LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-TR-737335

A Game Theoretic Model of Thermonuclear Cyberwar

B. C. Soper

August 23, 2017

Disclaimer

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

A GAME THEORETIC MODEL OF THERMONUCLEAR CYBERWAR

BRADEN C. SOPER

1. Introduction

In this paper we propose a formal game theoretic model of *thermonuclear cyberwar* based on ideas found in [1] and [2]. Our intention is that such a game will act as a first step toward building more complete formal models of Cross-Domain Deterrence (CDD). We believe the proposed thermonuclear cyberwar game is an ideal place to start on such an endeavor because the game can be fashioned in a way that is closely related to the classical models of nuclear deterrence [4–6], but with obvious modifications that will help to elucidate the complexities introduced by a second domain.

We start with the classical bimatrix nuclear deterrence game based on the game of chicken, but introduce uncertainty via a left-of-launch cyber capability that one or both players may possess. The Nash equilibria of the resulting game can be categorized based on the players' left-of-launch cyber capability. Informally that categorization is as follows:

(1) If neither player's cyber capabilities are advanced, then the resulting game is strategically equivalent to the original game of chicken.
(2) If a single player's cyber capabilities are advanced while the other player's cyber capabilities are not advanced, then the player with advanced cyber capabilities has a decisive strategic advantage in which standing firm is a strictly dominating strategy for him, and the other player is forced to submit.
(3) If both players have advanced cyber capabilities, standing firm is a strictly dominating strategy for both players.

We discuss various extensions of this simple matrix game. The first extension is to the incomplete information case where the players do not know the exact cyber capabilities of their opponents. A second extension is to consider the strategic nature of cyber intrusion and detection. Thus the strategy space of the game is increased to include how aggressively a player tries to infiltrate the opponent's nuclear command and control system as well as how sensitive a player should be in detecting possible intrusions. Finally, we examine the effects of the cyber capabilities on the players's resolve in the dynamic brinksmanship version of the nuclear deterrence game.

2. A Modified Game of Chicken

The classic game of "chicken" has been used as a simple model for nuclear deterrence and brinksmanship [4–6]. We borrow the notation and game set-up from Powell [5]. As

Date: October 18, 2017.

FIGURE 1. The Game of Chicken

		II: stand firm	II: submit
I	stand firm	d_I, d_{II}	w_I, s_{II}
	submit	s_I, w_{II}	c_I, c_{II}

noted by Powell, the game is probably too simple to shed much light on the dynamics of escalation and brinksmanship in nuclear crises. Nevertheless, it serves as a useful starting point for constructing more complex games.

The game of chicken is a 2x2 matrix game, meaning there are two players, labelled I and II, each having two strategies. The players are usually taken to be states (e.g. USA and USSR in the classic application). The states have to decide between standing firm and submitting in a nuclear stand-off. If both states stand firm, the game ends in a general nuclear exchange. Under this strategy profile, the payoff for player $i = I, II$ is d_i and it is the least desirable outcome for both players. If both players submit, the game ends in a compromise with payoff c_i for player $i = I, II$. Finally if one player stands firm and the other player submits, the game ends with the standing-firm player as the "winner" and the submitting player the "loser". Winning the standoff in this manner is the most desirable outcome for both players, while losing in this manner is undesirable but more preferred than the destruction of a general nuclear exchange. We label these outcomes w_i (if player $i = I, II$ stands firm and is the winner) and s_i (if player $i = I, II$ submits and is the loser). With these interpretations we have the ordering of payoffs as follows:

$$w_I > c_I > s_I > d_I,$$
$$w_{II} > c_{II} > s_{II} > d_{II}.$$

In matrix form, the game is shown in Figure 1.

The game has two pure Nash equilibria, namely the two strategy profiles in which one player stands firm and the other submits. There is also one Nash equilibria in mixed strategies:

$$\text{Player } I \text{ stands firm with probability } \phi_I = \frac{w_{II} - c_{II}}{w_{II} - c_{II} + s_{II} - d_{II}},$$
$$\text{Player } II \text{ stands firm with probability } \phi_{II} = \frac{w_I - c_I}{w_I - c_I + s_I - d_I}.$$

We are interested in understanding the effects on the above nuclear deterrence (ND) game when possible cyber attacks are included. To this end we will modify the above game to include uncertainty about a left-of-launch (LOL) cyber capability as described in [2]. A successful LOL cyber attack on a nuclear command and control (C&C) system

is assumed to undermine the opponent's ability to launch a nuclear weapon. To model a LOL cyber capability we introduce the two probabilities, p_I and p_{II}, which are defined as

$$p_I = P(\text{player } I\text{'s attack will fail due to player } II \text{ LOL intervention}),$$
$$p_{II} = P(\text{player } II\text{'s attack will fail due to player } I \text{ LOL intervention}).$$

As an example of how these probabilities will affect the outcome of the game, consider the case where both players choose "stand firm" as their strategy. Interpreting this as both players launching nuclear weapons, there are four cases to consider:

(1) Both players' attacks fail due to the other players' LOL preemption.
(2) Player I's attack fails while player II's succeeds.
(3) Player II's attack fails while player I's succeeds.
(4) Neither players' attacks fail.

One way to assign costs to each of these scenarios is to interpret each of the four above outcomes as strategically equivalent to one of the four outcomes in the original game of chicken. Specifically we equate outcomes (1),(2), (3) and (4) above with the original nuclear deterrence game outcomes resulting from the strategy profiles ("submit","submit"), ("submit","stand firm"), ("stand firm","submit"), and ("stand firm","stand firm") respectively. Thus the associated payoffs with each of the above outcomes are as follows.

(1) c_I, c_{II}
(2) s_I, w_{II}
(3) w_I, s_{II}
(4) d_I, d_{II}

For notational simplicity we will define the strategy "stand firm" as a and "submit" as s. Let $u_I(x, y)$ and $u_{II}(x, y)$ be the expected payoff under strategy profile (x, y) for player's I and II respectively. Then the expected payoffs under strategy profile (a, a) can be written as follows.

$$u_I(a,a) = d_I(1-p_I)(1-p_{II}) + w_I(1-p_I)p_{II} + s_I p_I(1-p_{II}) + c_I p_I p_{II}$$
$$u_{II}(a,a) = d_{II}(1-p_I)(1-p_{II}) + w_{II}(1-p_{II})p_I + s_{II} p_{II}(1-p_I) + c_{II} p_I p_{II}$$

FIGURE 2. "Left-of-Launch" Nuclear Deterrence Game

		II: stand firm	II: submit
I	stand firm	d_I^*, d_{II}^*	w_I^*, s_{II}^*
	submit	s_I^*, w_{II}^*	c_I^*, c_{II}^*

Similar reasoning gives us the following payoffs for the two players under all other possible strategy profiles.

$$\begin{aligned}
u_I(a,s) &= w_I(1-p_I) + c_I p_I \\
u_{II}(a,s) &= s_{II}(1-p_I) + c_{II} p_I \\
u_I(s,a) &= s_I(1-p_{II}) + c_I p_{II} \\
u_{II}(s,a) &= w_{II}(1-p_{II}) + c_{II} p_{II} \\
u_I(s,s) &= c_I \\
u_{II}(s,s) &= c_{II}
\end{aligned}$$

To analyze the effects of introducing the uncertainty of LOL capabilities we introduce the payoff variables $w_I^*, c_I^*, s_I^*, d_I^*, w_{II}^*, c_{II}^*, s_{II}^*, d_{II}^*$, which we define as the corresponding payoffs in a new left-of-launch nuclear deterrence (LOLND) game.

$$\begin{aligned}
d_I^* &= u_I(a,a) = d_I(1-p_I)(1-p_{II}) + w_I(1-p_I)p_{II} + s_I p_I(1-p_{II}) + c_I p_I p_{II} \\
d_{II}^* &= u_{II}(a,a) = d_{II}(1-p_I)(1-p_{II}) + w_{II}(1-p_{II})p_I + s_{II} p_{II}(1-p_I) + c_{II} p_I p_{II} \\
w_I^* &= u_I(a,s) = w_I(1-p_I) + c_I p_I \\
s_{II}^* &= u_{II}(a,s) = s_{II}(1-p_I) + c_{II} p_I \\
s_I^* &= u_I(s,a) = s_I(1-p_{II}) + c_I p_{II} \\
w_{II}^* &= u_{II}(s,a) = w_{II}(1-p_{II}) + c_{II} p_{II} \\
c_I^* &= u_I(s,s) = c_I \\
c_{II}^* &= u_{II}(s,s) = c_{II}
\end{aligned}$$

The expected payoff matrix for the LOLND game is shown in Figure 2.

3. Game Analysis

We now investigate what happens to the new LOLND game equilibria as the probabilities p_I and p_{II} are varied. Label the original ND game payoff matrix by M and the new LOLND expected payoff matrix by M^*. First consider the extreme cases. When $p_I = p_{II} = 0$, then $M^* = M$, and we have the original game of chicken. When $p_I = p_{II} = 1$ all payoffs in M^* are (c_I, c_{II}), thus both players are indifferent to all strategies as they all lead to the same outcome. When $p_I = 0$ and $p_{II} = 1$ we get the following M^*:

		II	
		stand firm	submit
I	stand firm	w_I, s_{II}	w_I, s_{II}
	submit	c_I, c_{II}	c_I, c_{II}

In this case "stand firm" is a strictly dominating strategy for player I, while player II is indifferent to her strategy. Thus the strategy profile ("stand firm", "submit") is the unique pure Nash equilibrium while there are an infinite number of mixed strategies for player II as long as player I stands firm. An analogous result is obtained in the case $p_I = 1$ and $p_{II} = 0$.

Ignoring these extreme cases, we focus on the more interesting case that both $0 < p_I < 1$ and $0 < p_{II} < 1$. The values of p_I and p_{II} are fixed parameters of the game, but for the purposes of analysis we consider them to be continuous, real variables in the open interval $(0, 1)$. Thus the payoffs $w_I^*, c_I^*, s_I^*, d_I^*, w_{II}^*, c_{II}^*, s_{II}^*, d_{II}^*$ are all differentiable, linear functions of p_I and p_{II}. Note that as $p_I \downarrow 0$ and $p_{II} \downarrow 0$ we have $M^* \to M$. Thus for small enough p_I and p_{II}, M^* is strategically equivalent to M, which we write as $M^* \sim M$. Intuitively this makes sense, since if the probability of a successful LOL preemption is low, it should have little effect on the game.

We begin by assuming that both p_I and p_{II} are sufficiently small so that $M^* \sim M$. Keeping p_I fixed and increasing p_{II} we observe several different regimes for the game parameters resulting in different Nash equilibria.

Differentiating the payoffs in M^* with respect to p_{II} gives us

$$\begin{aligned}
\frac{\partial d_I^*}{\partial p_{II}} &= (w_I - d_I)(1 - p_I) + (c_I - s_I)p_I, \\
\frac{\partial w_I^*}{\partial p_{II}} &= 0, \\
\frac{\partial s_I^*}{\partial p_{II}} &= c_I - s_I, \\
\frac{\partial c_I^*}{\partial p_{II}} &= 0, \\
\frac{\partial d_{II}^*}{\partial p_{II}} &= (s_{II} - d_{II})(1 - p_I) + (c_{II} - w_{II})p_I, \\
\frac{\partial w_{II}^*}{\partial p_{II}} &= c_{II} - w_{II}, \\
\frac{\partial s_{II}^*}{\partial p_{II}} &= 0, \\
\frac{\partial c_{II}^*}{\partial p_{II}} &= 0.
\end{aligned}$$

Given the relations on the payoffs in M, the above implies $\frac{\partial d_I^*}{\partial p_{II}} > 0$, $\frac{\partial s_I^*}{\partial p_{II}} > 0$ and $\frac{\partial w_{II}^*}{\partial p_{II}} < 0$. The sign of $\frac{\partial d_{II}^*}{\partial p_{II}}$ is ambiguous. Solving $\frac{\partial d_{II}^*}{\partial p_{II}} = 0$ and defining $p_I^* = \frac{s_{II} - d_{II}}{s_{II} - d_{II} + w_{II} - c_{II}}$ we see that

$$\begin{aligned}
\frac{\partial d_{II}^*}{\partial p_{II}} &> 0 \text{ if } p_I < p_I^*, \\
\frac{\partial d_{II}^*}{\partial p_{II}} &= 0 \text{ if } p_I = p_I^*, \\
\frac{\partial d_{II}^*}{\partial p_{II}} &< 0 \text{ if } p_I > p_I^*.
\end{aligned}$$

Recall we started assuming that p_I is "small enough". Making this more precise we assume $p_I < p_I^*$, from which we have $\frac{\partial d_{II}^*}{\partial p_{II}} > 0$.

With both p_I and p_{II} close enough to zero we can insure the relations

$$\begin{aligned}
w_I^* > c_I^* > s_I^* > d_I^*, \\
w_{II}^* > c_{II}^* > s_{II}^* > d_{II}^*,
\end{aligned}$$

both hold. Then as we increase p_{II} we will observe the following shifts in these relations.

$$\begin{aligned}
\lim_{p_{II}\uparrow 1} d_I^* &= w_I(1-p_I) + c_I p_I = w_I^*, \\
\lim_{p_{II}\uparrow 1} s_I^* &= s_I < c_I = c_I^*, \\
\lim_{p_{II}\uparrow 1} d_{II}^* &= s_{II}(1-p_I) + c_{II} p_I = s_{II}^*, \\
\lim_{p_{II}\uparrow 1} w_{II}^* &= c_{II} = c_{II}^*,
\end{aligned}$$

Given these limits and signs of derivatives, we see that with p_I fixed small, and increasing p_{II} we pass through several regimes. First note that the order $w_{II}^* > c_{II}^* > s_{II}^* > d_{II}^*$ is fixed in the limit. Thus we only need to consider the order of player I's payoffs. The two distinct regimes are defined by the relations $s_I^* > d_I^*$ and $d_I^* > s_I^*$. By symmetry the same analysis can be applied to the payoffs when p_{II} is held fixed and p_I increases. Thus there are two regimes for player II's payoffs as p_I varies defined by the relations $s_{II}^* > d_{II}^*$ and $d_{II}^* > s_{II}^*$.

With these observations it is possible to characterize all Nash equilibria in the LOLND game, which is done in Proposition 1 below. We first define the following values:

$$\begin{aligned}
p_I^* &= \frac{s_{II} - d_{II}}{s_{II} - d_{II} + w_{II} - c_{II}}, \\
p_{II}^* &= \frac{s_I - d_I}{s_I - d_I + w_I - c_I}, \\
\phi_I^* &= \frac{w_{II}^* - c_{II}^*}{(w_{II}^* - c_{II}^*) + (s_{II}^* - d_{II}^*)}, \\
\phi_{II}^* &= \frac{w_I^* - c_I^*}{(w_I^* - c_I^*) + (s_I^* - d_I^*)}.
\end{aligned}$$

The Nash equilibria of the game M^* can be defined in terms of $p_I^*, p_{II}^*, \phi_I^*, \phi_{II}^*$.

Proposition 1. Assuming $p_I, p_{II} \in (0,1)$, the Nash equilibria in the LOLND game M^* are as follows.

(1) If $p_I \leq p_I^*$ and $p_{II} \leq p_{II}^*$, then $M \sim M^*$ and the Nash equilibria are analogous to the ND game. Namely there are two pure Nash equilibria, ("stand firm", "submit") and ("submit", "stand firm"), and one mixed equilibria given by player i standing firm w.p. ϕ_i^* for $i = I, II$. [1]
(2) If $p_I < p_I^*$ and $p_{II} > p_{II}^*$, then ("stand firm", "submit") is the unique Nash equilibria.
(3) If $p_I > p_I^*$ and $p_{II} < p_{II}^*$, then ("submit", "stand firm") is the unique Nash equilibria.
(4) If $p_i > p_i^*$ for $i = I, II$, ("stand firm", "stand firm") is the unique Nash equilibria.

[1] Note that in the degenerate case where both inequalities obtain, the mixed equilibria is actually pure with both players standing firm w.p. 1. If one inequality obtains, then one player plays "stand firm" w.p. 1 in the mixed equilibria.

(5) If $p_i = p_i^*$ and $p_j > p_j^*$, then for any $\phi \in [0,1]$ there is a mixed equilibria given by player i standing firm w.p. 1 and player j standing firm w.p. ϕ for $i \neq j$ and $i, j \in \{I, II\}$.

Proof: Fix $p_I < p_I^*$. This implies $\frac{\partial d_{II}^*}{\partial p_{II}} > 0$, from which we have

$$s_{II}(1-p_I) + c_{II}p_I > d_{II}(1-p_I) + w_{II}p_I \implies s_{II}^* > d_{II}(1-p_I) + w_{II}p_I.$$

The strict monotonicity of d_{II}^* in p_{II}^* implies that $d_{II}^* \in (\lim_{p_{II}\downarrow 0} d_{II}^*, \lim_{p_{II}\uparrow 1} d_{II}^*)$. Taking limits in p_{II} we have

$$\lim_{p_{II}\downarrow 0} d_{II}^* = d_{II}(1-p_I) + w_{II}p_I,$$
$$\lim_{p_{II}\uparrow 1} d_{II}^* = s_{II}^*.$$

It follows that $d_{II}^* \in (d_{II}(1-p_I) + w_{II}p_I, s_{II}^*)$, giving us $d_{II}^* < s_{II}^*$.

Conversely if $p_I > p_I^*$ we have $\frac{\partial d_{II}^*}{\partial p_{II}} < 0$ and $s_{II}^* < d_{II}(1-p_I) + w_{II}p_I$. Again the strict monotonicity implies $d_{II}^* \in (\lim_{p_{II}\uparrow 1} d_{II}^*, \lim_{p_{II}\downarrow 0} d_{II}^*)$. From this we have $d_{II}^* \in (s_{II}^*, d_{II}(1-p_I) + w_{II}p_I)$, giving us $d_{II}^* > s_{II}^*$. Finally we consider the marginal case $p_I = p_I^*$. In this case $\frac{\partial d_{II}^*}{\partial p_{II}} = 0$, which implies $s_{II}^* = d_{II}(1-p_I) + w_{II}p_I$. Since d_{II}^* is constant we must have $d_{II}^* = s_{II}^*$.

By symmetry, analogous arguments show that

$$p_{II} < p_{II}^* \implies d_I^* < s_I^*,$$
$$p_{II} = p_{II}^* \implies d_I^* = s_I^*,$$
$$p_{II} > p_{II}^* \implies d_I^* > s_I^*.$$

Combing the above results we have the following, in which equalities imply equalities and inequalities imply inequalities:

$$p_I \leq p_I^* \text{ and } p_{II} \leq p_{II}^* \implies d_{II}^* \leq s_{II}^* \text{ and } d_I^* \leq s_I^*,$$
$$p_I \leq p_I^* \text{ and } p_{II} \geq p_{II}^* \implies d_{II}^* \leq s_{II}^* \text{ and } d_I^* \geq s_I^*,$$
$$p_I \geq p_I^* \text{ and } p_{II} \leq p_{II}^* \implies d_{II}^* \geq s_{II}^* \text{ and } d_I^* \leq s_I^*,$$
$$p_I \geq p_I^* \text{ and } p_{II} \geq p_{II}^* \implies d_{II}^* \geq s_{II}^* \text{ and } d_I^* \geq s_I^*.$$

Since the relations $w_I^* > c_I^* > s_I^*$ are maintained in all cases, the result follows from standard Nash equilibrium analysis of the game matrix M^*. □

Note the relationship between the threshold LOL probabilities and the mixed equilibria of the original nuclear deterrence game. If ϕ_I and ϕ_{II} are the equilibria probabilities of

players I and II standing firm, respectively, then we have the following relation:

$$p_I^* = 1 - \phi_I,$$
$$p_{II}^* = 1 - \phi_{II}.$$

This relation is perhaps not surprising since $1 - \phi_i$ is the mixed equilibria probability of player i submitting and p_I is the probability of player I's attack failing, which we have interpreted as being forced to submit. Furthermore since mixed Nash equilibria make opponents indifferent to their strategies it becomes clear why $p_I = p_I^* = 1 - \phi_I$ implies player II becomes indifferent when player I stands firm.

4. Discussion

If we look at the distinct Nash equilibria regimes in Proposition 1, we can gain some insight into the effects of cyber capabilities on the simple nuclear deterrence game. In case (1) both probabilities of successful LOL are lower than the thresholds p_I^* and p_{II}^*. As such there is not a significant effect on the original game. Even though payoffs are altered, the characterization of Nash equilibria is not changed. When one or both of p_I and p_{II} reach the thresholds p_I^* and p_{II}^*, the mixed equilibria changes in character. Whichever player obtains $p_i = p_i^*$ will play "stand firm" with probability 1 in the mixed equilibria.

Recall that p_i is the probability that player i's command and control (C&C) system will fail due to his opponent's LOL capabilities. Thus it is a measure of the opponent's cyber capabilities. For concreteness assume player II increases his cyber capability to the point that $p_I = p_I^*$. In this case player II is indifferent to player I's "stand firm" posture. This actually has the somewhat undesirable effect of forcing player I to "stand firm" with probability 1 in the unique mixed strategy equilibria. Thus we see the presence of increased cyber capabilities creates a more hostile posture in the mixed strategies equilibria. Note however that the LOL capabilities mitigate this increased hostility by decreasing the likelihood of successfully launching a nuclear weapon, thus increasing payoffs in expectation.

Once player II increases his cyber capabilities to the point that $p_I > p_I^*$ the nature of the game changes considerably. At this point player II has "stand firm" as a strictly dominating strategy forcing player I to submit in equilibria. Thus increasing cyber capabilities over the threshold creates a distinct strategic advantage, essentially guaranteeing a victory in the nuclear stand off.

If both players obtain significant cyber capabilities ($p_I > p_I^*$ and $p_{II} > p_{II}^*$) we find that they are no longer playing a game of chicken, but instead they are playing a game where they both have "stand firm" as a strictly domination strategy. A further distinction can be made in this regime. If cyber capabilities p_I and p_{II} satisfy the following bounds,

$$p_I^* < p_I < \frac{(c_{II} - s_{II})p_{II} + (c_{II} - d_I)(1 - p_{II})}{(c_{II} - s_{II})p_{II} + (w_{II} - d_{II})(1 - p_{II})},$$
$$p_{II}^* < p_{II} < \frac{(c_I - s_I)p_I + (c_I - d_I)(1 - p_I)}{(c_I - s_I)p_I + (w_I - d_I)(1 - p_I)},$$

then the following relations will hold for both $i = I, II$:

$$w_i^* > c_i^* > d_i^* > s_i^*.$$

This implies that the game being played is the the Prisoner's Dilemma. Thus this mid range cyber capability results in an inefficient allocation of payoffs: Both players choose to stand firm, but both could collectively improve their payouts by coordinating to both "submit".

Note that the size of the regimes depends on the payoff matrix M from the original nuclear deterrence game via the definitions of p_I^* and p_{II}^*. Rewriting the definitions and dropping subscripts we have

$$p^* = \frac{1}{1 + \frac{w-c}{s-d}}.$$

Thus we see that the relation between $w - c$ and $s - d$ determines the size of the various Nash regimes. We can interpret $w - c$ as the additional benefit from "winning" a game of chicken versus both players submitting and $s - d$ the additional benefit of submitting in the face of a challenger versus a general nuclear exchange. In the case of nuclear stand-offs one would expect $s - d >> w - c$, in which case $p^* \approx 1$. Thus the left of launch capabilities will have to be extremely robust to have any impact on the game. On the other hand if being a dominant winner is valuable enough, i.e. $s - d << w - c$ then we have $p^* \approx 0$, in which case a modest degree of LOL capabilities will transform the game into one of the non-chicken regimes.

Finally we consider the overall probability of a general nuclear exchange, which we denote by $P(GNE)$ and $P(GNE^*)$ in the ND and LOLND games respectively. In the original game this depends on the the equilibrium being played. Specifically if one of the pure Nash equilibria is being played we have $P(GNE) = 0$. If the mixed equilibria is being played, then we have $P(GNE) = \phi_I \phi_{II}$, i.e. the probability that both players stand firm.

For the LOLND game we consider each case separately. We ignore case (5) since it is degenerate. Since case (1) is strategically equivalent to the ND game, we again have $P(GNE^*) = 0$ in both pure equilibria. The mixed equilibria case, on the other hand, is not quite analogous to the ND game. In this case we have $P(GNE^*) = \phi_I^* \phi_{II}^* (1 - p_I)(1 - p_{II})$ since for a general nuclear exchange to occur we must have both players standing firm and both LOL attacks failing. From the definitions of ϕ_I^* and ϕ_{II}^* it can be shown that

$$\phi_I^* = \frac{\phi_I}{1 - p_I} \geq \phi_I,$$

$$\phi_{II}^* = \frac{\phi_{II}}{1 - p_{II}} \geq \phi_{II}.$$

It follows that $\phi_I^* \phi_{II}^* (1 - p_I)(1 - p_{II}) = \phi_I \phi_{II}$, which implies $P(GNE) = P(GNE^*)$. Thus even though LOL attacks *decrease* the probability of a successful nuclear launch, they *increase* the probability of players standing firm in such a way that the overall probability of a general nuclear exchange is unchanged.

In cases (2) and (3) there is no chance of a general nuclear exchange, thus $P(GNE^*) = 0$.

Finally in case (4) we have both players standing firm w.p. 1, giving us $P(GNE^*) = (1-p_I)(1-p_{II})$. But in this case $p_i > p_i^*$ for $i = I, II$. Thus we have

$$(1-p_I)(1-p_{II}) < (1-p_I^*)(1-p_{II}^*) = \phi_I \phi_{II}.$$

It follows that $P(GNE^*) < P(GNE)$. Thus if both players can achieve cyber capabilities that exceed the threshold, the probability of a general nuclear exchange is decreased even though both players are certain to stand firm. As noted above, in nuclear standoffs one could reasonably expect the thresholds $p^* \approx 1$, making this scenario very unlikely.

5. Game Extensions

Several extensions of the above LOLND game may be considered. We start with considering an incomplete information game where players are not aware of their opponent's cyber capabilities. We then consider several extensions where players make strategic choices regarding the LOL attack and detections. Finally we consider the effects of cyber attacks on the dynamic bargaining version of the nuclear deterrence game.

5.1. **Incomplete Information.**

Consider a Bayesian game in which the types of the players are defined in terms of LOL capabilities. We introduce alternate notation for ease of exposition. Instead of subscripts I and II we will use subscripts 1 and 2, respectively. Also, for any variable x_i associated with player i we use x_{-i} to denote the same variable associated with the opponent to player i. We assume a discrete-type game in which player i is of type $\theta_i \in \{\theta_i^-, \theta_i^+\}$ where $0 < \theta_i^- < p_{-i}^* < \theta_i^+ < 1$.

We assume players do not know the type of their opponents. As such the players do not know which game they are playing with their adversary. For example, if player 1 is of type θ_1^-, then he knows he is in one of two regimes: Regime (1) $d_2^* < s_2^*$ and $d_1^* < s_1^*$ or regime (2) $d_2^* > s_2^*$ and $d_1^* < s_1^*$, but he does not know which. Thus he is either playing a game of chicken or a game in which player 2 has a strictly dominant strategy of "stand firm". On the other hand if player 1 is of type θ_1^+, he will know that "stand firm" is a strictly dominating strategy for him, but he will not know if it is for his opponent.

Strategies in the Bayesian game are maps from types to pure strategies of the complete information game. Denote the type spaces by $\Theta_i = \{\theta_i^-, \theta_i^+\}$ and the types by $\theta_i \in \Theta_i$ for $i = 1, 2$. Then we denote incomplete information strategies by the maps $\sigma_i : \Theta_i \to \{0, 1\}$, where we have taken the pure strategy space from the ND game to be $\{0, 1\}$, where 0 denotes "stand firm" and 1 denotes "submit" . We also assume that player i is of type θ_i^+ w.p. $\pi_i \in (0, 1)$ for $i = 1, 2$. This prior distribution on player types is assumed to be common knowledge among both players.

Note that the payoffs for players are now functions of type. Specifically, for $i = 1, 2$ we have $d_i^* = d_i^*(\theta_1, \theta_2), s_i^* = s_i^*(\theta_i)$ and $w_i^* = w_i^*(\theta_{-i})$, while c_i^* is constant across types.

The following proposition characterizes the pure Bayesian equilibria in the incomplete information LOLND game. We define the parameters π_i^- for $i = 1, 2$, which will be needed

in the sequel, as follows:

$$\pi_i^- = \frac{w_i^*(\theta_{-i}^-) - c_i^*}{w_i^*(\theta_{-i}^-) - c_i^* + s_i^*(\theta_i^-) - d_i^*(\theta_1^-, \theta_2^+)}.$$

Under our assumptions on types and previous analyses, we are guaranteed for $\pi_i^- \in (0,1)$. We also define two pure Bayesian strategies, $\sigma_\emptyset$ and σ_{01}, that will characterize all possible pure Bayesian Nash equilibria. We drop the subscripts on θ and Θ to denote the possibility of defining the maps on either type space.

$$\sigma_\emptyset(\theta) = 0 \quad \text{for all } \theta \in \Theta.$$

$$\sigma_{01}(\theta) = \begin{cases} 0 & \text{if } \theta = \theta^+, \\ 1 & \text{if } \theta = \theta^-. \end{cases}$$

Proposition 2. Assuming $\pi_i \in (0,1,)$ for $i = 1,2$ the pure Bayesian Nash equilibria in the incomplete information LOLND game are as follows.

- If $\pi_i < \pi_i^-$ for both $i = 1,2$, then there are two pure Bayesian Nash equilibria:
  (1) $\sigma_1(\cdot) = \sigma_\emptyset(\cdot)$ and $\sigma_2(\cdot) = \sigma_{01}(\cdot)$,
  (2) $\sigma_1(\cdot) = \sigma_{01}(\cdot)$ and $\sigma_2(\cdot) = \sigma_\emptyset(\cdot)$.
- If $\pi_i < \pi_i^-$ and $\pi_{-i} > \pi_{-i}^-$ then there is a unique pure Bayesian Nash equilibria given by $\sigma_i(\cdot) = \sigma_{01}(\cdot)$ and $\sigma_{-i}(\cdot) = \sigma_\emptyset(\cdot)$.
- If $\pi_i > \pi_i^-$ for both $i = 1,2$, then there is a unique pure Bayesian Nash equilibria given by $\sigma_i(\cdot) = \sigma_{01}(\cdot)$ and $\sigma_{-i}(\cdot) = \sigma_{01}(\cdot)$.

Proof: Suppose $\sigma_i(\cdot) = \sigma_\emptyset(\cdot)$. Then no matter the type, or opponent strategy, player i will stand firm. We will show that the best response to this strategy is for player $-i$ to choose $\sigma_{-i}(\cdot) = \sigma_{01}(\cdot)$. Letting $x_i, x_{-i} \in \{0,1\}$ denote the strategies for players i and $i-$ respectively, where 0 denotes "stand firm" and 1 denotes "submit", the posterior expected payoff for player $-i$, denoted by $u_{-i}(x_i, x_{-i}|\theta_{-i})$ can be written as

$$\begin{aligned} &u_{-i}(x_i, x_{-i}|\theta_{-i}) = \\ &\pi_i\left((1-x_{-i})(1-x_i)d_{-i}^*(\theta_{-i}, \theta_i^+) + (1-x_{-i})x_i w_{-i}^*(\theta_i^+) + x_{-i}(1-x_i)s_{-i}^*(\theta_{-i}) + x_{-i}x_i c_{-i}^*\right) \\ &+(1-\pi_i)\left((1-x_{-i})(1-x_i)d_{-i}^*(\theta_{-i}, \theta_i^-) + (1-x_{-i})x_i w_{-i}^*(\theta_i^-) + x_{-i}(1-x_i)s_{-i}^*(\theta_{-i}) + x_{-i}x_i c_{-i}^*\right) \end{aligned}$$

Since $\sigma_i(\cdot) = \sigma_\emptyset(\cdot)$ we need only consider $x_i = 0$. In this case we have

$$\begin{aligned} u_{-i}(0, x_{-i}|\theta_{-i}) = \pi_i&\left((1-x_{-i})d_{-i}^*(\theta_{-i}, \theta_i^+) + x_{-i}s_{-i}^*(\theta_{-i})\right) \\ &+ (1-\pi_i)\left((1-x_{-i})d_{-i}^*(\theta_{-i}, \theta_i^-) + x_{-i}s_{-i}^*(\theta_{-i})\right) \end{aligned}$$

From earlier analysis we know

$$\begin{aligned} \theta_{-i} = \theta_{-i}^- &\implies s_{-i}^*(\theta_{-i}) > d_{-i}^*(\theta_{-i}, \theta_i^+) > d_{-i}^*(\theta_{-i}, \theta_i^-), \\ \theta_{-i} = \theta_{-i}^+ &\implies s_{-i}^*(\theta_{-i}) < d_{-i}^*(\theta_{-i}, \theta_i^+) < d_{-i}^*(\theta_{-i}, \theta_i^-). \end{aligned}$$

From this it follows that

$$0 = \arg\max_{x_{-i}\in\{0,1\}} u_{-i}(0, x_{-i}|\theta^-_{-i}),$$
$$1 = \arg\max_{x_{-i}\in\{0,1\}} u_{-i}(0, x_{-i}|\theta^+_{-i}).$$

This is equivalent to $\sigma_{01}(\cdot)$ being a best response to the strategy $\sigma_{\emptyset}(\cdot)$ in the incomplete information LOLND game.

We now consider each case separately. First suppose $\pi_i < \pi^-_i$ for both $i = 1, 2$. Furthermore suppose $\sigma_i(\cdot) = \sigma_{01}(\cdot)$. It suffices to show that in this case $\sigma_{-i}(\cdot) = \sigma_{\emptyset}(\cdot)$ is a best response. Since $\sigma_i(\cdot) = \sigma_{01}(\cdot)$ the posterior expected payoff for player $-i$ can be written as

$$u_{-i}(\sigma_{01}(\cdot), x_{-i}|\theta_{-i}) = \pi_i\left((1 - x_{-i})d^*_{-i}(\theta_{-i}, \theta^+_i) + x_{-i}s^*_{-i}(\theta_{-i})\right) + (1 - \pi_i)\left((1 - x_{-i})w^*_{-i}(\theta^-_i) + x_{-i}c^*_{-i}\right).$$

Recall that $w^*_{-i}(\theta_i) > c^*_{-i}$ for any θ_i. If $\theta_{-i} = \theta^+_{-i}$ then $d^*_{-i}(\theta_{-i}, \theta^+_i) > s^*_{-i}(\theta_{-i})$. Thus

$$0 = \arg\max_{x_{-i}\in\{0,1\}} u_{-i}(\sigma_{01}(\cdot), x_{-i}|\theta^+_{-i}).$$

On the other hand, if $\theta_{-i} = \theta^-_{-i}$, then $\pi_i < \pi^-_i$ implies $u_{-i}(\sigma_{01}(\cdot), 0|\theta^-_{-i}) > u_{-i}(\sigma_{01}(\cdot), 1|\theta^-_{-i})$, giving us

$$0 = \arg\max_{x_{-i}\in\{0,1\}} u_{-i}(\sigma_{01}(\cdot), x_{-i}|\theta^-_{-i}).$$

Thus $\sigma_{-i}(\cdot) = \sigma_{\emptyset}(\cdot)$ is a best response to $\sigma_i(\cdot) = \sigma_{01}(\cdot)$. By symmetry$\sigma_i(\cdot) = \sigma_{\emptyset}(\cdot)$ is a best response to $\sigma_{-i}(\cdot) = \sigma_{01}(\cdot)$ since $\pi_{-i} < \pi^-_{-i}$.

Now suppose $\pi_i < \pi^-_i$ and $\pi_{-i} > \pi^-_{-i}$. In this case the same analysis as above applies to proving that $\sigma_i(\cdot) = \sigma_{01}(\cdot)$ and $\sigma_{-i}(\cdot) = \sigma_{\emptyset}(\cdot)$ is a pure strategy Bayesian equilibrium. However, the analysis for the second equilibria is no longer valid. In this case player i's best response to $\sigma_{01}(\cdot)$ is $\sigma_{01}(\cdot)$, not $\sigma_{\emptyset}(\cdot)$.

Finally consider the case that $\pi_i > \pi^-_i$ for both $i = 1, 2$. Again, following the same line of reasoning as above we can arrive at the answer. The only difference in the analysis is that when $\theta_{-i} = \theta^-_{-i}$ we have $\pi_i < \pi^-_i$ implies $u_{-i}(\sigma_{01}(\cdot), 0|\theta^-_{-i}) < u_{-i}(\sigma_{01}(\cdot), 1|\theta^-_{-i})$, giving us

$$1 = \arg\max_{x_{-i}\in\{0,1\}} u_{-i}(\sigma_{01}(\cdot), x_{-i}|\theta^-_{-i}).$$

The result follows.

Because the type space and number of players is small, there are not too many pure strategies in the incomplete information game. Furthermore, since any best response strategy must satisfy $\theta^+ \to 0$, the number of feasible best response strategies is cut in half. Thus one can check all other feasible pure strategy profiles to conclude the above pure equilibria are unique. □

Note that while it is possible to derive mixed Bayesian equilibria in the incomplete information LOLND game, the formula are quite complex and difficult to interpret. Thus we omit them here.

FIGURE 3. Dis-coordination Game

		II: don't raise alarm	II: do raise alarm
I	don't intrude	A_I, A_{II}	B_I, C_{II}
	do intrude	C_I, B_{II}	D_I, D_{II}

5.2. **Strategic Left-of-launch.** In CDD there are strategic actions associated with each domain. Thus we should explicitly model the strategic aspects of using the LOL cyber capability. To introduce such a feature we again start with a simple model, and extend it in a straight forward way. Consider again the LOLND game above. We will incorporate strategic actions that will determine the values p_I and p_{II}.

5.2.1. *Simple Dis-coordination Game.* A simple two-stage game we can consider is the following. Model the detection game as a dis-coordination game, similar to matching pennies. In matrix form the game is shown in Fig. 3. The game of matching pennies has the following orderings of payoffs:

$$\begin{aligned} C_I &> A_I, \\ B_I &> D_I, \\ C_{II} &< A_{II}, \\ B_{II} &< D_{II}. \end{aligned}$$

This bimatrix game is the simplest game that can capture the strategic nature of an intruder and a defender trying to detect the presence of the intruder. The intruder must decide whether or not to intrude while the defender must decide whether to raise and alarm or not. The intruder clearly prefers to either intrude and not get caught ("intrude", "don't raise alarm") or to not intrude when a defender raises an alarm ("don't intrude", "raise alarm"). On the other hand the defender prefers to avoid these exact scenarios which correspond to a false negative and false positive, respectively.

There are no pure Nash equilibria in this game. Let ρ_i be the probability that player i plays "do". Then the unique mixed Nash equilibria are given by the two probabilities

$$\begin{aligned} \rho_I &= \frac{1}{1 + \frac{D_{II} - B_{II}}{A_{II} - C_{II}}}, \\ \rho_{II} &= \frac{1}{1 + \frac{B_I - D_I}{C_I - A_I}}. \end{aligned}$$

Recall that by definition, p_I is a false negative in the LOLND game. Thus we have

$$p_{II} = \rho_I(1-\rho_{II}) = \left(\frac{1}{1+\frac{D_{II}-B_{II}}{A_{II}-C_{II}}}\right)\left(1-\frac{1}{1+\frac{B_I-D_I}{C_I-A_I}}\right) = \left(\frac{1}{1+\frac{D_{II}-B_{II}}{A_{II}-C_{II}}}\right)\left(\frac{1}{1+\frac{C_I-A_I}{B_I-D_I}}\right).$$

Reversing the roles of intruder and defender gives us an analogous game but with different parameters, say $\tilde{A}_i, \tilde{B}_i, \tilde{C}_i, \tilde{D}_i$ for $i = I, II$. Similar reasoning gives the mixing probabilities $\tilde{\rho}_i$ for $i = I, II$ which we can relate to the parameter p_I.

$$p_I = \tilde{\rho}_{II}(1-\tilde{\rho}_I) = \left(\frac{1}{1+\frac{\tilde{D}_I-\tilde{B}_I}{\tilde{A}_I-\tilde{C}_I}}\right)\left(\frac{1}{1+\frac{\tilde{C}_{II}-\tilde{A}_{II}}{\tilde{B}_{II}-\tilde{D}_{II}}}\right).$$

We have thus related the LOL cyber capability parameters p_I and p_{II} with the strategic actions of attacking and defending nuclear C&C systems. Further analysis of this relationship will be taken up at a later time. In particular we would like to investigate an incomplete information version and a dynamic brinksmanship version of this game.

5.2.2. *Fixed N Observations.* In this section we consider a more involved strategic model of C&C intrusion detection based on making a fixed number of observations of the C&C system on which the defender will make their decision as to whether it is compromised or not. The defender wishes to avoid detection thus must decide how aggressively to attack the C&C system. We assume that whether a player is able to infiltrate the other player's cyber command and control system is random. Thus we define

$$X_I = \mathbb{1}\{\text{ player } I \text{ infiltrates player } II\text{'s C\&C }\} \sim \text{Bernoulli}(\theta_1),$$
$$X_{II} = \mathbb{1}\{\text{ player } I \text{ infiltrates player } II\text{'s C\&C }\} \sim \text{Bernoulli}(\theta_2),$$

for some given parameters $\theta_1, \theta_2 \in [0,1]$. If a player is successful in infiltrating their opponents C&C, they will make a strategic decision as to how aggressively they interfere with the system. We model this by the strategic parameter $\alpha_I, \alpha_{II} \in [0,1]$, which we interpret as the probability of successfully stopping a C&C operation. We further assume there is some natural C&C failure rates $\beta_I, \beta_{II} \in [0,1)$ that do not depend on any interference by an adversary, i.e. this is a C&C failure rate independent of all other variables (strategic or stochastic) in the game. Note that if there is a successful infiltration then the overall probability of failure is $1-(1-\beta_I)(1-\alpha_{II})$ and $1-(1-\beta_{II})(1-\alpha_I)$ for player's I and II respectively. If there is no infiltration then the failure probability is simply the natural failure rates β_I, β_{II}.

Each player makes $N > 1$ observations of C&C operations. Each operation either fails or succeeds. Based on these observations the players must decide whether or not their C&C systems are compromised. The detection strategy of each player will be a threshold value $t_I, t_{II} \in \{0, 1, ..., N\}$. If m_I C&C operations fail for player I, then player I's observed failure rate is $\frac{m_I}{N}$. If $m_I \geq t_I$ then player I raises an alarm and decides his system is compromised. Similarly if $m_{II} \geq t_{II}$, then player II raises an alarm and decides his system is compromised. If a player detects an infiltration, we assume appropriate measures are taken to remove the adversaries C&C access.

Assuming each C&C operation potential failure is i.i.d., the above detection problem is equivalent to a simple hypothesis testing problem. Namely

$$H_0 : m_I \sim \text{Binomial}(N, \beta_I) \text{ w.p. } 1 - \theta_{II}$$
$$H_1 : m_I \sim \text{Binomial}(N, 1 - (1 - \beta_I)(1 - \alpha_{II})) \text{ w.p. } \theta_{II}.$$

We denote the cumulative distribution functions for a random variable $X \sim \text{Binomial}(n, \theta)$ by $F_{n,\theta}(\cdot)$, i.e.

$$F_{n,\theta}(x) = \sum_{i=1}^{\lfloor x \rfloor} \left(\begin{array}{c} n \\ i \end{array} \right) \theta^i (1 - \theta)^{n-i}.$$

We assume the players want to minimize the costs from false negatives and false positives in their detections. We derive the expected cost for player I. Player II's is analogous. Let $P(\text{FN})$ denote the probability of a false negative and $P(\text{FP})$ denote the probability of a false positive. Furthermore let c_{FN} and c_{FP} be the costs associated with a false negative and a false positive respectively. Then the expected cost associated with the detection for player I, which we denote by D_I, is

$$D_I(\alpha_I, \alpha_{II}, t_I, t_{II}) = c_{\text{FN}} P(\text{FN}) + c_{\text{FP}} P(\text{FP}).$$

The probability of false negatives and false positives can be computed given the above modeling assumptions.

$$\begin{aligned} P(\text{FN}) &= P(\text{ no detection , successful infiltration }) \\ &= P\left(m_I < t_I, X_{II} = 1\right) \\ &= P\left(m_I < t_I \middle| X_{II} = 1\right) P\left(X_{II} = 1\right) \\ &= F_{N,1-(1-\beta_I)(1-\alpha_{II})}(t_I)\theta_{II} \end{aligned}$$

$$\begin{aligned} P(\text{FP}) &= P(\text{ detection , unsuccessful infiltration }) \\ &= P\left(m_I \geq t_I, X_{II} = 0\right) \\ &= P\left(m_I \geq t_I \middle| X_{II} = 0\right) P\left(X_{II} = 0\right) \\ &= \left(1 - F_{N,\beta_I}(t_I)\right)(1 - \theta_{II}) \end{aligned}$$

Thus we arrive at the following expected cost function for player I:

$$D_I(\alpha_{II}, t_I) = c_{\text{FN}} F_{N,1-(1-\beta_I)(1-\alpha_{II})}(t_I)\theta_{II} + c_{\text{FP}} \left(1 - F_{N,\beta_I}(t_I)\right)(1 - \theta_{II}).$$

With this functional form for the probabilities of false negatives and false positives, we can augment the previous matrix game to include the larger strategy space

$$\mathcal{A} = [0, 1] \times \{0, 1, 2, ..., N\} \times \{\text{submit, stand firm}\}.$$

Thus a strategy is a triple (α, t, s) where α is a measure of how aggressively a player manipulates an opponents C&C operations, t is a measure of how sensitive a player is to detecting anomalous failures and s is the classic chicken strategy.

We would now like to relate the parameter p to the strategic variables α and t. Note that in order for the intruder to have any chance at utilizing his cyber LOL capability, the detector must experience a false negative in its observations of the previous CC actions. Assuming the final game of chicken is played after the detection game, the probability of a successful LOL attack will be

$$p_{II}(\alpha_I, t_{II}) = (1 - (1 - \beta_{II})(1 - \alpha_I))F_{N,1-(1-\beta_{II})(1-\alpha_I)}(t_{II})\theta_I,$$
$$p_I(\alpha_{II}, t_I) = (1 - (1 - \beta_I)(1 - \alpha_{II}))F_{N,1-(1-\beta_I)(1-\alpha_{II})}(t_I)\theta_{II}.$$

The probabilities of LOL failure are now functions of players strategies α and t. Recall the outcomes from the matrix game M^* were denoted $u_I(\cdot)$ and $u_{II}(\cdot)$. These values are now functions of the entire strategy profile $(\alpha_I, t_I, s_I, \alpha_{II}, t_{II}, s_{II})$. Combing the detection game and the nuclear deterrence game we can write the complete expected utility for each player, denoted by $U_I((\alpha_I, t_I, s_I, \alpha_{II}, t_{II}, s_{II}))$ and $U_{II}((\alpha_I, t_I, s_I, \alpha_{II}, t_{II}, s_{II}))$.

$$U_I(\alpha_I, t_I, s_I, \alpha_{II}, t_{II}, s_{II}) = u_I(\alpha_I, t_I, s_I, \alpha_{II}, t_{II}, s_{II}) - D_I(\alpha_{II}, t_I)$$
$$U_{II}(\alpha_I, t_I, s_I, \alpha_{II}, t_{II}, s_{II}) = u_{II}(\alpha_I, t_I, s_I, \alpha_{II}, t_{II}, s_{II}) - D_{II}(\alpha_I, t_{II})$$

We will consider a mixed strategy to be a profile $\sigma_I = (\alpha_I, q_I^1, q_I^2, ..., q_I^{N-1}, r_I^1) \in [0,1]^{N+1}$. Here α_I is the pure strategy for intruder aggressiveness. We need not consider a mixed strategy for this strategy since it is already in the unit interval. Standard equilibrium results will apply. It will also simplify our search for mixed equilibria to do this. The parameters $q_I^1, q_I^2, ..., q_I^{N-1}$ are probabilities on the set $\{1, 2, .., N-1\}$ and define the mixed strategy on threshold t_I. To complete this mixed strategy we define $q_I^N = 1 - \sum_i^{N-1} q_I^i$. Finally we have r_I^1 as the probability of playing "stand firm". Again to complete the mixed strategy we note that the probability of playing "submit" is $r_I^2 = 1 - r_I^1$. The mixed strategy for player II, $\sigma_{II} = (\alpha_{II}, q_{II}^1, q_{II}^2, ..., q_{II}^{N-1}, r_{II}^1) \in [0,1]^{N+1}$, is defined analogously.

To simplify the notation we will consider the pure strategies from the matrix game M^* to be the set $\{1, 2\}$ for both players. Thus playing "submit" is equivalent to playing 1 and playing "stand firm" is equivalent to playing 2. The expected utilities given a mixed strategy profile $\sigma = (\sigma_I, \sigma_{II})$ can then be written as follows:

$$U_I(\sigma) = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{2}\sum_{\ell=1}^{2}\left(u_I(\alpha_I, i, k, \alpha_{II}, j, \ell) - D_I(\alpha_{II}, i)\right) q_I^i q_{II}^j r_I^k r_{II}^\ell$$

$$U_{II}(\sigma) = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{2}\sum_{\ell=1}^{2}\left(u_{II}(\alpha_I, i, k, \alpha_{II}, j, \ell) - D_{II}(\alpha_I, j)\right) q_I^i q_{II}^j r_I^k r_{II}^\ell.$$

The following propositions establish the existence of mixed Nash equilibria in two versions of the game.

Proposition 2: There exists a Nash equilibria in mixed strategies for the fixed-N-observation, strategic left-of-launch, nuclear deterrence game.

Proof: This follows directly from Glicksberg's theorem for infinite games. See [3].

□

Proposition 3: There exists a Nash equilibria with pure strategies for α and mixed strategies for t and s for the fixed-N-observation, strategic left-of-launch, nuclear deterrence game.

Proof: This follows indirectly from another of Glicksberg's theorems also found in [3]. First note that the strategy space is $[0,1]^{N+1}$, which is a nonempty, compact, convex subset of $\mathbb{R}^{N+1}$. Furthermore, the utility functions are continuous in all strategies when we consider mixed strategies for s and t. Since the payoffs are linear (thus quasi-concave) in the mixing strategies, all that is left to prove is that U_i is quasi-concave in the α_i for $i = I, II$. First note that

$$\frac{\partial F_{n,\phi}(x)}{\partial \phi} = -(n - \lfloor x \rfloor)\left(\begin{array}{c} n \\ \lfloor x \rfloor \end{array} \right)\phi^{\lfloor x \rfloor}(1-\phi)^{n-\lfloor x \rfloor - 1} \leq 0.$$

Setting $\phi = 1 - (1-\beta_{II})(1-\alpha_I)$ we then have

$$\begin{aligned}
\frac{\partial p_{II}}{\partial \alpha_I} &= \frac{\partial}{\partial \alpha_I}\left[(1-(1-\beta_{II})(1-\alpha_I))F_{N,\phi}(t_{II})\theta_I\right] \\
&= \frac{\partial (1-(1-\beta_{II})(1-\alpha_I))}{\partial \alpha_I}F_{N,\phi}(t_{II})\theta_I + \alpha_I \frac{\partial}{\partial \alpha_I}\left[F_{N,\phi}(t_{II})\theta_I\right] \\
&= (1-\beta_{II})F_{N,\phi}(t_{II})\theta_I + \alpha_I \frac{\partial}{\partial \alpha_I}\left[F_{N,\phi}(t_{II})\theta_I\right] \\
&= (1-\beta_{II})F_{N,\phi}(t_{II})\theta_I + \frac{\partial F_{N,\phi}(t_{II})}{\partial \phi}\frac{\partial \phi}{\partial \alpha_I}\theta_I \\
&= (1-\beta_{II})F_{N,\phi}(t_{II})\theta_I + \frac{\partial F_{N,\phi}(t_{II})}{\partial \phi}(1-\beta_{II})\theta_I \\
&= (1-\beta_{II})F_{N,\phi}(t_{II})\theta_I - (N - \lfloor t_{II} \rfloor)\left(\begin{array}{c} N \\ \lfloor t_{II} \rfloor \end{array} \right)\phi^{\lfloor t_{II} \rfloor}(1-\phi)^{N-\lfloor t_{II} \rfloor - 1}(1-\beta_{II})\theta_I.
\end{aligned}$$

Define the functions $r(\phi)$ and $\ell(\phi)$ as follows:

$$\begin{aligned}
r(\phi) &= (N - \lfloor t_{II} \rfloor)\left(\begin{array}{c} N \\ \lfloor t_{II} \rfloor \end{array} \right)\phi^{\lfloor t_{II} \rfloor}(1-\phi)^{N-\lfloor t_{II} \rfloor - 1}, \\
\ell(\phi) &= \sum_{i=0}^{\lfloor t_{II} \rfloor}\left(\begin{array}{c} N \\ i \end{array} \right)\phi^{i}(1-\phi)^{N-i}.
\end{aligned}$$

With these definitions we have

$$\operatorname{sign}\left(\frac{\partial p_{II}}{\partial \alpha_I}\right) = \operatorname{sign}\Big(\ell(\phi) - r(\phi)\Big)$$

Note the following properties:

(1) $r(\phi) = \ell'(\phi) \geq 0$
(2) $r(\phi) = 0 \iff \phi \in \{0, 1\}$
(3) $r'(\phi) = \ell''(\phi) = 0 \iff \phi = \frac{\lfloor t_{II} \rfloor}{N-1}$
(4) $r(0) = \ell(0)$

For $t_{II} < 1$ or $t_{II} = N - 1$ we can show that $r(\phi) > \ell(\phi)$ and $r(\phi) < \ell(\phi)$, respectively, in which case $\frac{\partial p_{II}}{\partial \alpha_I} > 0$ and $\frac{\partial p_{II}}{\partial \alpha_I} < 0$, respectively. In these cases, p_{II} is monotonic in α_I, hence it is quasi-concave. The other cases require a little work, but it is possible to show that for $t_{II} \in [1, N)$, there is at most one value ϕ^* such that $r(\phi^*) = \ell(\phi^*)$. This fact, combined with the above properties (1)-(4), implies p_{II} is quasi-concave in α_I. Since U_I only depends on α_I through u_I and u_I is non-decreasing in p_{II} we must have

$$\operatorname{sign}\left(\frac{\partial U_I}{\partial \alpha_I}\right) = \operatorname{sign}\left(\frac{\partial p_{II}}{\partial \alpha_I}\right).$$

It follows that U_I is quasi-concave in α_I. An analogous result can be obtained for player II giving us the desired result.

□

5.2.3. *Approximate Game.* Due to the larger strategy space in this game, characterizing the equilibria is much more difficult. One alternative representation of the game that may be more tractable is to use the normal distribution approximation to the Binomial distribution for the C&C failure process. If N is large enough and the failure rate for the system is π, the distribution of failures can be approximated by a normal distribution with mean $\mu = N\pi$ and standard deviation $\sigma = \sqrt{N\pi(1-\pi)}$. In this case we can reformulate the hypothesis testing problem as follows:

$$H_0 : m_I \sim \text{Normal}(N\beta_I, N\beta_I(1-\beta_I)) \text{ w.p. } 1 - \theta_{II}$$
$$H_1 : m_I \sim \text{Normal}\left(N(1-(1-\beta_I)(1-\alpha_I)), N(1-(1-\beta_I)(1-\alpha_I))(1-\beta_I)(1-\alpha_I)\right) \text{ w.p. } \theta_{II}$$

We can then make the strategy a continuous threshold $T \in \mathbb{R}$. We can use differential calculus to look for Nash equilibria and we can visualize the results better, since the strategy is in a lower dimension when considering pure threshold strategies. However, it's not clear that pure equilibria will exist in this game.

5.3. **Dynamic Bargaining.** We now consider the dynamic, incomplete-information, brinksmanship game in [5]. In this game the resolve of player i is not known to player $-i$, where resolve R_i is defined for player i as

$$R_i = \frac{w_i - s_i}{w_i - d_i}.$$

In the complete information game, the sequential equilibria solution is completely determined by the players' resolve: The state with the greatest effective resolve will prevail. The sequential equilibria in the two-sided incomplete information game are more subtle and not so easily characterized. We refer readers to [5] for a detailed description.

In either the complete or incomplete information game, it turns out that cyber capabilities have a profound effect on the resolve of the players. In particular increasing one's cyber capability is equivalent to increasing one's resolve. This is obvious from the definition of resolve. In the LOLND game, resolve becomes

$$R_i^* = \frac{w_i^* - s_i^*}{w_i^* - d_i^*}.$$

Note that this definition makes resolve in the LOLND game a function of cyber capability: $R_i^* = R_i^*(p_i, p_{-i})$.

For simplicity we consider a complete information game with slight modifications from the static LOLND game considered earlier. We consider the following definition for the d^* payoffs:

$$\begin{aligned} d_I^* &= u_I(a,a) = d_I(1-p_I)(1-p_{II}) + w_I(1-p_I)p_{II} + d_I p_I(1-p_{II}) + s_I p_I p_{II} \\ d_{II}^* &= u_{II}(a,a) = d_{II}(1-p_I)(1-p_{II}) + w_{II}(1-p_{II})p_I + d_{II} p_{II}(1-p_I) + s_{II} p_I p_{II} \end{aligned}$$

The rational behind this form of the game is that if both players launch a nuclear attack and both player's LOL preemption is successful, then the payoff should be less than if both players had submitted. On the other hand, if one player's LOL preemption does not work, then that player should suffer the same consequences as the general nuclear exchange. Furthermore we assume that LOL preemption only takes effect in the place of a general nuclear exchange. The reasoning is that if one player submits prior to a nuclear exchange then there is no need to use the LOL preemption. The resulting payoffs are as

follows:

$$\begin{aligned}
d_I^* &= u_I(a,a) = d_I(1-p_{II}) + w_I(1-p_I)p_{II} + s_I p_I p_{II} \\
d_{II}^* &= u_{II}(a,a) = d_{II}(1-p_I) + w_{II}(1-p_{II})p_I + s_{II} p_I p_{II} \\
w_I^* &= u_I(a,s) = w_I \\
s_{II}^* &= u_{II}(a,s) = s_{II} \\
s_I^* &= u_I(s,a) = s_I \\
w_{II}^* &= u_{II}(s,a) = w_{II} \\
c_I^* &= u_I(s,s) = c_I \\
c_{II}^* &= u_{II}(s,s) = c_{II}
\end{aligned}$$

Analysis of this game is more straight forward. Notice that d_i^* is monotonically increasing in p_{-i} and monotonically decreasing in p_i and

$$\begin{aligned}
\lim_{p_{-i}\to 1} d_i^* &= w_i(1-p_i) + s_i p_i > s_i = s_i^* \\
\lim_{p_i\to 1} d_i^* &= d_i(1-p_{-i}) + s_i p_{-i} < s_i = s_i^*.
\end{aligned}$$

It follows that

$$\lim_{p_{-i}\to 1} R_i^* = \frac{w_i^* - s_i^*}{w_i^* - (w_i(1-p_i) + s_i p_i)} = \frac{w_i^* - s_i^*}{w_i - s_i}\frac{1}{p_i} = \frac{w_i^* - s_i^*}{w_i^* - s_i^*}\frac{1}{p_i} = \frac{1}{p_i} \geq 1$$

and

$$\lim_{p_i\to 1} R_i^* = \frac{w_i^* - s_i^*}{w_i^* - d_i(1-p_{-i}) - s_i p_{-i}} < 1.$$

By definition resolve is a probability, thus we must have $R_i^* \leq 1$. For this reason in the sequential LOLND game we set

$$R_i^* = \min\left\{\frac{w_i^* - s_i^*}{w_i^* - d_i^*}, 1\right\}$$

Notice that it is in the limit as $p_{-i} \to p_{-i}^*$ that we have $R_i^*(p_i, p_{-i}) \to 1$. Thus we can interpret the cyber capability threshold p_{-i}^* as the point at which player i obtains infinite resolve in the dynamic brinksmanship game.

Notice that R_i^* is monotonic in both player's cyber capabilities. Increasing one's own cyber capabilities increases one's own resolve, while an increase in one's opponent's cyber capability decreases one's own resolve. Thus uncertainty in cyber capability is equivalent to uncertainty in resolve in the dynamic brinksmanship game. More importantly players must consider the possibility that the opposing player has unlimited resolve, i.e. $R_i^* = 1$. In this case the opponent will never back down.

Examining the cases in which one or both players possibly have a resolve equal to one in the incomplete information, dynamic brinksmanship game is the next step in better understanding the effects of cyber LOL capabilities on nuclear deterrence. We hope to take up this task at a later time.

5.4. **Dynamic brinksmanship with strategic left-of-launch and incomplete information.** The ultimate goal of this modeling framework would be to combine all of the game extensions above. Namely we would like to consider a dynamic brinksmanship game with strategic left-of-launch and incomplete information. This will undoubtedly be a difficult task and may require numerical explorations to explore the games sequential Bayesian equilibria. We hope to take up this endeavor at a later time.

Acknowledgement

LLNL-TR-737335
This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

References

[1] Erik Gartzke and Jon Lindsay. The U.S. wants to stop North Korean missiles before they launch. That may not be a great idea. *The Washington Post*, March 2017.
[2] Erik Gartzke and Jon R. Lindsay. Thermonuclear cyberwar. *Journal of Cybersecurity*, February 2017.
[3] I. L. Glicksberg. A further generalization of the kakutani fixed point theorem, with application to nash equilibrium points. *Proceedings of the American Mathematical Society*, 3(1):170–174, 1952.
[4] Herman Kahn. *On Escalation: Metaphors and Scenarios*. Praeger, 1965.
[5] Robert Powell. *Nuclear deterrence theory: the search for credibility*. Cambridge University Pr, Cambridge, 1990. OCLC: 20089122.
[6] Thomas C Schelling. *Arms and influence*. Yale University Press, 2008. OCLC: 929632479.

Lawrence Livermore National Laboratory
E-mail address: soper3@llnl.gov