

LA-UR- 98-4476

Title:

**INCENTIVES FOR CHEATING GIVEN  
IMPERFECT DETECTION**

Author(s):

Gregory H. Canavan, P-DO

Submitted to:

For discussions outside the Laboratory

Date: 1 October 1998

*RECEIVED  
AUG 18 1999  
@ S.T.I.*



**Los Alamos**  
NATIONAL LABORATORY

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. The Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy.

Form No. 836 R5  
ST 2629 10/91

## **DISCLAIMER**

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, make any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

## **DISCLAIMER**

**Portions of this document may be illegible  
in electronic image products. Images are  
produced from the best available original  
document.**

## INCENTIVES FOR CHEATING GIVEN IMPERFECT DETECTION

Gregory H. Canavan

The incentives for cheating given imperfect detection can be discussed within the context of first strike stability. The cost reduction due to is balanced against the sanctions that would be imposed if cheating was detected. For small political sanctions, the optimum level is at high levels of cheating. For large sanctions, the optimum is at quite low levels, which discourages cheating

A model due to Immele treats the incentives for cheating—and hence benefits of transparency—for imperfect detection. In it the cost reduction due to cheating, i.e., building or hiding additional weapons, are balanced against the sanctions that would be imposed if cheating was detected.<sup>1</sup> This note adapts that model to the costs derived to discuss first strike stability. The proper balancing of the benefits and risks for cheating can be addressed with the costs derived for first strike stability. The advantages of first strike cost reduction are balanced against the losses that would be incurred if the cheating were detected. The cheating weapons affect the allocation of weapons in first and second strikes. The minimization of the overall cost of is performed by observing the expected losses for each level of cheating and choosing that which minimizes overall costs. From a nominal post-START III single-weapon baseline configuration, cheating rapidly degrades stability. For small political sanctions, the optimum level of cheating is at high levels of cheating. For large political sanctions, the optimum is at quite low levels, which discourages cheating.

Exchanges are treated by optimization techniques derived in companion notes.<sup>2</sup> For this example it is assumed that the side identified as “unprime” has  $M$  vulnerable missiles with  $m$  weapons each and  $N$  survivable missiles with  $n$  weapons each for a total of  $W = mM + nN$  weapons. If he strikes first, he delivers  $f$   $W$  weapons on “prime”’s  $M'$  vulnerable missiles and the rest in a first strike on military value targets of magnitude

$$F = (1 - f)W, \quad (1)$$

This counter force strike delivers an average of  $r = fW/M'$  weapons on each vulnerable missile, which gives them a survival probability  $Q' \approx q^r$ , where  $p = 1 - q$  is the missile single shot probability of kill. Prime’s second strike is thus

$$S' = (Q'm'M' + n'N'), \quad (2)$$

which is delivered on value. Prime’s first strike  $F'$  and unprime’s second strike  $S'$  are obtained by conjugating the expressions above, i.e., by interchanging primed and unprimed symbols.

First and second strikes are converted into costs through exponential approximations to the value of military value targets destroyed, assuming that each side has  $1/k \sim 1,000$  value

targets. The cost of damage to self and incomplete damage to other are joined with a weighting parameter  $L$ , which measures the attacker's relative preference for damage to other and prevention of damage to self.<sup>3</sup> The costs of unprime striking first and second are thus

$$C_1 \approx (1 - e^{-kS} + Le^{-kF}) / (1 + L), \quad (3)$$

$$C_2 \approx (1 - e^{-kF} + Le^{-kS}) / (1 + L). \quad (4)$$

The conventional cost ratio stability index for unprime is

$$I = C_1 / C_2, \quad (5)$$

The composite index is the product of the index for each side.

**Cheating.** The model assumes that prime is able to add additional weapons through stealth fabrication or concealment. Those weapons would reduce his first strike cost in a crisis. However, if the deception was detected, unprime could take precautionary measures, up to and including a preemptive strike. Thus, the expected value  $EV$  of prime's cheating is

$$EV = (1 - P)C_1 + PQ(C_2 + U), \quad (6)$$

where  $P$  is the probability that the cheating will be detected,  $C_1$  and  $C_2$  are the first and second strike costs derived above,  $Q$  is the probability of unprime launching a preemptive strike given that it discovers prime's deception, and  $U$  is the additional cost other than military action that would be imposed on prime, given discovery

The payoff to prime for cheating is the reduction of  $C_1$ , which can be significant. The costs are the military and non-military losses in  $C_2 + U$ . The cheater should balance them by choosing a level of weapons additions for which the probability of deception is not too great. It is plausible that the probability of detection increases with the number of weapons concealed. Below, it is assumed that

$$P = (A - 500) / 2,000, \quad (7)$$

which would increase from zero to unity as the number of weapons added,  $A$ , increased from the initial 500 weapons to 2,500 weapons on  $M = 500$  singlet missiles.

**Results.** Figure 1 shows the individual and composite stability indices produced by prime's increase from  $1 \times 500 = 500$  weapons to  $5 \times 500 = 2,500$  weapons. Unprime's index initially falls more rapidly than prime's, but for  $m > 2$  weapons per missile, unprime no longer sees an incentive to strike, prime does, and the composite index follows prime's rapid reduction.

Figure 2 shows the probability of detection and  $EV$ . For  $U = 0$ , i.e., military sanctions only,  $EV$  has a shallow minimum at  $m \approx 5$ . That is, such sanctions would provide little augmentation above the costs of being preempted. For  $U = 0.25$ , which is about an equal mix of military and non-military sanctions,  $EV$  has a minimum at  $m \approx 2.5$ . For  $U = 0.5$  and larger values, which reflect large non-military sanctions, the optimum shifts to  $m \approx 1$ , i.e. against cheating.

These calculations vary  $U$  for a fixed  $Q = 0.5$ , that is, a probability of 0.5 that unprime will feel it necessary to preempt if serious cheating is uncovered. Obviously, increasing or decreasing

$Q$  would just increase or decrease prime's losses proportionally, which would have much the same effect as increasing  $U$ .

**Summary and conclusions.** The proper balancing of the risks for cheating can be addressed with the costs derived for discussion of first strike stability in the context of a model due to Immele. In it, the advantages of first strike cost reduction are balanced against the losses that would be incurred if the cheating was detected. Treated in this way, the cheating weapons do affect the allocation of weapons in first and second strikes, but the minimization of the overall cost of cheating does not. Instead, minimization is performed by observing the expected losses for each level of cheating and choosing that which minimizes overall costs. From a nominal post-START III single-weapon baseline configuration, cheating rapidly degrades stability. For small political sanctions, the optimum level of cheating is at high levels of weapons. For large political sanctions, the optimum is at quite low levels, which discourages cheating.

## References

---

<sup>1</sup>. J. Immele, private communication, 30 September 1998.

<sup>2</sup>. G. Canavan, "Considerations in Missile Reductions and De-alerting," L. Kruger ed, Proceedings of the World Federation of Scientists Working Group on Missile Proliferation and Defense (World Federation of Scientists, Lausanne, 1998); Los Alamos National Laboratory Report LA-UR-98-1426, April 1988.

<sup>3</sup>. G. Kent and R. DeValk, "Strategic Defenses and the Transition to Assured Survival," RAND Report R-3369-AF, October, 1986.

**Fig. 1. Stability indices**

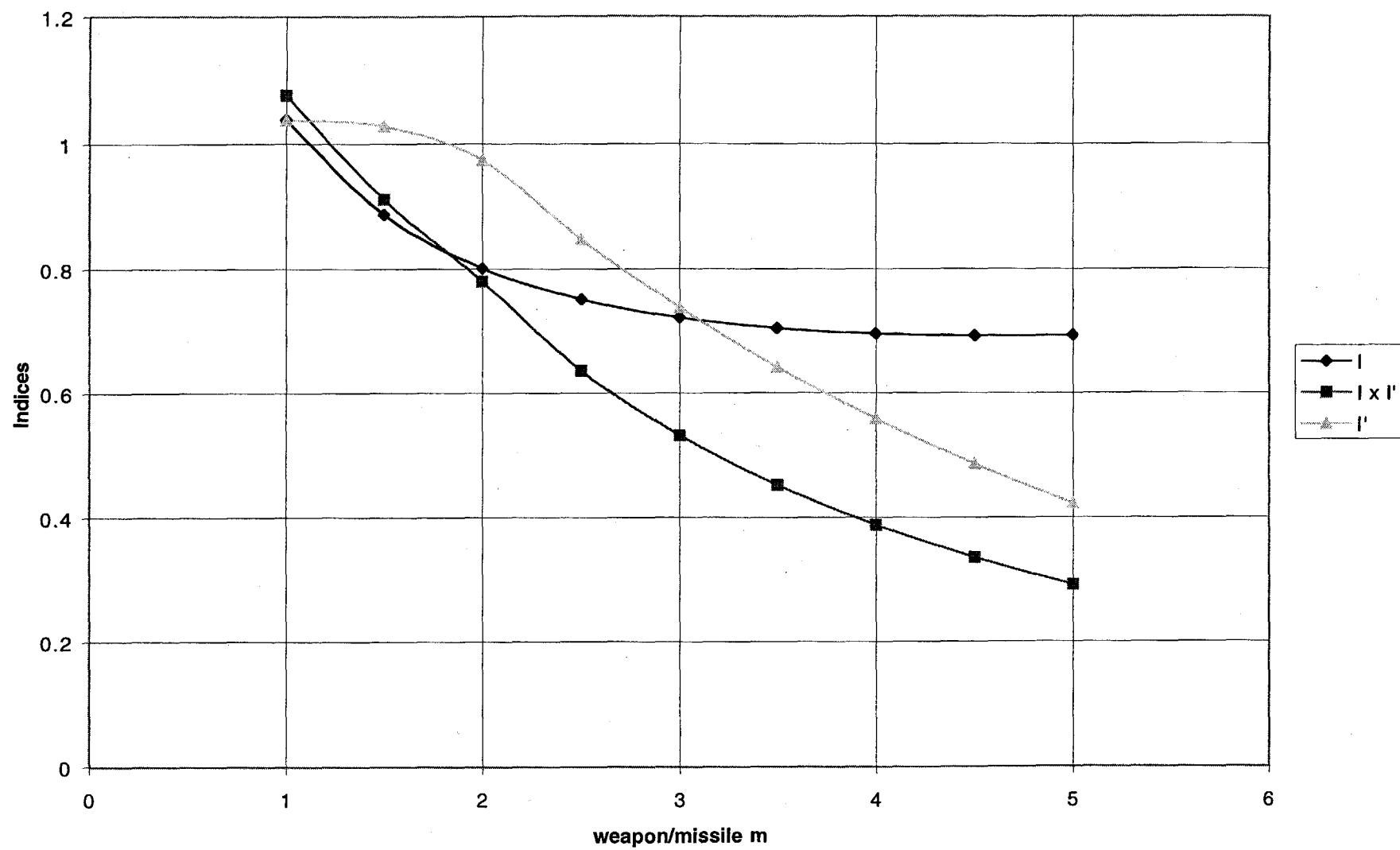


Fig. 2. EV vs weapons for  $Q = \sqrt{0.5}$

