\underline{\theta}) = 0$$

where the last inequality follows from the assumption  $\underline{\theta} < p_{\max} D$ .  $\square$

Recall that the first best may require  $\mu^0 = 0$ , i.e., there should be no attack by the terrorist. The lemma shows therefore that the targets' problem cannot be solved entirely by contracting. The question is then whether the targets may achieve the first best if it does not require a complete deterrence. A necessary condition is given by the next lemma.

**Lemma 6** If  $p_i^{**} = p^{**}$  for all  $i$ , then  $p^{**}(D + B) + (1 - p^{**})B' = p^{**}b + (1 - p^{**})b'$ .

*Proof* Suppose not. Then there are two cases to consider: i)  $p^{**}(D + B) + (1 - p^{**})B' > p^{**}b + (1 - p^{**})b'$  and ii)  $p^{**}(D + B) + (1 - p^{**})B' < p^{**}b + (1 - p^{**})b'$ . In the first case, a target may deviate by setting  $p' = p^{**} - \varepsilon$  for a small  $\varepsilon > 0$ . By doing so, the target becomes the strongest target with certainty. The net gain from the deviation becomes

$$\begin{aligned} & \{c(p^{**}) - c(p')\} + \mu^{**} \frac{n-1}{n} \{p^{**}b + (1 - p^{**})b' - p'b - (1 - p')b'\} \\ & + \mu^{**} \frac{1}{n} \{p^{**}(D + B) + (1 - p^{**})B' - p'b - (1 - p')b'\} \end{aligned}$$

As  $\varepsilon$  gets smaller, the first two terms approach zero but the last term remains positive. This contradicts the assumption that  $p^{**}$  is the targets' equilibrium strategy. A similar argument can be applied to the case  $p^{**}(D + B) + (1 - p^{**})B' < p^{**}b + (1 - p^{**})b'$ . One may consider a deviation  $p' = p^{**} + \varepsilon$ . The deviant becomes the weakest target in this case. Then the net gain is given by

$$\begin{aligned} & \{c(p^{**}) - c(p')\} + \mu^{**} \frac{1}{n} \{p^{**}(D + B) + (1 - p^{**})B' - p'(D + B) - (1 - p')B'\} \\ & + \mu^{**} \frac{n-1}{n} \{p^{**}b + (1 - p^{**})b' - p'(D + B) - (1 - p')B'\} \end{aligned}$$

As  $\varepsilon$  becomes smaller, the first two terms approach zero but the last term remains positive. Again, this leads to a contradiction.  $\square$

The meaning of the condition becomes clear once a target's objective function is examined more closely. If the targets set the same  $p^{**}$  in equilibrium, which must be the case if the equilibrium achieves the first best, target  $i$ 's expected loss is given by

$$\begin{aligned} & E^s[L_i(p_i, p_{-i}^{**}, s)] \\ & = \begin{cases} c(p_i) + \mu^{**}\{p_i(D + B) + (1 - p_i)B'\} & \text{if } p_i > p^{**} \\ c(p_i) + \mu^{**} \left[ \frac{1}{n} \{p_i(D + B) + (1 - p_i)B\} + \frac{n-1}{n} \{p_i b + (1 - p_i)b'\} \right] & \text{if } p_i = p^{**} \\ c(p_i) + \mu^{**}\{p^{**}b + (1 - p^{**})b'\} & \text{if } p_i < p^{**} \end{cases} \end{aligned}$$

The function  $E^s[L_i(p_i, p_{-i}^{**}, s)]$  is continuous except for a potential discontinuity at  $p_i = p^{**}$ . Conditional on there being an attack, a target incurs a loss of  $\{p_i(D + B) + (1 - p_i)B'\}$  if it is the victim. Otherwise, the loss becomes  $p^{**}b + (1 - p^{**})b'$ . Hence the condition in the lemma requires that a targets should be indifferent between being and not being the victim of an attack. If not, a target

will deviate by slightly lowering (or raising) its  $p$  to get the better payoff. Once the condition is satisfied, the function  $E^S[L_i(p_i, p_{-i}^{**}, s)]$  becomes continuous for all  $p_i$ .

A contract induces the first-best if  $p_i^{**} = p^o$ , i.e., each target sets  $p^o$  in equilibrium. Then it must be that

$$p^o(D + B) + (1 - p^o)B' = p^o b + (1 - p^o)b' \quad (1)$$

For the strategy to be optimal, any deviation must result in higher expected loss. Examining the target's expected loss reveals that setting any  $p < p^o$  is never profitable. Given that all other targets set  $p^o$ , lowering  $p$  unilaterally will only increase the cost of investment without changing the probability that an attack becomes successful.

A deviation by setting a  $p > p^o$ , on the other hand, may reduce a target's expected loss. One way to prevent such a deviation is to align the optimality condition of individual targets with the one that minimizes their joint expected loss. Assuming that an attack occurs with a positive probability, i.e.,  $\mu^o = G(p^o D) > 0$ , the first best solution is determined by

$$nc'(p^o) + G(p^o D)D + G'(p^o D)p^o D^2 = 0$$

or

$$c'(p^o) + G(p^o D) \frac{D}{n} \left\{ 1 + \frac{G'(p^o D)}{G(p^o D)} p^o D \right\} = 0 \quad (2)$$

On the other hand, a target's expected loss will be minimized at  $p^o$  in equilibrium if it is increasing in  $p_i$  for all  $p_i > p^o$ . Differentiating the function  $E^S[L_i(p_i, p_{-i}^{**}, s)]$  with respect to  $p_i$  and evaluating it at  $p^{**} = p^o$  gives  $c'(p_i) + \mu^o(D + B - B')$ . This implies that a target will never deviate from the first-best solution if

$$c'(p_i) + \mu^o(D + B - B') \geq 0$$

for  $p_i > p^o$ . Given the convexity of  $c(\cdot)$ , the left-hand side of the inequality increases with  $p_i$ . Hence the condition will be satisfied for all  $p_i > p^o$  if the inequality holds at  $p_i = p^o$ , i.e.,

$$c'(p^o) + \mu^o(D + B - B') \geq 0$$

Comparing the expression with (2) reveals that this inequality is reduced to

$$D + B - B' \geq \frac{D}{n} \left\{ 1 + \frac{G'(p^o D)}{G(p^o D)} p^o D \right\} \quad (3)$$

By construction, the first best can be supported as an equilibrium if a contract satisfies (1) and (3) along with the two budget-balancing conditions. As long as other targets are doing the same, it will be optimal for a target to set  $p = p^o$ . Given the targets' strategy, the terrorist should initiate an operation only when  $\theta \leq \theta^o = p^o D$ . The targets' belief  $\mu^o$  is then consistent with the terrorist's strategy because  $\mu^o = G(p^o D)$ . Given that the condition (3) is in inequality form, however, the optimal contract is not determined uniquely. In case (3) holds as a strict inequality, each target's expected loss must be strictly increasing at  $p = p^o$ . This implies that the targets would reduce their expected loss further if they could collectively choose to lower their  $p$ 's. Then a strong refinement in the spirit of Aumann (1959) may eliminate such a possibility and hence resolve the indeterminacy.

Now, for the proof of proposition 4, notice first that any budget-balancing contract that implements the first best must satisfy

$$B + (n - 1)b = 0$$

$$B' + (n - 1)b' = 0$$

$$p^o(D + B) + (1 - p^o)B' = p^o b + (1 - p^o)b'$$

Solving for  $b$ ,  $B$ , and  $B'$  gives,

$$b = \frac{D}{n} - \frac{1-p^o}{p^o} b'$$

$$B = -(n - 1)b = -(n - 1) \left( \frac{D}{n} - \frac{1-p^o}{p^o} b' \right)$$

$$B' = -(n - 1)b'$$

Then the expected loss of target  $i$  can be expressed in terms of  $b'$  only. Given that  $p_j = p^o$  for all  $j \neq i$ , this gives

$$\begin{aligned} & c(p_i) + \mu^o \{p_i(D + B) + (1 - p_i)B'\} \\ &= c(p_i) + \mu^o p_i \left\{ \frac{D}{n} + (n - 1) \frac{b'}{p^o} \right\} - \mu^o (n - 1)b' \equiv \tilde{L}(p_i; b'). \end{aligned}$$

For  $p_i = p^o$  to be optimal,  $\tilde{L}(p_i; b')$  must be increasing in  $p_i$  at  $p_i = p^o$ , i.e.,

$$\frac{\partial \tilde{L}(p^o; b')}{\partial p_i} = c'(p^o) + \mu^o \left\{ \frac{D}{n} + (n - 1) \frac{b'}{p^o} \right\} \geq 0.$$

It is easy to confirm that the inequality holds as an equality at

$b' = \frac{1}{n(n-1)} \frac{G'(p^o D)}{G(p^o D)} (p^o D)^2 > 0$ . Given that  $\frac{\partial L(p_i; b')}{\partial p_i}$  is increasing in  $p_i$ , this implies that any contract with  $b' \geq \frac{1}{n(n-1)} \frac{G'(p^o D)}{G(p^o D)} (p^o D)^2$  can support the first best given that  $b$ ,  $B$ , and  $B'$  are determined accordingly.  $\square$

#### Proof of Corollary

For the contract to implement the first best, it must satisfy (1), which is now reduced to

$$D + B = b.$$

Combined with the remaining budget-balancing constraint  $B + (n - 1)b = 0$ , the contract can be determined completely, which is given by

$$b = \frac{D}{n}$$

$$B = -(n - 1) \frac{D}{n}.$$

Then it is straightforward to verify that the contract does not implement the first best. This is because the contract violates (3), i.e.,

$$D + B - B' = b = \frac{D}{n} < \frac{D}{n} \left\{ 1 + \frac{G'(p^o D)}{G(p^o D)} p^o D \right\}. \quad \square$$