APPENDIX

A. PROOF OF LEMMA 3

LEMMA 3. *The formulation* (4) *is equivalent with its relaxed formulation as follows, where the variable* $\mathbf{x}$ *is continuous:*

$$\min_{\mathbf{x},\mathbf{B}} \quad \sum_{C_k \in \mathcal{C}} v_k x_k + \sum_{(i,j)\in E} B_{ij}\lambda_{ij} \tag{11a}$$

$$s.t. \quad \sum_{C_k \in \mathcal{C}} \alpha_{ki} x_k = 1 \qquad i \in N \tag{11b}$$

$$\sum_{C_k \in \mathcal{C}} \alpha_{ki}\alpha_{kj} x_k \geq 1 - B_{ij} \quad (i,j) \in E \tag{11c}$$

$$\mathbf{x} \in [0,1]^{|\mathcal{C}|}, \mathbf{B} \in \{0,1\}^{|N|\times|N|} \tag{11d}$$

An auxiliary Proposition is provided for proving Lemma 3 as follows:

PROPOSITION 1. *For any feasible solution* $\mathbf{x}$ *of MILP* (11), x_k *is non-zero only if the induced subgraph of* C_k *is a connected component itself or consists of several connected components of graph* $G(\mathbf{B})$ *(i.e., the remaining graph after defender's blocking).*

PROOF OF PROPOSITION 1. To prove Proposition 1, we show that for a feasible solution $\mathbf{x}$ with an $x_{\hat{k}} > 0$ corresponding to coalition $C_{\hat{k}}$, if a vertex $i \in C_{\hat{k}}$, then all vertices connected to i in $G(\mathbf{B})$ is also in $C_{\hat{k}}$. Considering the contrary, given a solution $\mathbf{x}$ with $x_{\hat{k}} > 0$ for $C_{\hat{k}}$, if there exists an edge $(\hat{i},\hat{j})$ not blocked by the defender (i.e., $B_{\hat{i}\hat{j}} = 0$), such that $\hat{i} \in C_{\hat{k}}$ while $\hat{j} \notin C_{\hat{k}}$, i.e., $\alpha_{\hat{k}\hat{i}} = 1$ and $\alpha_{\hat{k}\hat{j}} = 0$, then according to Eq.(11c), we have: $\sum_k \alpha_{k\hat{i}}\alpha_{k\hat{j}} x_k = \sum_{k:k\neq\hat{k}} \alpha_{k\hat{i}}\alpha_{k\hat{j}} x_k \geq 1$ since $\alpha_{\hat{k}\hat{j}} = 0$. Therefore, we have $\sum_k \alpha_{ki} x_k = \sum_{k:k\neq\hat{k}} \alpha_{ki} x_k + x_{\hat{k}} \geq \sum_{k:k\neq\hat{k}} \alpha_{k\hat{i}}\alpha_{k\hat{j}} x_k + x_{\hat{k}} \geq 1 + x_{\hat{k}} > 1$, which is a contradiction since $\sum_k \alpha_{ki} x_k = 1$ according to Eq.(11b). □

PROOF OF LEMMA 3. We will prove Lemma 3 by showing that for any binary solution $\mathbf{B}$, the optimal solution $\mathbf{x}$ for the formulation (11) is always binary.

Let $\mathcal{K}$ be the set of connected components of $G(\mathbf{B})$, and let $\mathcal{C}'$ be the set of coalitions whose induced subgraphs are connected components themselves or consist of several connected components of $G(\mathbf{B})$. Let K_l denote the l-th connected component in $\mathcal{K}$, and for a coalition $C \in \mathcal{C}'$, let $\beta_l^C = 1$ if the induced subgraph of C is K_l itself or consists of K_l, and $\beta_l^C = 0$ otherwise. According to Proposition 1, a feasible $\mathbf{x}$ can only take positive values on coalitions in $\mathcal{C}'$. Therefore, we can treat the connected components as the basic elements of forming coalitions, and reformulate the solution $\mathbf{x}$ as $\tilde{\mathbf{x}}$ defined on $\mathcal{C}'$. Thus, we have that every connected component K_l has a total coverage of 1,

$$\sum_{C \in \mathcal{C}'} \beta_l^C \tilde{x}_C = 1 \quad \forall K_l \in \mathcal{K} \tag{12}$$

and the minimized objective becomes: $\sum_{C\in\mathcal{C}'} v(C)\tilde{x}_C + \sum_{(i,j)\in E} B_{ij}\lambda_{ij}$. For any fixed defender strategy $\mathbf{B}$, since the coalition's value is superadditive, we have that $v(C) \geq \sum_{K_l\in\mathcal{K}} \beta_l^C \tilde{v}(K_l)$ where $\tilde{v}(K_l)$ is the value of coalition whose induced subgraph is K_l. In this case, for any $\mathbf{x}$, we have: $\sum_{C\in\mathcal{C}'} v(C)\tilde{x}_C \geq \sum_{K_l\in\mathcal{K}} \tilde{v}(K_l) \sum_{C\in\mathcal{C}'} \beta_l^C \tilde{x}_C = \sum_{K_l\in\mathcal{K}} \tilde{v}(K_l)$ according to Eq.(12), which means that the binary solution $\mathbf{x}$, which takes value of 1 for those coalitions whose induced subgraphs are connected components of $G(\mathbf{B})$ and 0 otherwise, is optimal to relaxed formulation (11). □

B. INTERIOR POINT STABILIZATION

Algorithm 2: Interior Point Stabilization (IPS)

Input: M
Output: an interior point $(\mathbf{f}^{int}, \mathbf{g}^{int})$ of $\mathcal{D}$

```
Counter = 0, f^int = 0, g^int = 0;
while Counter < M do
    Counter = Counter + 1;
    randomly generate objective coefficient (μ, ω);
    solve the (D^{μ,ω}) and (D^{-μ,-ω}) to get two extreme
    points (f*_{μ,ω}, g*_{μ,ω}) and (f*_{-μ,-ω}, g*_{-μ,-ω});
    f^int = f^int + (f*_{μ,ω} + f*_{-μ,-ω})/2M;
    g^int = g^int + (g*_{μ,ω} + g*_{-μ,-ω})/2M;
return (f^int, g^int);
```

Stabilization is a critical issue of column generation, as the standard procedure returns the optimal dual solution of RMP, an extreme point of the dual polyhedron, which is characterized by very large values for some weights f_i and g_{ij} while others are zero, and thus far away from the optimal dual solution of the unrestricted master problem. Much better approximations of the optimal dual solution would be obtained if the dual variables would take values in the center (or at least the interior) of the optimal dual polyhedron of master problem [3].

Once the master problem is solved to optimality, let $\tilde{\mathcal{C}}$ and $\tilde{E}$ be defined as the set of coalitions and edges for which $x_k > 0$ and $B_{ij} > 0$ (i.e., some of the basic columns), and let $\hat{E}$ be the set of edges for which the constraint (5c) is not tight. Using complementary slackness conditions, the optimal dual polyhedron $\mathcal{D}$ of master problem containing all optimal values for $\mathbf{f}$ and $\mathbf{g}$ is defined by

$$\sum_{i\in C_k} f_i + \sum_{(i,j)\in E: i,j\in C_k} g_{ij} \leq v_k \quad \forall C_k \in \mathcal{C}'\backslash\tilde{\mathcal{C}} \tag{13a}$$

$$\sum_{i\in C_k} f_i + \sum_{(i,j)\in E: i,j\in C_k} g_{ij} = v_k \quad \forall C_k \in \tilde{\mathcal{C}} \tag{13b}$$

$$g_{ij} \leq \lambda_{ij} \qquad \forall (i,j) \in E\backslash\tilde{E} \tag{13c}$$

$$g_{ij} = \lambda_{ij} \qquad \forall (i,j) \in \tilde{E} \tag{13d}$$

$$g_{ij} \geq 0 \qquad \forall (i,j) \in E\backslash\hat{E} \tag{13e}$$

$$g_{ij} = 0 \qquad \forall (i,j) \in \hat{E} \tag{13f}$$

$$f_i \in \mathbb{R} \qquad \forall i \in N \tag{13g}$$

To obtain an extreme point of $\mathcal{D}$, we can define a random objective function $\boldsymbol{\mu}^{\boldsymbol{T}}\mathbf{f} + \boldsymbol{\omega}^{\boldsymbol{T}}\mathbf{g}$ where $\boldsymbol{\mu}, \boldsymbol{\omega} \sim U(0,1)$, i.e., each element of $\boldsymbol{\mu}$ and $\boldsymbol{\omega}$ is uniformly distributed between 0 and 1, and solve the following LP, denoted by $(\mathcal{D}^{\boldsymbol{\mu},\boldsymbol{\omega}})$, with a simplex based method:

$$\min_{\mathbf{f},\mathbf{g}} \quad \sum_{i\in N} \mu_i f_i + \sum_{(i,j)\in E} \omega_{ij} g_{ij} \tag{14a}$$

$$\text{s.t.} \quad \text{Eqs.(13a)–(13g)} \tag{14b}$$

Different instances of $(\mathcal{D}^{\boldsymbol{\mu},\boldsymbol{\omega}})$ can be generated by defining multiple objective coefficients $(\boldsymbol{\mu}, \boldsymbol{\omega})$ and thus several extreme points of $\mathcal{D}$ can be obtained. Since $\mathcal{D}$ is a convex set, any convex combination of its extreme points will lie within it. In particular, if we take the average of all obtained extreme points, we should obtain an interior point of $\mathcal{D}$ that gives much more centered dual values. The corresponding method is called *Interior Point Stabilization* (IPS), as shown in Algorithm 2. Noticed that in Algorithm 2, for each objective coefficient $(\boldsymbol{\mu}, \boldsymbol{\omega})$ generated randomly, both problems $(\mathcal{D}^{\boldsymbol{\mu},\boldsymbol{\omega}})$ and $(\mathcal{D}^{-\boldsymbol{\mu},-\boldsymbol{\omega}})$ are solved in order to favor the identification of distant extreme points.

C. GENETIC ALGORITHM

Genetic algorithm (GA) is a popular probabilistic search algorithm applied in complex optimization problem [2]. The idea of GAs is based on an evolutionary process, where populations evolve according to natural selection and survival of the fittest. A GA simulates these process by creating an initial population of solutions and applying genetic operators on it iteratively.

We provide a GA method to solve larger coalitional security games. Here each defender strategy is encoded into a binary string of length $|E|$, where 1 represents "blocking" and 0 represent "not blocking". The fitness of each solution is the defender utility $U_d(\mathbf{B})$. Each generation consists of a fixed number g of defender strategies. There are four genetic operators: selection, crossover, mutation and reproduction. During the selection processes, since the fitness is negative, we use the tournament selection framework [1] instead of fitness proportionate selection. 2-point crossover is used to recombine the selected two defender strategies and produce the "offspring" strategies. A bitwise mutation procedure with mutation probability q is applied after the crossover is finished. The new generation is reproduced by copying g strategies with the highest utilities among parent and child generations.

The selection-crossover-mutation-reproduction cycle is repeated until the terminal criterion is met: the gap of average fitness between parent generation and child generation is smaller than a fixed constant δ.

The parameters of GA tested in our experiments are set as follows: the size of each generation is 100, and the mutation probability $q = 1/|E|$. Once the difference of fitness between the child generation and current generation is smaller than $\delta = 10^{-6}$, GA terminates.

REFERENCES

[1] D. E. Goldberg, B. Korb, and K. Deb. Messy genetic algorithms: Motivation, analysis, and first results. *Complex systems*, 3(5):493–530, 1989.

[2] M. Mitchell. *An Introduction to Genetic Algorithms*. MIT press, 1998.

[3] L. Rousseau, M. Gendreau, and D. Feillet. Interior point stabilization for column generation. *Operations Research Letters*, 35(5):660–668, 2007.