Coalition Formation in Multi-defender Security Games*

Dolev Mutzari,[1] Jiarui Gan,[2] Sarit Kraus[1]
[1]Department of Computer Science, Bar-Ilan University
[2]Max Planck Instutite for Software Systems
dolevmu@gmail.com, jrgan@mpi-sws.org, sarit@cs.biu.ac.il

Abstract

We study Stackelberg security game (SSG) with multiple defenders, where heterogeneous defenders need to allocate security resources to protect a set of targets against a strategic attacker. In such games, coordination and cooperation between the defenders can increase their ability to protect their assets, but the heterogeneous preferences of the self-interested defenders often make such cooperation very difficult. In this paper, we approach the problem from the perspective of cooperative game theory and study coalition formation among the defenders. Our main contribution is a number of algorithmic results for the computation problems that arise in this model. We provide a poly-time algorithm for computing a solution in the *core* of the game and show that all of the elements in the core are Pareto efficient. We show that the problem of computing the entire core is NP-hard and then delve into a special setting where the size of a coalition is limited up to some threshold. We analyse the parameterized complexity of deciding if a coalition structure is in the core under this special setting, and provide a poly-time algorithm for computing successful deviation strategies for a given coalition.

1 Introduction

The Multi-defender Stackelberg Security Game (SSG) is a model developed for studying real-world scenarios where multiple defenders protect a set of targets against a strategic attacker, who is their common enemy. The game is a generalization of the classic single-defender SSG, which was studied extensively in recent years in the multi-agent community (Paruchuri et al. 2008; Tambe 2011; Nguyen et al. 2013; Sinha et al. 2018; An and Tambe 2017). The augmentation of defenders makes the model fundamentally different from single-defender SSGs; understanding the relation between the defenders becomes the key to analysing the game.

Though there is very limited research, some recent work proposed *non-cooperative* game models for multi-defender SSGs with corresponding equilibrium concepts for independent movement of the defenders (Lou and Vorobeychik 2015; Gan, Elkind, and Wooldridge 2018), which lay foundations for further development of research along this line. In this paper, we approach this model from the perspective of *cooperative* game theory, and aim at understanding formation of defender coalitions in the game.

Indeed, collaboration in the form of coalition formation is a natural way of achieving efficient defense in the real world. In multi-defender SSGs, defenders' movements can be synchronized by collaboration, avoiding overspending of resources. Nevertheless, a recent attempt by Gan et al. (2020) to use mechanism design to promote defense collaboration in multi-defender SSGs shows that it's often not possible to achieve voluntary collaboration among *all* defenders in a multi-defender SSG, no matter what mechanism is used. The result is primarily due to the heterogeneous preferences of the defenders over candidate joint strategies to be carried out, which are hard to tame. Indeed, it would be very demanding to expect defenders with conflicting preferences to collaborate, but this does not mean that nothing can be done with those who share similar views or when external enforcement of cooperative behavior (e.g., via contracting) is possible. This motivates us to study multi-defender SSGs using cooperative game theory—a powerful tool for analyzing how agents form stable groups to compete based on their collective payoffs.

To fit the multi-defender SSG model into the framework of cooperative game theory, we face two major challenges. First, SSG is a game with externalities, meaning that the payoff value of a coalition is dependent on the other coalitions. Second, the game is a non-transferable utility game (NTU), meaning that the defenders' payoffs cannot be transferred between each other as monetary payoffs. These two properties introduce great difficulties in generalizing the model to allow for coalition formation. To begin with, we cannot define a utility value for a coalition that depends only on who are in the coalition; instead, we need to define a utility value for each defender separately, where a coalition structure is specified in advance. Furthermore, coalition's value depends on the attacker's strategy (which target they attack), who is a strategic player but will not be a participant in the coalition

*This work has been supported in part by the Israel Science Foundation under grant 1958/20 and the EU project TAILOR under Grant 992215. Jiarui Gan was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 945719).

formation. Basically, the attacker will take the best response that maximizes her utility. However, in general there may be multiple targets that are equally optimal for the attacker but each of which results in different utilities for every defender, making it a challenge to define a defender's utility.

Our Contribution We make the following contributions in this paper. First, we present a multi-defender SSG model that allows coalition formation. The *core* of this game consists of all coalition structures in which no defender can benefit from deviating from their current coalition in order to join another. Since it's a game with externalities, the strategies of the defenders that aren't deviating must be taken into account when considering "successful" deviations. We proposed a β-core definition, where these defenders join together to take revenge on the deviating defenders.

We then study the problem of computing a coalition structure in the core. Our main result is that there is always a strategy that results in the grand coalition structure being in the core, and this strategy can be computed in polynomial time. In addition, all elements in the core are Pareto efficient, i.e., a centralized mechanism that can enforce the agents to follow a given strategy that will give at least defender a higher utility while not making the other defenders worse off. On the other hand, the problem of computing the entire core is NP-hard even in the simple setting with only two defenders; we present a hardness result. We also present an efficient algorithm to decide if a coalition structure is stable against a given deviating coalition. The algorithm can be used to decide the stability of a coalition structure against any possible deviating coalitions; we analyze the parameterized complexity of this problem and consider in particular a setting with a small number of defender types.

Related Work Multi-agent SSGs have only recently started to attract attention from the AI community. Lou and Vorobeychik (2015) first proposed a non-cooperative game model and analyzed the Nash equilibrium and the *price of anarchy* of the game. More recently, Gan, Elkind, and Wooldridge (2018) studied a variant of the model that assumes that the defenders have common interests in protecting the targets and relaxed the constraint that each of them only protects a disjoint set of targets; they showed the existence of Nash equilibrium in their model. Based on this model, they further considered coordination between defenders and approached it by designing coordination mechanisms (Gan et al. 2020). Our model is largely based on these three pieces of work, though we focus on cooperative games. There are other papers on the applications of multi-defender SSGs or similar models (Jiang et al. 2013; Laszka, Lou, and Vorobeychik 2016; Castiglioni, Marchesi, and Gatti 2019), which are either conceptually very different from our model or are meant for very specific scenarios. To the best of our knowledge, our paper is the first to study multi-defender SSGs from the perspective of cooperative game theory.

As we've mentioned, our model is an NTU game with externalities. There have been different approaches in the literature for studying coalition formation in such games. Yi (1996) studied how the sign of external effects affects the stability of coalition structures, and how to leverage this information to provide a useful organizing principle when examining coalition structures. Finus and Rundshagen (2003) derived a non-cooperative foundation of core-stability for positive externality NTU-games. Dunne et al. (2010), Chander (2010), and Rahwan et al. (2012) developed solution concepts for coalitional resource games and supplied some techniques for finding solutions under certain assumptions. Skibski, Michalak, and Wooldridge (2018) offered a generalization of the stochastic Shapley value for coalitional games with externalities. Our work differs from previous work as our underlying model is an SSG that has a leader-follower action sequence among the players. The attacker is a special player in the game, who is strategic but will never join any coalition. This introduces many challenges to modeling and analyzing the game as we will show in this paper.

2 Preliminaries: Multi-defender SSGs

In a multi-defender SSG, there are n defenders $1, \dots, n$, and a set T of m targets, which the defenders want to protect; there is an attacker who wants to attack the targets. For any integer $n > 0$, we write $[n] = \{1, \dots, n\}$ in this paper. Each defender $i \in [n]$ has $k_i \in \mathbb{N}$ security resources that can be allocated to the targets to protect them. Using randomized allocation strategies, the resources of a defender can be distributed continuously, resulting in a vector $\mathbf{x} = (x_t)_{t \in T}$, such that x_t is the probability of target t being protected by some (at least one) resource. We call such vectors *coverage (vectors)*; they will be crucial for defining the players' utilities, which we will do shortly. With k_i resources, each defender i can use a coverage vector $\mathbf{x} \in \mathcal{C}_{k_i}$ as their allocation strategy, where for any integer $k > 0$, it is defined that $\mathcal{C}_k = \{\mathbf{x} \in [0,1]^m : \sum_{t \in T} x_t \leq k\}$; provably, such vectors can be implemented by a distribution over deterministic allocation strategies, each using at most k resources. For example, a defender with two resources can use a coverage vector $(0.6, 1, 0.4) \in \mathcal{C}_2$ over three targets; this vector can be implemented by allocating the resources to the first two targets w.p. 0.6, and to the last two targets w.p. 0.4.

Suppose that each defender i uses a coverage vector $\mathbf{x}_i = (x_{it})_{t \in T}$. When the defenders are uncoordinated, they act independently; hence, each target $t \in T$ will be protected by some defender with probability

$$c_t = 1 - \prod_{i \in [n]} (1 - x_{it}),$$

i.e., $\prod_{i \in [n]} (1 - x_{it})$ is the probability that t is not protected by any defender. We will refer to the vector $\mathbf{c} = (c_t)_{t \in T}$ as the *overall coverage (vector)*. As in a standard SSG model, the attacker wants to attack the targets, and they can conduct surveillance on the defenders' joint strategy beforehand and best respond to it. Suppose the attacker chooses to attack some target $t \in T$. Following the standard SSG model, the attack will be successful if no resource is allocated to protect t, which happens with probability $1 - c_t$ (we assume that the defenders' resources are homogeneous). In this case, the attacker receives a *reward* value $r^a(t)$ and each defender i receives a *penalty* value $p_i^d(t)$. On the other hand, the attack will fail if at least one resource is allocated to protect it, which happens with probability c_t, and in which case the attacker receives a penalty $p^a(t)$ and each defender i receives a reward $r_i^d(t)$. Thus, we can write the utility functions of each defender and the attacker as follows:

$$U_i^d(\mathbf{c}, t) = c_t \cdot r_i^d(t) + (1 - c_t) \cdot p_i^d(t) \quad (1)$$
$$U^a(\mathbf{c}, t) = (1 - c_t) \cdot r^a(t) + c_t \cdot p^a(t) \quad (2)$$

Note that for every target t, it's assumed that $r_i^d(t) > p_i^d(t)$ and $r^a(t) > p^a(t)$, so all defenders prefer an unsuccessful attack to a successful one and the attacker prefers the opposite.

Thus, as a rational player, after knowing the defenders' strategies and hence, the overall coverage $\mathbf{c}$, the attacker will best respond by attacking a target in the set

$$\mathsf{BR}(\mathbf{c}) := \arg\max_{t \in T} U^a(\mathbf{c}, t)$$

that maximizes their expected utility in this situation. When there are multiple targets in $\mathsf{BR}(\mathbf{c})$, we follow the modeling approach by Gan, Elkind, and Wooldridge (2018) and assume that the defenders will adopt the pessimistic assumption about the attacker's tie-breaking behavior. Hence, we define

$$\mathsf{br}_i(\mathbf{c}) = \arg\min_{t \in \mathsf{BR}(\mathbf{c})} U_i^d(\mathbf{c}, t);$$

that is, the target which defender i believes will be selected by the attacker.

3 Coalition Formation

When the defenders are uncoordinated, independent movement inevitably causes inefficient resource use. More specifically, if two defenders both protect a target with some positive probability, that results in a positive probability when both defenders appear on this target, which is not necessary since according to the SSG model, one resource is sufficient for thwarting the attacker's attack. We study how the defenders can form coalitions to improve efficiency of resource use and gain advantages in the game. When a subset $S \subseteq [n]$ of defenders form a coalition, they can allocate their resources jointly and play as a single defender in the game, who has $k_S = \sum_{i \in S} k_i$ resources. For example, if two defenders move independently and each of them has one resource, the set of overall coverage vectors over three targets, which can be generated from their joint strategy, is

$$\left\{\mathbf{c} \in \mathbb{R}^3 \colon c_t = 1-(1-x_{1t})(1-x_{2t}) \, \forall t \in T, \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{C}_1\right\}$$

It is not hard to verify that this set is strictly contained in $\mathcal{C}_2$, which is the set of coverage vectors these two defenders can use if they coordinate their resource allocation, allocating resources jointly as if they are one defender. A coalition is called the *grand coalition* if it contains all the defenders.

Therefore, coalition formation results in a partition of the set of defenders, which we refer to as a *coalition set*, denoted as $\mathcal{P} = \{P_1, \ldots, P_l\}$, such that $\bigcup_{i=1}^l P_i = [n]$. Each coalition P_i then chooses a joint strategy $\mathbf{x}_i \in \mathcal{C}_{k_{P_i}}$. Intuitively, the game now becomes a multi-defender SSG played by the coalitions. Similarly, the strategy profile $X = (\mathbf{x}_1, \ldots, \mathbf{x}_l)$ of the coalitions now induces the following overall coverage for each target t:

$$\mathsf{cov}_t(X) := 1 - \prod_{i \in [l]} (1 - x_{it}). \quad (3)$$

Similarly, we write the overall coverage vector as $\mathsf{cov}(X) = (\mathsf{cov}_t(X))_{t \in T}$. Given a strategy profile X and an attacker response t, the players' utilities are defined in the same way as in (1) and (2). Since a strategy profile X defines a unique coverage, for notational simplicity, we will sometimes write X instead of $\mathsf{cov}(X)$ in places where a coverage vector is expected, e.g., we write $U_i^d(X, t) = U_i^d(\mathsf{cov}(X), t)$ and $\mathsf{BR}(X) = \mathsf{BR}(\mathsf{cov}(X))$. We call a coalition set equipped with a strategy profile a *coalition structure*, and denote it by $\mathcal{CS} = \langle \mathcal{P}, X \rangle = \{(P_1, \mathbf{x}_1), \ldots, (P_l, \mathbf{x}_l)\}$.

Definition 3.1 (Coalition structure). A coalition structure $\mathcal{CS}$ is a coalition set $\mathcal{P} = (P_1, \ldots, P_l)$ (i.e., a partition of $[n]$) equipped with a strategy profile $X = (\mathbf{x}_1, \ldots, \mathbf{x}_l)$, such that $\mathbf{x}_i \in \mathcal{C}_{k_{P_i}}$ for each $i = 1, \ldots, l$. We denote $\mathcal{CS} = \langle \mathcal{P}, X \rangle = \{\langle P_1, \mathbf{x}_1 \rangle, \ldots, \langle P_l, \mathbf{x}_l \rangle\}$.

Similarly, since $\mathcal{CS}$ corresponds to a unique overall coverage vector, we will also view the players' utilities and best response as functions of $\mathcal{CS}$. More concisely, we let

$$U_i^d(\mathcal{CS}) := U_i^d\left(\mathcal{CS}, \mathsf{br}_i(\mathcal{CS})\right)$$

be the utility a coalition structure $\mathcal{CS}$ yields for defender i. We remark that our game is an NTU game with externalities, which is different to many other coalitional games where a value function $v : 2^n \to \mathcal{R}_{\geq 0}$ determines the core uniquely under a relatively reasonable set of axioms. We define the core concept for our game next.

3.1 Core of Coalition Structures

Given a coalition structure $\mathcal{CS}$, a subset of defenders can choose to deviate from the coalitions in $\mathcal{CS}$ and form a new coalition to improve their utilities. We define a *deviation* as a subset of defenders $D \subseteq [n]$ and a joint strategy $\mathbf{x} \in \mathcal{C}_{k_D}$ of them, which will be played after the deviation.

Definition 3.2 (Deviation). A deviation $\langle D, \mathbf{x} \rangle$ from a coalition structure $\mathcal{CS}$ is a nonempty subset of defenders $D \subseteq [n]$ and a joint strategy $\mathbf{x} \in \mathcal{C}_{k_D}$ of these defenders, which will be played after the deviation.

To further describe the outcome after a deviation and evaluate the benefit of making that deviation, we need to specify how the other defenders will react to a deviation (again, this is because the externalities in our game, so the utilities generated by a deviation do not depend on the deviating coalition alone). We consider the concept of *β-core* which assumes that the other defenders will take revenge against any deviation to protect the status quo (Moulin and Peleg 1982; Chalkiadakis, Elkind, and Wooldridge 2011). More concretely, we assume that after a deviation we expect to see a deviators' coalition D and a revengers' coalition $R = [n] \setminus D$ formed by the other defenders. The revenger coalition aims at making at least one deviator worse off, so as to prevent the deviation from happening. A deviation cannot be successful if the revengers can indeed achieve this.

Definition 3.3 (ϵ-successful deviation). For any $\epsilon > 0$, a deviation $\langle D, \mathbf{x} \rangle$ from a coalition structure $\mathcal{CS}$ is *ϵ-successful* if every defender $i \in D$ gets a utility improvement of at least ϵ no matter what strategy the revengers'

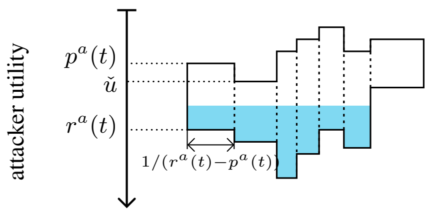

Figure 1: The tank-filling visualization (Gan, Elkind, and Wooldridge 2018).

coalition $R = [n] \setminus D$ uses, i.e., for all $i \in D$ and all $\mathbf{y} \in \mathcal{C}_{k_R}$, it holds that

$$U_i^d(\mathcal{CS}') \geq U_i^d(\mathcal{CS}) + \epsilon,$$

where $\mathcal{CS}' = \{\langle D, \mathbf{x}\rangle, \langle R, \mathbf{y}\rangle\}$.

In other words, a defender does not cooperate in a deviation unless she gets a strict utility improvement over the previous coalition structure. We also call a 0-successful deviation a *successful deviation*.

Definition 3.4 (**ϵ-core).** A coalition structure $\mathcal{CS}$ is in the ϵ-core if there exists no ϵ-successful deviation from it.

In fact, the above definition still leaves one modeling issue for us to deal with. We assume that the utility of each defender i is determined by br_i, but under this assumption, a defender can always reduce the protection to some target to make the attacker strictly prefer attacking that target, thus avoiding the worst target br_i to be chosen. Thus, as long as BR is not a singleton, the coalition structure will normally not be stable. To deal with this modeling issue, we view the core as a limit point in a similar vein to the 0^+-NSE defined by Gan, Elkind, and Wooldridge (2018) and introduce the following definition.

Definition 3.5 (0^+**-core).** A coalition structure $\mathcal{CS}$ is in the 0^+-core if there exists a sequence of coalition structures $\left(\mathcal{CS}^\ell\right)_{\ell=1}^\infty$ and a sequence of real numbers $(\epsilon^\ell)_{\ell=1}^\infty$, such that every $\mathcal{CS}^\ell$ is in the ϵ^ℓ-Core, $\lim_{\ell\to\infty} \mathcal{CS}^\ell = \mathcal{CS}$ and $\lim_{\ell\to\infty} \epsilon^\ell = 0$.

4 Non-Emptyness and Efficiency of 0^+-Core

In this section, we will show that the 0^+-core is nonempty and it provides efficiency guarantees. To present our results, we will use the tank-filling model introduced by Gan, Elkind, and Wooldridge (2018). We introduce the following notions.

Definition 4.1 (**Height of a coverage).** The height of a coverage vector $\mathbf{c} \in \mathbb{R}^m$, denoted $\mathsf{height}(\mathbf{c})$, is the optimal attacker utility it yields, i.e., $\mathsf{height}(\mathbf{c}) := \max_j U^a(\mathbf{c}, j)$.

Definition 4.2 (**Level coverage).** A coverage vector $\mathbf{c}$ is said to be level, if $c_t = 0$ for all $t \notin BR(\mathbf{c})$. A strategy profile X is level if $\mathsf{cov}(X)$ is level.

The tank-filling model is illustrated in Figure 1. Briefly speaking, each tank in the figure corresponds to a target $t \in T$, and the amount of water in it represents the coverage of that target. The width of tank t is exactly $\frac{1}{r^a(t)-p^a(t)}$, so the height of the water surface as indicated on the left axis represents exactly the attacker's utility of attacking target t. The intuition of this model is that in a stable state, the water in the tanks must form a level across all of the tanks and the corresponding coverage vector is level as defined in Definition 4.2; indeed, if the coverage is not level, some defender would rather shift resources from overly protected targets to the least protected ones (which will be attacked by the attacker) to increase their utility.

For simplicity, in the remainder of this paper, we will focus only on *canonical games*, where the *maximal level coverage* exists (see Definitions 4.3 and 4.4). Intuitively, if a game is not a canonical game, by putting $k_{[n]}$ units of water in the tank-filling model, the water surface will reach the top of some tank (i.e., $\breve{u}$ as shown in Figure 1). This introduces additional complexities for proving results to be presented while in practice we expect non-canonical games to be rare as resources are usually insufficient to allow a target in the attacker's best response set to be fully covered. All our results can also be extended to non-canonical games by using an approach similar to the one used by Gan, Elkind, and Wooldridge (2018) to show the existence of equilibrium in the uncoordinated situation.

Definition 4.3 (**Maximal level coverage).** A level coverage $\overline{\mathbf{c}}$ is called the *maximal level coverage* if $\sum_{t\in T} \overline{c} = k_{[n]}$.

Definition 4.4 (**Canonical game).** A game is called a *canonical game* if there exists a maximal level coverage $\overline{\mathbf{c}}$.

4.1 Non-Emptiness of 0^+-Core

We first present a characterization of coalition structures in the 0^+-core in the lemma below. The proof of this lemma is omitted due to space limitations.[1]

Lemma 4.5. *Suppose that a coalition structure* $\mathcal{CS} = \langle \mathcal{P}, X\rangle$ *results in a coverage vector* $\mathbf{c}$. *If the game is canonical, then* $\mathcal{CS}$ *is in the* 0^+*-core if and only if it satisfies the following two conditions:*

i. $c_t = \overline{c}_t$ *for all* $t \in T$, *where* $\overline{\mathbf{c}}$ *is the maximal level coverage.*

ii. There exists t^* *such that* $c_{t^*} > 0$ *and for every deviation* $\langle D, \mathbf{x}\rangle$, *it holds that* $U_i^d(\{\langle D, \mathbf{x}\rangle, \langle R, \mathbf{y}\rangle\}) \leq U_i^d(\mathbf{c}, t^*)$ *for some* $i \in [n]$ *and* $\mathbf{y} \in \mathcal{C}_{k_R}$.

Intuitively, the first condition is due to the fact that if the coverage is not level, then some defender(s) could always re-balance the coverage to improve the coverage of targets in the attacker's best response set; this will improve the defender's utility accordingly. In the second condition, the target t^* corresponds to the limit point of the attacker's best response in the sequence of coalition structure that defines $\mathcal{CS}$; this condition can help simplify our analysis about the limit point in the definition of the 0^+-core (Definition 3.5), so that we only need to find a target t^* in order to show that a coalition structure is in the 0^+-core.

[1]Omitted proofs can be found in the full version of this paper.

Algorithm 1: Construct a coalition structure in the 0^+-core.

Input : a canonical game instance;
Output: $\mathcal{CS} = \{\langle [n], \mathbf{c} \rangle\}$, t^*, and $\mathbf{z}_1, \ldots, \mathbf{z}_n$.

1. Let $\mathbf{z}_i = (0, \ldots, 0) \in \mathbb{R}^m$ for each $i \in [n]$.
   Let $c_j := \sum_{i \in [n]} z_{ij}$ for all $j \in T$ throughout.
   Let $\overline{\mathbf{c}}$ be the maximal coverage.
2. For each defender $i = 1, \ldots, n$:
   - Sort targets in T, so that $U_i^d(\overline{\mathbf{c}}, t_1) \le U_i^d(\overline{\mathbf{c}}, t_2) \le \cdots \le U_i^d(\overline{\mathbf{c}}, t_m)$.
   - For each $j = t_1, \ldots, t_m$:
     Increase z_{ij} until $c_j = \overline{c}_j$, or $\sum_{j' \in T} z_{ij'} = k_i$.
     If $z_{ij} > 0$, let $t^* = j$.

Given Lemma 4.5, our approach to proving the non-emptyness of the 0^+-core is to construct a coalition structure which satisfies the conditions in this lemma. The way we construct this coalition is inspired by the approach of Gan, Elkind, and Wooldridge (2018) for constructing a Nash-like equilibrium; we present it in Algorithm 1. The difference is that here we let the defenders form a grand coalition, so they can allocate their resources as if they are one defender; this results in the coverage contributed by each defender to be additive, instead of sub-additive as in the situation where they move independently.

In Algorithm 1, each defender is invited to distribute resources; the goal is for the resulting coverage to reach the maximal level coverage $\overline{\mathbf{c}}$. The defenders come one by one and distribute their resources according to the order determined by the values $U_i^a(\overline{\mathbf{c}}, t)$; once the coverage of a target t is improved to $\overline{c}_t$, they move to the next target and repeat until their resources are used up (i.e., when $\sum_{j \in T z_{ij}} = k_i$). Hence, z_{ij} indicates the amount of resources defender i contributes to target j; these variables and t^* are set to help us verify the second condition of Lemma 4.5 in the proof of the next theorem, which shows that the 0^+-core is non-empty.

Theorem 4.6. *The 0^+-core is non-empty and it always contains a grand coalition structure.*

Proof. We show that the grand coalition structure $\mathcal{CS}$ generated by Algorithm 1 satisfies the two conditions in the statement of Lemma 4.5, so it is in the 0^+-core.

Note that Algorithm 1 always ends with $\mathbf{c} = \overline{\mathbf{c}}$. Indeed, suppose that this is not the case, according to the way $\mathbf{c}$ is updated in the algorithm we always have $c_j \le \overline{c}_j$ for all j, so the only possibility when we have $\mathbf{c} \ne \overline{\mathbf{c}}$ is that $c_\ell < \overline{c}_\ell$ for some target ℓ. This implies that

$$\textstyle\sum_{j \in T} c_j < \sum_{j \in T} \overline{c}_j = k_{[n]}$$

given that $\overline{\mathbf{c}}$ is the maximal level coverage. Hence,

$$\textstyle\sum_{i \in [n], j \in T} z_{ij} = \sum_{j \in T} c_j < k_{[n]},$$

which means that $\sum_{j \in T} z_{ij} < k_i$ for some defender i. This is a contradiction because we would expect $c_\ell = \overline{c}_\ell$ when we increase $z_{i\ell}$ in Step 2. Thus, $\mathcal{CS}$ satisfies Condition (i) of Lemma 4.5.

It remains to show that the second condition is also satisfied. We let t^* be the target in the output of the algorithm. Indeed, we have $c_{t^*} > 0$ as it is always points to a target that receives a positive improvement in its coverage as in Step 2. Pick arbitrary $D \subseteq [n]$ and deviating strategy $\mathbf{x} \in k_D$. We show that the strategy $\mathbf{y}$ with

$$y_j = \textstyle\sum_{i \in R} z_{ij} \quad \text{for all } j \in T$$

will make the inequality in Condition (ii) hold. Note that according to Step 2, we have $\sum_{j \in T} z_{ij} \le k_i$ for all j, so it follows that $\sum_{j \in T} y_j = \sum_{i \in R} \sum_{j \in T} z_{ij} \le k_R$ and indeed $\mathbf{y} \in \mathcal{C}_{k_R}$.

Let $\mathcal{CS}' = \{\langle D, \mathbf{x} \rangle, \langle R, \mathbf{y} \rangle\}$ and $\mathbf{c}'$ be the coverage resulting from $\mathcal{CS}'$, i.e., $c'_j = 1 - (1 - x_t)(1 - y_t)$. If $\mathbf{c}' = \mathbf{c}$, then we have

$$U_i^d(\mathcal{CS}') = \min_{t \in \mathsf{BR}(\mathbf{c})} U_i^d(\mathbf{c}, t) \le U_i^d(\mathbf{c}, t^*),$$

where we use the fact that $t^* \in \mathsf{BR}(\mathbf{c})$ given that $\mathbf{c} = \overline{\mathbf{c}}$ is a level coverage and $c_{t^*} > 0$ (see Definition 4.2). Hence, Condition (ii) holds. In what follows we assume that $c'_t \ne c_t$ for some t.

We must have $c'_\ell < c_\ell$ for some $\ell \in T$: otherwise, $c'_t \ge c_t$ for all t while at least one of these inequalities must be strictly satisfied, which leads to the following contradiction:

$$k_{[n]} = \sum_{t \in T} \overline{c}_t = \sum_{t \in T} c_t < \sum_{t \in T} c'_t = \sum_{t \in T} (1 - (1 - x_t)(1 - y_t)) \le \sum_{t \in T} (x_t + y_t) \le k_{[n]}.$$

Pick an arbitrary $\ell^* \in \mathsf{BR}(\mathbf{c}')$. It must be that $c'_{\ell^*} < c_{\ell^*}$: otherwise, $c'_{\ell^*} \ge c_{\ell^*}$, so we can establish the following inequalities, which contradicts the fact that $\ell^* \in \mathsf{BR}(\mathbf{c}')$:

$$U^a(\mathbf{c}', \ell^*) \le U^a(\mathbf{c}, \ell^*) \le U^a(\mathbf{c}, \ell) < U^a(\mathbf{c}', \ell),$$

where the first and third transitions follow by monotonicity of the utility function; and the second transition is due to the fact that $\ell \in \mathsf{BR}(\mathbf{c})$ (as $\mathbf{c}$ is level and $c_\ell > 0$).

Observe that for all j with $\sum_{i \in D} z_{ij} = 0$, we have

$$c'_j \ge y_j = \textstyle\sum_{i \in R} z_{ij} = \sum_{i \in [n]} z_{ij} = c_t.$$

Thus, we must have $\sum_{i \in D} z_{i\ell^*} > 0$, which means $z_{i\ell^*} > 0$ for some $i \in D$. This implies that $U_i^d(\overline{\mathbf{c}}, \ell^*) \le U_i^d(\overline{\mathbf{c}}, t^*)$: if the opposite holds, it must be that when $z_{i\ell^*}$ is set to a positive value at Step 2 of Algorithm 1, we already have $c_{t^*} = \overline{c_{t^*}}$; this contradicts the fact that t^* is the last target that reaches coverage $\overline{\mathbf{c}}$. Consequently, we have

$$U_i^d(\mathcal{CS}') = \min_{t \in \mathsf{BR}(\mathbf{c}')} U_i^d(\mathbf{c}', t) \le U_i^d(\mathbf{c}', \ell^*) < U_i^d(\overline{\mathbf{c}}, \ell^*) \le U_i^d(\overline{\mathbf{c}}, t^*) = U_i^d(\mathbf{c}, t^*)$$

(recall that $c'_{\ell^*} < c_{\ell^*} = \overline{c}_{\ell^*}$) so Condition (ii) holds, too. □

Nevertheless, according to the following theorem, the problem becomes hard if we exclude the grand coalition from our consideration. In other words, to compute the entire core is computationally hard.

Theorem 4.7. *It is NP-hard to decide if the 0^+-core contains any coalition structure $\mathcal{CS} = \langle \mathcal{P}, X \rangle$ where $\mathcal{P}$ does not contain the grand coalition even when $n = 2$.*

4.2 Efficiency of 0^+-Core

It turns out that the 0^+-core also guarantees some efficiency properties given that the coverage is always "maximized" by Lemma 4.5. The following result says that a coalition structure in the 0^+-core is *weakly Pareto efficient*, meaning that the situation cannot be strictly improved for every defender in any other coalition structure.

Proposition 4.8. *For every coalition structure sequence $\left(\mathcal{CS}^\ell\right)_{\ell=1}^\infty$ associated with a $\mathcal{CS}$ in the 0^+-core, there exists no coalition structure $\mathcal{CS}'$ such that for every defender $i \in [n]$, it holds that $U_i^d(\mathcal{CS}') > \lim_{\ell\to\infty} U_i^d(\mathcal{CS}^\ell)$.*

Indeed, a stronger notion of the 0^+-core, call it *strict* 0^+-core, can be defined by using a weaker notion of ϵ-successful deviation, which only requires *some* (instead of all) deviator to have an ϵ advantage as long as the other deviators will not be hurt by the deviation (i.e., the others have advantages of at least 0). The strict 0^+-core may be empty (see Example B.1). However, when it is not empty, any coalition structure in it is (strongly) Pareto efficient. In other words, given a coalition structure in the strict 0^+-core, one cannot hope that, by using a centralized mechanism that enforces the defenders to follow a given strategy, we can improve the utility of some defender without hurting the others. We present this result below.

Proposition 4.9. *Every coalition structure in the strict 0^+-core is* Pareto efficient. *For every coalition structure sequence $\left(\mathcal{CS}^\ell\right)_{\ell=1}^\infty$ associated with a $\mathcal{CS}$ in the strict 0^+-core, there exists no coalition structure $\mathcal{CS}'$ and defender i such that $U_i^d(\mathcal{CS}') > \lim_{\ell\to\infty} U_i^d(\mathcal{CS}^\ell)$ and for every other defender $j \in [n] \setminus \{i\}$ it holds that $U_j^d(\mathcal{CS}') \geq \lim_{\ell\to\infty} U_j^d(\mathcal{CS}^\ell)$.*

5 Stability of a Coalition Structure

We have shown that the grand coalition is always in the 0^+-core. In this section, we investigate the problem of deciding the stability of a given coalition structure; namely, whether it is in the 0^+-core. Our main result in this section is an efficient algorithm for deciding the stability of a coalition structure against a given deviating coalition D. To present this result, we first define the following useful notions.

Definition 5.1 (**ϵ-safety demand**)**.** The *ϵ-safety demand* ($\epsilon > 0$) of a defender $i \in [n]$ with respect to a coalition structure $\mathcal{CS}$ is a vector $\mathbf{s}_i^\epsilon = (s_{it}^\epsilon)_{t\in T} \in [0,1]^m$, such that

$$s_{it}^\epsilon = \max\left\{0,\ \frac{U_i^d(\mathcal{CS}) + \epsilon - p_i^d(t)}{r_i^d(t) - p_i^d(t)}\right\}. \quad (4)$$

The safety demand of a set $D \subseteq [n]$ of defenders with respect to $\mathcal{CS}$ is a vector $\mathbf{s}^\epsilon = (s_t^\epsilon)_{t\in T} \in [0,1]^m$ such that

$$s_t^\epsilon = \max_{i\in D} s_{it}^\epsilon.$$

Algorithm 2: Solve (5).

Input : $\mathcal{CS}, D$;
Output: $\mathbf{x}$, v.

1. Compute the ϵ-safety demand $\mathbf{s}$ of coalition D with respect to $\mathcal{CS}$ according to (4). Sort targets in T, so that $U^a(\mathbf{s}, t_1) \geq U^a(\mathbf{s}, t_2) \geq \cdots \geq U^a(\mathbf{s}, t_m)$;
2. For each $j = 0, 1, \ldots, m$:
   - Let $T^+ = \{t_\ell \in T : \ell \leq j\}$ and $T^- = T \setminus T^+$;
   - Solve the following LP (linear program):

$$\begin{aligned} \min_{\mathbf{x},v} \quad & v & \\ \text{subject to} \quad & v \geq U^a(x_t, t) & \text{for all } t \in T^- \\ & 0 \leq x_t \leq 1 & \text{for all } t \in T^- \\ & s_t^\epsilon \leq x_t \leq 1 & \text{for all } t \in T^+ \\ & \textstyle\sum_{t\in T^-} x_t \leq q & \end{aligned}$$

3. Find the LP in Step 2 with the smallest optimal objective value.[2] Output the optimal solution $\mathbf{x}$ and v of this LP.

Definition 5.2 (**ϵ-safe deviation strategy and safety value**)**.** With respect to $\mathcal{CS}$, an *ϵ-safe* ($\epsilon > 0$) deviation strategy $\mathbf{x}$ for a coalition D is the solution to the following optimization:

$$\min_{\mathbf{x}'\in\mathcal{C}_{k_D}} \max_{t\in T: x'_t < s_t^\epsilon} U^a(\mathbf{x}', t), \quad (5)$$

where $\mathbf{s}$ is the safety demand of D and when $\{t \in T : x'_t < s_t^\epsilon\} = \varnothing$ we define $\max_{t\in T: x'_t < s_t^\epsilon} U^a(\mathbf{x}', t) = -\infty$. We call the optimal value of (5) the *ϵ-safety value* of D.

In other words, Definition 5.1 ensures that a target t being attacked does not remove the deviating coalition's incentive to deviate as long as the coverage of t is not below the safety demand. As for targets that receive coverage below the safety demand, the deviating coalition can only hope that they are not the best response of the attacker. Hence, the idea behind (5) is to minimize the attacker's utility for attacking these targets.

In what follows, for any strategy $\mathbf{x}$, we define $T_\mathbf{x}^+ = \{t \in T : x_t \geq s_t^\epsilon\}$ and $T_\mathbf{x}^- = \{t \in T : x_t < s_t^\epsilon\}$. The following lemmas provide useful characterizations for ϵ-safe deviation strategies and ϵ-successful strategies. Using these lemmas, we prove our main result Theorem 5.5.

Lemma 5.3. *Let $\mathbf{x}$ be an ϵ-safe deviation strategy of a coalition $D \subseteq [n]$ and v be the ϵ-safety value. If $v > -\infty$, then it holds that:*

i. $v = U^a(\mathbf{x}, t_1) \leq U^a(\mathbf{x}, t_2)$, for any $t_1 \in T_\mathbf{x}^-, t_2 \in T_\mathbf{x}^+$;
ii. $x_t = s_t^\epsilon$ for all $t \in T_\mathbf{x}^+$.

Lemma 5.4. *Let $\mathbf{x}$ be an ϵ-safe deviation strategy of a coalition $D \subseteq [n]$ with respect to $\mathcal{CS}$, and v be the ϵ-safety value. Let $\gamma_t = \frac{r^a(t)-v}{r^a(t)-p^a(t)}$ for each $t \in T$ (so that $U^a(\gamma_t, t) = v$). Then D has an ϵ-successful deviation from $\mathcal{CS}$ if and only if one of the following conditions is true:*

[2] We let the optimal objective value be ∞ when the program is infeasible, and $-\infty$ when it is unbounded.

- $v = -\infty$;
- $\gamma_t > 1$ *for some* $t \in T^+_{\mathbf{x}}$; *or*
- $\sum_{t \in T^+_{\mathbf{x}}} \frac{\gamma_t - s^\epsilon_t}{1 - s^\epsilon_t} > k_R$, *where* $R = [n] \setminus D$ *is the revergers' coalition.*

Theorem 5.5. *There is a polynomial-time algorithm to decide whether a coalition* $D \subseteq [n]$ *has an* ϵ*-successful deviation strategy from a coalition structure* $\mathcal{CS}$.

Proof. Using Lemma 5.4, our approach to deciding the stability of a coalition structure against a coalition D is to compute an ϵ-safe deviation strategy and check whether any of the conditions in Lemma 5.4 holds. Indeed, given $\mathbf{x}$ and v, to verify the conditions is trivial, so the key is to solve the optimization problem defined in (5). To this end, we present Algorithm 2 (which is obviously a polynomial-time algorithm) and show that it outputs an optimal solution to (5) correctly. To distinguish, let $\mathbf{x}^*$ be an optimal solution to (5) and $v^* = \max_{t \in T: x'_t < s^\epsilon_t} U^a(\mathbf{x}^*, t)$ be the corresponding objective value.

For the output $\mathbf{x}$ and v of Algorithm 2, the constraints of the LP in Step 2 of the algorithm ensures that $T^-_{\mathbf{x}} \subseteq T^-$, so we have

$$v = \max_{t \in T^-} U^a(\mathbf{x}, t) \geq \max_{t \in T: x'_t < s^\epsilon_t} U^a(\mathbf{x}, t) \geq v^*. \quad (6)$$

On the other hand, observe that the LP is equivalent to: $\min_{\mathbf{x}' \in \mathcal{C}_{k_D}} \max_{t \in T^-} U^a(\mathbf{x}', t)$. Hence, as long as in some round it holds that $T^- = T^-_{\mathbf{x}^*}$, we will have

$$\begin{aligned} v &= \min_{\mathbf{x}' \in \mathcal{C}_{k_D}} \max_{t \in T^-} U^a(\mathbf{x}', t) \\ &\leq \max_{t \in T^-} U^a(\mathbf{x}^*, t) = \max_{t \in T^-_{\mathbf{x}^*}} U^a(\mathbf{x}^*, t) = v^* \end{aligned}$$

and hence, $v = v^*$, which will then complete the proof. We show next that indeed $T^- = T^-_{\mathbf{x}^*}$ in some round.

If $T^-_{\mathbf{x}^*} = \varnothing$ or $T^+_{\mathbf{x}^*} = \varnothing$, then in round $j = m$ or $j = 0$, we have $T^- = T^-_{\mathbf{x}^*}$. Otherwise, $T^-_{\mathbf{x}^*} \neq \varnothing$ implies that $v^* > -\infty$, so Lemma 5.3 is applicable, according to which the following inequality holds for any $a \in T^-_{\mathbf{x}^*}$ and $b \in T^+_{\mathbf{x}^*}$:

$$U^a(\mathbf{s}, a) < U^a(\mathbf{x}^*, a) \leq U^a(\mathbf{x}, b) = U^a(\mathbf{s}, b).$$

Since $t_1, \ldots, t_m$ are ordered according to the value $U^a(\mathbf{s}, t)$, the above inequality implies that there exists $j \in [m-1]$ such that $T^+_{\mathbf{x}^*} = \{t_1, \ldots, t_j\}$ and $T^-_{\mathbf{x}^*} = \{t_{j+1}, \ldots, t_m\}$. In round j, we have exactly $T^- = T^-_{\mathbf{x}^*}$. □

We have presented an efficient way to decide the stability of a coalition structure against a given deviating coalition. Generally speaking, if a deviating coalition is not provided, we do not know yet how to deal with this task; we leave this problem open in this paper. Despite this unanswered problem, we will next discuss a special setting that allow us to adapt the above results to decide the stability of a coalition structure against *any* possible deviating coalition.

Trivially, now that we know how to decide if a given coalition has a successful deviating strategy, when there is only a small number of defenders we can enumerate all $2^n - 1$ coalitions to check the stability. Similarly, if the size of coalitions the defenders can form is bounded by a small number, the same approach applies. A slightly more complex realistic setting is one in which the number of defenders is unbounded but the number of possible defender *types* is small. Such games were studied by Shrot, Aumann, and Kraus (2010). In our model, two defenders have the same type if they have the same payoff parameters, i.e., $p^d_i(t) = p^d_j(t)$ and $r^d_i(t) = r^d_j(t)$ for every $t \in T$. Theorem 5.6 shows that if the number of types is fixed, we can decide the stability of a coalition structure in polynomial time. The key idea for proving this result is to show that it suffices to enumerate all the possible combinations of defender types.

Theorem 5.6. *There is an algorithm for deciding whether a coalition structure is in the* ϵ*-core or not that runs in time* $O(2^\lambda \operatorname{poly}(m, n))$*, where* λ *is the number of defender types.*

6 Conclusion

We provide a model of a coalitional multi-defender security game that enables defenders to form coalitions and distribute their resources jointly. We focus on the solution concept of the core. We prove that the 0^+-core is none-empty, give a polynomial-time algorithm for computing a grand coalition structure in the 0^+-core, and show that it is NP-hard to compute the entire core even when there is a fixed number of defenders. We also discuss the stability of a coalition structure against a given deviating coalition and the parameterized complexity of validating that a coalition structure is in the core.

There are other two possible approaches that could have been considered, namely, coordination mechanism and negotiation. However, in coordination mechanism, we assume the existence of a reliable third party, and that all of the defenders report their properties honestly to that party and that they all accept the resultant decision. None of these assumptions are required in our coalition formation framework. Moreover, even with the strong assumptions that coordination mechanism requires, it cannot yield a better solution for the parties than our stable grand coalition structure.

As for negotiation, a challenge to deal with is when a defender opts out of negotiation. One can consider frameworks in which the defenders that reach an agreement are able to punish the defenders that opted out. Conversely, there are games where the opting out defenders can take advantage of the agreed upon strategies of the other defenders. In order for this approach to work, it has to be efficient, to have a utility guarantee and for the complexity of finding the equilibrium negotiation strategies to be low.

The idea of a coalitional Stackelberg game can evolve in many directions. It is interesting to consider settings where the game does not have complete information, that is, the attacker may not be aware of the defenders' allocation strategies. Furthermore, the dependencies between targets may be considered. Protecting one target may also protect other targets close to it. Moreover, a more relaxed definition for the core should be considered, such as γ-core. We assumed that after deviation the rest of the defenders take revenge, however this may, in general, not be a credible threat.

References

An, B.; and Tambe, M. 2017. *Stackelberg Security Games (SSG) Basics and Application Overview*, 485–507. Cambridge University Press.

Castiglioni, M.; Marchesi, A.; and Gatti, N. 2019. Be a Leader or Become a Follower: The Strategy to Commit to with Multiple Leaders. In *IJCAI*, 123–129.

Chalkiadakis, G.; Elkind, E.; and Wooldridge, M. 2011. Computational aspects of cooperative game theory. *Synthesis Lectures on Artificial Intelligence and Machine Learning* 5(6): 1–168.

Chander, P. 2010. Cores of games with positive externalities. *CORE DP* 4: 2010.

Dunne, P. E.; Kraus, S.; Manisterski, E.; and Wooldridge, M. 2010. Solving coalitional resource games. *Artificial Intelligence* 174(1): 20–50.

Finus, M.; and Rundshagen, B. 2003. A Non-Cooperative Foundation of Core-Stability in Positive Externality NTU-Coalition Games. *FEEM* .

Gan, J.; Elkind, E.; Kraus, S.; and Wooldridge, M. 2020. Mechanism Design for Defense Coordination in Security Games. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS'20)*, 402–410.

Gan, J.; Elkind, E.; and Wooldridge, M. 2018. Stackelberg security games with multiple uncoordinated defenders. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, 703–711. International Foundation for Autonomous Agents and Multiagent Systems.

Jiang, A. X.; Procaccia, A. D.; Qian, Y.; Shah, N.; and Tambe, M. 2013. Defender (mis) coordination in security games. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI'13)*, 220–226.

Laszka, A.; Lou, J.; and Vorobeychik, Y. 2016. Multi-Defender Strategic Filtering Against Spear-Phishing Attacks. In *AAAI*, 537–543. Citeseer.

Lou, J.; and Vorobeychik, Y. 2015. Equilibrium Analysis of Multi-Defender Security Games. In *IJCAI*, 596–602.

Moulin, H.; and Peleg, B. 1982. Cores of effectivity functions and implementation theory. *Journal of Mathematical Economics* 10(1): 115 – 145.

Nguyen, T. H.; Yang, R.; Azaria, A.; Kraus, S.; and Tambe, M. 2013. Analyzing the Effectiveness of Adversary Modeling in Security Games. In *AAAI*.

Paruchuri, P.; Pearce, J. P.; Marecki, J.; Tambe, M.; Ordonez, F.; and Kraus, S. 2008. Efficient Algorithms to Solve Bayesian Stackelberg Games for Security Applications. In *AAAI*, 1559–1562.

Rahwan, T.; Michalak, T.; Wooldridge, M.; and Jennings, N. R. 2012. Anytime coalition structure generation in multi-agent systems with positive or negative externalities. *Artificial Intelligence* 186: 95–122.

Shrot, T.; Aumann, Y.; and Kraus, S. 2010. On agent types in coalition formation problems. In *Proceedings of the 9th International Conference on Autonomous Agents and Multi-agent Systems: volume 1-Volume 1*, 757–764.

Sinha, A.; Fang, F.; An, B.; Kiekintveld, C.; and Tambe, M. 2018. Stackelberg security games: Looking beyond a decade of success. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*. IJCAI.

Skibski, O.; Michalak, T. P.; and Wooldridge, M. 2018. The Stochastic Shapley Value for coalitional games with externalities. *Games and Economic Behavior* 108: 65–80.

Tambe, M. 2011. *Security and game theory: algorithms, deployed systems, lessons learned*. Cambridge university press.

Yi, S.-S. 1996. Stable coalition structures with externalities. *Available at SSRN 2071* .

A Omitted Proofs

A.1 Proof of Lemma 4.5

First, we show that any coalition structure that satisfies the conditions of the lemma is in the 0^+-core.

Let $\mathcal{CS} = \{([n], \pi)\}$ be a coalition structure that satisfies the conditions of the theorem, with some fixed target t^*. Since by the first condition $\mathbf{c}$ is a level coverage, and since $c_{t^*} > 0$, by definition $t^* \in BR$. Therefore, for any $\delta > 0$, if we reduce the the coverage of target t^* by δ, then $BR = \{t^*\}$. Let this coalition structure be $\mathcal{CS}_\delta$, and consider a series $0 < \delta_\ell \to 0$. Then we have that $\mathcal{CS}_{\delta_\ell}$ converges to $\mathcal{CS}$.

Therefore, by definition of the 0^+-core, it is enough to prove that $\mathcal{CS}_{\delta_\ell}$ is in some ϵ_ℓ-core, where $0 < \epsilon_\ell \to 0$. Let $(D, \mathbf{x})$ be some deviation. Then by the 2[nd] condition, there exist a defender i and a revenge strategy $\mathbf{y}$ for the revenge coalition R, such that $U_i^d(\mathcal{CS}') \leq U_i^d(\mathbf{c}, t^*)$, where $\mathcal{CS}' = \{\langle D, \mathbf{x}\rangle, \langle R, \mathbf{y}\rangle\}$.

Note that $U_i^d(\mathcal{CS}_{\delta_\ell}) = U_i^d(\mathcal{CS}_{\delta_\ell}, t^*) = U_i^d(\mathbf{c}, t^*) - \delta_\ell \cdot (r_i^d(t^*) - p_i^d(t^*))$. If we let $\epsilon_\ell := \delta_\ell \cdot (r_i^d(t^*) - p_i^d(t^*))$, then:

$$U_i^d(\mathcal{CS}') \leq U_i^d(\mathcal{CS}_{\delta_\ell}) + \epsilon_\ell,$$

and therefore $\mathcal{CS}_{\delta_\ell}$ is in the ϵ_ℓ-core, and $\epsilon_\ell \to 0$.

Next, let $\mathcal{CS}$ be a coalition structure in the core. Let $\{\mathcal{CS}_\ell\}_{\ell=1}^\infty$ be a corresponding sequence of coalition structures in the ϵ_ℓ-core, where $\epsilon_\ell \to 0$. We must show that it satisfies the both conditions of the lemma. Denote by $\mathbf{c}^{(\ell)}$ the coverage vector of coalition structure $\mathcal{CS}_\ell$.

As for the 1[st] condition, if it doesn't hold, then there is a target $t \in T$ such that $c_t < \bar{c}_t$. Observe the following deviation of the grand coalition $[n]$: We take $c'_j = 0.5 \cdot c_j + 0.5 \cdot \bar{c}_j$. Intuitively, this increases the coverage of less protected targets and decreases the coverage of overly protected targets (while preserving the preference order of the attacker). Note that since $\sum_j c'_j = 0.5 \cdot \sum_j c_j + 0.5 \cdot \sum_j \bar{c}_j \leq 0.5 \cdot \sum_j k_j + 0.5 \cdot \sum_j k_j = \sum_j k_j$, it is a legal coverage vector.

Let $A = \{t : c_t < \bar{c}_t\}$, and $A_\ell = \{t : c_t^{(\ell)} < \bar{c}_t\}$. By limit definition, there exist some ℓ_0 such that for every $\ell > \ell_0$, $A_\ell = A$. Let $B = \{t : c_t > \bar{c}_t\}$, and $B_\ell = \{t : c_t^{(\ell)} > \bar{c}_t\}$. By limit definition, there exist some ℓ_1 such that for every $\ell > \ell_1$, $B_\ell = B$.

Observe that, for each target t, we have

$$U_i^d(\mathbf{c}', t) = U_i^d(\mathbf{c}^{(\ell)}, t) + (1/2\bar{c}_t - 1/2c_t^{(\ell)}) \cdot (r_i^d(t) - p_i^d(t)).$$

By limit definition, there exist some $\ell_2 \geq \max(\ell_0, \ell_1)$, such that for every $\ell \geq \ell_2$, and for every target $t \in A$, $\bar{c}_t - c_t^{(\ell)} > (\bar{c}_t - c_t)/2$.

Let $\epsilon = \min_{t \in A} \frac{(\bar{c}_t - c_t)}{2} \cdot \min_i (r_i^d(t) - p_i^d(t))$, then we will show that $\mathbf{c}'$ is an ϵ successful deviation for every $\ell \geq \ell_2$. Let i be some defender and let $t \in \mathsf{br}_i(\mathbf{c}^{(\ell)})$. Then $t \notin B$. If $t \in A$, then the advantage of i from the deviation is indeed more than ϵ. The case left is when $c_t = \bar{c}_t$. However, this suggests that $\mathbf{c}^(\ell) = \bar{\mathbf{c}}$. However, that can happen only for finitely many times, or else we will have $\bar{\mathbf{c}}$ as a partial limit, and therefore $\mathbf{c} = \bar{\mathbf{c}}$. Therefore, there exists $\ell_2 \geq \ell_1$ such for each $\ell \geq \ell_2$, we have $\mathsf{BR}(\mathbf{c}^{(\ell)}) \subseteq A$. Now let $\ell_3 \geq \ell_2$ such that $\mathcal{CS}_{\ell_2}$ is in the $\epsilon_{\ell_3} < \epsilon$-core. This results in a contradiction.

Now assume the first condition holds, but the second doesn't. In that case, let t be some target with $c_t > 0$. Then we have a deviation $\langle D, \mathbf{x}\rangle$ such that $U_i^d(\{\langle D, \mathbf{x}\rangle, \langle R, \mathbf{y}\rangle\}) > U_i^d(\mathbf{c}, t)$ for every $i \in [n]$ and $\mathbf{y} \in \mathcal{C}_{k_R}$. Let $\epsilon = \min_i(U_i^d(\{\langle D, \mathbf{x}\rangle, \langle R, \mathbf{y}\rangle\}) - U_i^d(\mathbf{c}, t))$.

There exist some ℓ_0 such that for every $\ell \geq \ell_0$, $|U_i^d(\mathbf{c}, t) - U_i^d(\mathbf{c}^{(\ell)}, t)| < \epsilon/2$. Therefore, we have:

$$U_i^d(\{\langle D, \mathbf{x}\rangle, \langle R, \mathbf{y}\rangle\}) - U_i^d(\mathbf{c}^{(\ell)}, t) = (U_i^d(\{\langle D, \mathbf{x}\rangle, \langle R, \mathbf{y}\rangle\}) - U_i^d(\mathbf{c}, t)) +$$

Denote by $C = \{t : c_t > 0\}$, and $C_\ell = \{t : c_t^{(\ell)} > 0\}$. By limit definition, there exist some $\ell_1 > \ell_0$ such that for every $\ell > \ell_1$, $C_\ell = C$.

Now there are 2 cases: in the first case, $\mathsf{BR}(\mathbf{c}) \cap C =\neq \emptyset$. In this case, $\langle D, \mathbf{x}\rangle$ is an ϵ successful deviation from every $\mathcal{CS}_\ell$ where $\ell \geq \ell_1$. This is because for each defender, $U_i^d(\mathbf{c}) \leq U_i^d(\mathbf{c}, t)$ for some $t \in \mathsf{BR}(\mathbf{c}) \cap C$. Therefore, taking $\ell_2 \geq \ell_1$ such that $\epsilon_\ell < \epsilon$ will give a contradiction to the fact that $\mathcal{CS}_{\ell_2}$ is in the ϵ_{ℓ_2}-core.

Therefore, it must be the second case, where each target $t \in \mathsf{BR}(\mathbf{c})$ yields $c_t = 0$. However, this suggests that $\mathbf{c} = 0$, which is clearly not in the core.

A.2 Proof of Theorem 4.7

We show a reduction from the PARTITION problem, which is one of the classic NP-complete problems. An instance of the PARTITION problem is given by a set of positive integers $I = \{a_1, \ldots, a_l\}$. It is a yes-instance if there exists a partition $\{I_1, I_2\}$ of I, such that $\sum_{\ell \in I_1} a_\ell = \sum_{\ell \in I_2} a_\ell$; otherwise, it is a no-instance. Without loss of generality, we can assume that $\sum_{\ell \in I} a_\ell$ is always even.

We construct the following game for a given instance I of the PARTITION problem. Let $\gamma = \frac{1}{2} \sum_{\ell \in I} a_\ell$. We let $n = 2$ and $k_1 = k_2 = 1$, i.e., there are two defenders, each having one resource. Let there be l targets: $T = \{t_1, \ldots, t_l\}$. The attacker's payoff parameters are $r^a(t_\ell) = \frac{a_\ell}{\gamma}$ and $p^a(t_\ell) = \frac{a_\ell}{\gamma} - 1$ for each $\ell \in [l]$. Defenders 1 and 2 have the same payoff parameters: $r_i^d(t_\ell) = 1 - \frac{a_\ell}{\gamma}$ and $p_i^d(t_\ell) = -\frac{a_\ell}{\gamma}$ for each $i \in \{1, 2\}$ and $\ell \in [l]$. Note these parameters makes the game zero-sum between the attacker and each individual defender, i.e., we have

$$U^a(\mathbf{c}, t) = -U_i^d(\mathbf{c}, t) \tag{7}$$

for any coverage $\mathbf{c}$ and target $t \in T$. We will argue that I is a yes-instance if and only if there is a coalition structure in the 0^+-core, which is not a grand coalition structure.

The "only if" direction. Suppose that I is a yes-instance; there exists a partition $\{I_1, I_2\}$ of I, with $\sum_{\ell \in I_1} a_\ell = \sum_{\ell \in I_2} a_\ell = \gamma$. We argue that the coalition structure $\mathcal{CS} = \{\langle\{1\}, \mathbf{x}_1\rangle, \langle\{2\}, \mathbf{x}_2\rangle\}$ is in the 0^+-core, where for each $i = \{1, 2\}$ and $\ell \in [l]$ we let $x_{i\ell} = \frac{a_\ell}{\gamma}$ if $\ell \in I_i$ and $x_{i\ell} = 0$, otherwise. Indeed, $\mathbf{x}_1$ and $\mathbf{x}_2$ satisfy the budget constraint:

it holds that

$$\sum_{\ell\in[l]} x_{i\ell} = \sum_{\ell\in I_i} x_{i\ell} = \frac{1}{\gamma}\sum_{\ell\in I_i} a_\ell = 1 \le k_i,$$

so they are feasible strategies. We will next show that $\mathcal{CS}$ satisfies the conditions in Lemma 4.5.

First, since each ℓ is contained in exactly one of I_1 and I_2, each target t_ℓ is protected by exactly one defender; hence, $\mathcal{CS}$ results in coverage $c_\ell = \frac{a_\ell}{\gamma}$ for each target t_ℓ. Taking this value into the attacker's utility function gives exactly

$$U^a(\mathbf{c}, t) = 0 \quad \text{for all } t \in T, \tag{8}$$

so $\mathsf{BR}(\mathbf{c}) = T$, which implies that $\mathbf{c}$ is a level coverage. In addition, we have $\sum_{\ell\in[l]} c_\ell = \sum_{\ell\in I_1} x_{1\ell} + \sum_{\ell\in I_2} x_{2\ell} = 2 = k_1 + k_2$. This means that $\mathbf{c} = \overline{\mathbf{c}}$ and the first condition in Lemma 4.5 holds.

To see that the second condition in Lemma 4.5 also holds, we pick an arbitrary $t^* \in T$. By our calculation above, we have $c_{t^*} > 0$ (we have $c_t > 0$ for all $t \in T$). Consider any deviation $\langle D, \mathbf{x}\rangle$ and let the coverage resulting from this deviation be $\mathbf{c}'$ (assuming that the revenger, if any, sticks to their strategy in $\mathcal{CS}$, which suffices to thwart the deviation in this particular game as we will show). Since $\mathbf{c} = \overline{\mathbf{c}}$, for at least one target $t \in T$ it holds that $c'_t \le c_t$, i.e., its coverage will not increase after the deviation; indeed, if this is not the case, we would have $\sum_{t\in T} c'_t > \sum_{t\in T} c_t = k_1 + k_2$, which violates the budget constraint. Moreover, since $\mathbf{c}$ is level, we must also have $c'_t \le c_t$ for all $t \in \mathsf{BR}(\mathbf{c}')$. If this is not the case, we would have $c'_t > c_t$ and hence $U^a(\mathbf{c}', t) < U^a(\mathbf{c}, t)$ by monotonicity. It follows that $U^a(\mathbf{c}', t') < U^a(\mathbf{c}', t) < U^a(\mathbf{c}, t) = U^a(\mathbf{c}, t')$ for all $t' \notin \mathsf{BR}(\mathbf{c}')$ (where the last equality is by (8)), which then implies that $c'_{t'} > c_{t'}$. Consequently, $\sum_{t\in T} c'_t > \sum_{t\in T} c_t = k_1 + k_2$, which violates the budget constraint, too. Therefore, we have

$$U_1^d(\mathbf{c}', t) \le U_1^d(\mathbf{c}, t) = U_i^d(\mathbf{c}, t^*),$$

where $U_1^d(\overline{\mathbf{c}}, t) = U_i^d(\mathbf{c}, t^*)$ is due to the fact that $U_1^d(\mathbf{c}, t) = 0$ for all $t \in T$ by (7) and (8).

The"if" direction. Conversely, suppose that there exists a coalition structure $\mathcal{CS}$ in the 0^+-core which is not a grand coalition structure. Indeed, in this constructed game, if $\mathcal{CS}$ is not a grand coalition structure, it has to be in the form $\mathcal{CS} = \{\langle\{1\}, \mathbf{x}_1\rangle, \langle\{2\}, \mathbf{x}_2\rangle\}$. Let $\mathbf{c}$ be the coverage vector resulting from $\mathcal{CS}$. By Lemma 4.5, $\mathbf{c} = \overline{\mathbf{c}}$. Observe that the maximal level coverage $\overline{\mathbf{c}}$ must be unique in a canonical game. Hence, according to our calculation above, we have $c_\ell = \overline{c}_\ell = \frac{a_\ell}{\gamma}$ for each $\ell \in [l]$, by which $\overline{\mathbf{c}}$ is level (see (8)) and $\sum_{\ell\in[l]} c_\ell = k_1 + k_2 = 2$.

By (3), we now have $c_\ell = 1-(1-x_{1\ell})(1-x_{2\ell}) = \frac{a_\ell}{\gamma}$. Let $I_1 = \{\ell \in [l] : x_{1\ell} > 0\}$ and $I_2 = \{\ell \in [l] : x_{2\ell} > 0\}$. Note that we can establish the following chain of inequalities:

$$\begin{aligned} 2 = \sum_{\ell\in[l]} c_\ell &= \sum_{\ell\in[l]}(1-(1-x_{1\ell})(1-x_{2\ell})) \\ &= \sum_{\ell\in[l]}(x_{1\ell} + x_{2\ell} - x_{1\ell}\cdot x_{2\ell}) \\ &\le \sum_{\ell\in[l]}(x_{1\ell} + x_{2\ell}) \\ &\le k_1 + k_2 = 2, \end{aligned}$$

which must therefore all be equalities. In particular, for the transition in the third line to be an equality, it must be that $x_{1\ell}\cdot x_{2\ell} = 0$ for all ℓ; hence, $I_1 \cap I_2 = \varnothing$. For the transition in the last line to be an equality, it must be that $\sum_{\ell\in[l]} x_{i\ell} = k_i = 1$ for both $i = 1$ and 2.

Therefore, for each $\ell \in I_1$, we have $x_{2\ell} = 0$; hence, we have $c_\ell = x_{1\ell} = \frac{a_\ell}{\gamma}$, and

$$\sum_{\ell\in I_1} \frac{a_\ell}{\gamma} = \sum_{\ell\in I_1} x_{1\ell} = \sum_{\ell\in[l]} x_{1\ell} = 1,$$

which means that $\sum_{\ell\in I_1} a_\ell = \gamma$. Similarly, we also have $\sum_{\ell\in I_2} a_\ell = \gamma$. Since $I_1 \cap I_2 = \varnothing$, we conclude that I is a yes-instance.

A.3 Proof of Proposition 4.8

Assume in contradiction that there exist such $\mathcal{CS}, \mathcal{CS}'$. Let $\mathcal{CS}_\ell$ be a series of coalition structure, each one some ϵ_ℓ-Core, and $\epsilon_\ell \to 0$. Consider the deviation $D = ([n], \pi)$ which simulates the coalition structure $\mathcal{CS}'$. Let $\epsilon > 0$ be the minimal defender advantage of deviation D from $\mathcal{CS}$.

By limit definition, there exist $\ell_i \in \mathcal{N}$ such that for each $\ell \ge \ell_i$, $U_i^d(\mathcal{CS}_\ell) < U_i^d(\mathcal{CS}) + \epsilon/2$. Taking $\ell_0 = \max \ell_1, \dots, \ell_n$, we get that for each defender i, $U_i^d(\mathcal{CS}_\ell) < U_i^d(\mathcal{CS}) + \epsilon/2 \le U_i^d(\mathcal{CS}') - \epsilon/2$. This means that D is an $\epsilon/2$-successful deviation, for any coalition structure $\mathcal{CS}_\ell$ where $\ell \ge \ell_0$. However, for some large enough $\ell \ge \ell_0$, we have that $\epsilon_\ell < \epsilon/2$. Then considering $\mathcal{CS}_\ell$, D is an $\epsilon/2$ successful deviation, in contradiction to the fact that $\mathcal{CS}_\ell$ is in the ϵ_ℓ-Core.

A.4 Proof of Proposition 4.9

In this proof, we let to $U_i^d(\mathcal{CS}) := \lim_{\ell\to\infty} U_i^d(\mathcal{CS}_\ell)$ for brevity.

Assume in contradiction that there exist such $\mathcal{CS}, \mathcal{CS}', i$. Then consider a deviation of the grand coalition structure, with a strategy that simulates the strategy of $\mathcal{CS}'$, $D = ([n], \pi)$. Also, let $\mathcal{CS}_\ell \to \mathcal{CS}$, where each $\mathcal{CS}_\ell$ is in some weak ϵ_ℓ-Core, and $\epsilon_\ell \to 0$. We will show that D is a weak successful deviation, for some sub-series of $\mathcal{CS}_\ell$. This will lead to a contradiction to the fact that $\mathcal{CS}$ is in the weak 0^+-Core.

Indeed, let $\epsilon = U_i^d(\mathcal{CS}') - U_i^d(\mathcal{CS})$. By definition of the limit, there exist some ℓ_1 such that for any $\ell \ge \ell_1$, $|U_i^d(\mathcal{CS}) - U_i^d(\mathcal{CS}_\ell) < \epsilon/2$. Therefore,

$$U_i^d(\mathcal{CS}') > U_i^d(\mathcal{CS}) + \epsilon > U_i^d(\mathcal{CS}_\ell) + \epsilon/2$$

In order to have a contradiction, we must find infinitely many ℓ's such that for any defender $j \neq i$,

$$U_j^d(\mathcal{CS}') \geq U_j^d(\mathcal{CS}) \geq U_j^d(\mathcal{CS}_\ell)$$

This will make D a weak $\epsilon/2$ successful deviation from such $\mathcal{CS}_\ell$. Then, as $\epsilon_\ell \to 0$, there will be some sufficiently large ℓ_2 such that for any $\ell \geq \ell_2$, $\epsilon_\ell < \epsilon/2$. Taking $\ell_0 = \max \ell_1, \ell_2$ will give a contradiction to the fact that $\mathcal{CS}_{\ell_0}$ is in the weak ϵ_ℓ-Core.

We divide into cases. Consider the following set L:

$$L := \{\ell \in \mathbb{N} | \forall i \in [n] : U_i^d(\mathcal{CS}_\ell) \leq U_i^d(\mathcal{CS})\}$$

If this set is infinite, then consider the subseries $\{\mathcal{CS}_\ell\}_{\ell \in L}$. It is also a series of coalition structure, each in some weak ϵ'_ℓ-Core, and $\epsilon'_\ell \to 0$, $\mathcal{CS}'_\ell \to \mathcal{CS}$. Moreover, for any ℓ we have $U_j^d(\mathcal{CS}) \geq U_j^d(\mathcal{CS}'_\ell)$, and we get the contradiction we wanted.

Else, the set L is finite. This means that there exist ℓ_2, such that for any $\ell \geq \ell_2$, there exist a defender i_ℓ with a defender utility greater than its utility in the limit point. Now, as this happens infinitely many times, there is a defender i such that for infinitely many indices, this condition happens. Therefore, there is a sub series $L' \subseteq \mathbb{N}$ such that the utility of defender i is non-increasing.

However, in SSG this means that the utility of all defenders is non-increasing. Let us denote the series of coalition structures we got by $\mathcal{CS}''_\ell$. Then since the utility of each defender is non-increasing, we have that $\mathcal{CS}''_\ell$ is in the weak $\epsilon''_{\ell'}$-Core for any ℓ'. Hence, by considering the constant series of the coalition structures of itself, we get that $\mathcal{CS}''_\ell$ is in the 0^+-Core, for each ℓ.

Now, consider two consecutive coalition structures, $\mathcal{CS}''_\ell, \mathcal{CS}''_{\ell+1}$. As the utility of each defender can only decrease, there are two options. In the first option, there is at least one defender i who's utility is strictly decreasing, by $\delta > 0$. In that case, the grand coalition structure that simulates $\mathcal{CS}''_\ell$ is a weak δ successful deviation for all coalition structures $\mathcal{CS}''_{\ell'}$ where $\ell' > \ell$. By limit definition, there exist ℓ''_2 such that for any $\ell'' \geq \ell''_2$, $\epsilon''_\ell < \delta$. Taking $\ell''_0 = \max \ell''_1, \ell''_2$ will result in a coalition $\mathcal{CS}''_{\ell''_0}$ which is not in the weak ϵ''_{ℓ_0}-Core. This results in a contradiction.

In the second option, the utilities of all defenders are independent of ℓ. In that case, for any defender j we have an equality, namely $U_j^d(\mathcal{CS}) = U_j^d(\mathcal{CS}_\ell)$, which is enough for us to get our contradiction.

A.5 Proof of Lemma 5.3

First, we argue that when $v > -\infty$, there does not exists an ϵ-safe deviation $\mathbf{y} \in \mathcal{C}_{k_D}$ for D, such that $y_t \leq x_t$ for all t and $y_{t'} < x_{t'}$ for some $t' \in T$. Indeed, if this is the case, we can construct a strategy $\mathbf{z}$ such that

$$z_t = \begin{cases} y_t + \delta & \text{if } t \in T_{\mathbf{y}}^- \\ y_t & \text{if } t \in T_{\mathbf{y}}^+. \end{cases}$$

Note that $y_t < s_t^\epsilon \leq 1$ and $\sum_{t \in T} y_t < \sum_{t \in T} x_t \leq k_D$, so when $\delta > 0$ is sufficiently small, we have $z_t \in [0, 1]$ for all $t \in T$ and $\sum_{t \in T} z_t \leq k_D$, which means that $\mathbf{z} \in \mathcal{C}$. On the other hand, $T_{\mathbf{z}}^+ = T_{\mathbf{y}}^+$ and $T_{\mathbf{z}}^- = T_{\mathbf{y}}^-$, so by monotonicity of the utility function we have

$$\max_{t \in T_{\mathbf{z}}^-} U^a(\mathbf{z}, t) = \max_{t \in T_{\mathbf{y}}^-} U^a(\mathbf{z}, t) < \max_{t \in T_{\mathbf{y}}^-} U^a(y_t, t) = v,$$

which contradicts the assumption that the ϵ-safety value is v.

Now pick an arbitrary $t_1 \in T_{\mathbf{x}}^-$. By definition, $v = \max_{t \in T_{\mathbf{x}}^-} U^a(\mathbf{x}, t) \geq U^a(\mathbf{x}, t_1)$, so it remains to show that $v \leq U^a(\mathbf{x}, t_1)$. Suppose for the sake of contradiction that $v > U^a(\mathbf{x}, t_1)$. Then it is not hard to see that $\mathbf{x}$ will remain an optimal solution to (5) if we reduce x_{t_1} by a sufficiently small amount: indeed, this modification does not change $T_{\mathbf{x}}^-$, while it keeps $v \geq U^a(\mathbf{x}, t)$ for all $T_{\mathbf{x}}^-$. This contradicts our observation above.

Similarly, pick an arbitrary $t_2 \in T_{\mathbf{x}}^+$. If $v < U^a(\mathbf{x}, t_2)$, then we can reduce x_{t_2} by a small amount δ; while this may cause t_2 to become an element in $T_{\mathbf{x}}^-$ we can ensure that $v < U^a(\mathbf{x}, t_2)$ still hold by choosing a sufficiently small δ, so $\mathbf{x}$ remain optimal for (5), which is a contradiction. Thus, it must be that $v \leq U^a(\mathbf{x}, t_2)$.

Finally, to see (ii), note that by definition $x_t \geq s_t^\epsilon$ for all $t \in T_{\mathbf{x}}^+$. Suppose that $x_{t'} > s_{t'}^\epsilon$ for some $t' \in T_{\mathbf{x}}^+$. Then we can reduce $x_{t'}$ by a small amount while still keep $x_{t'} > s_{t'}^\epsilon$; after this modification, $\mathbf{x}$ will still be optimal for (5), which, too, contradicts our observation.

A.6 Proof of Lemma 5.4

We first prove the "if" direction.

The "if" direction. We show that the ϵ-safe strategy $\mathbf{x}$ is also an ϵ-successful strategy for coalition D. Suppose that D uses $\mathbf{x}$, the revengers' coalition R uses some strategy $\mathbf{y} \in \mathcal{C}_{k_R}$, and $\mathbf{x}$ and $\mathbf{y}$ jointly result in a coverage vector $\mathbf{c}$, i.e., $c_t = 1 - (1 - x_t)(1 - y_t)$ for each $t \in T$. We argue that no matter which $\mathbf{y}$ coalition R selects, every $i \in D$ will achieve a utility improvement of at least ϵ.

First, consider the first condition. Suppose that $v = -\infty$. Then by definition we have $x_t \geq s_t^\epsilon$ for all $t \in T$; hence, $c_t \geq x_t \geq s_t^\epsilon$. For each $i \in D$, let $t^* = \mathsf{br}_i(\mathbf{c})$; by monotonicity of the utility function and (4), we have

$$U_i^d(\mathbf{c}, t^*) \geq U_i^d(\mathbf{s}^\epsilon, t^*) \geq U_i^d(\mathbf{s}_i^\epsilon, t^*) \geq U_i^d(\mathcal{CS}) + \epsilon,$$

so $\mathbf{x}$ is indeed ϵ-successful for D.

Next, if one of the second and third conditions hold, we first argue that $\mathsf{BR}(\mathbf{c}) \cap T_{\mathbf{x}}^- = \varnothing$. Indeed, suppose for the sake of contradiction that there exists $t' \in \mathsf{BR}(\mathbf{c}) \cap T_{\mathbf{x}}^+$. We can establish the following inequality for all $t \in T$:

$$U^a(\mathbf{c}, t) \leq U^a(\mathbf{c}, t') \leq U^a(\mathbf{x}, t') = v,$$

where the first transition is due to $t' \in \mathsf{BR}(\mathbf{c})$, the second is due to $c_{t'} \geq x_{t'}$, and the third follows by Lemma 5.3 (i). Expanding $U^a(\mathbf{c}, t)$ and rearrange the terms, we obtain that

$$c_t \geq \frac{r^a(t) - v}{r^a(t) - p^a(t)} = \gamma_t.$$

Moreover, expanding c_t using $c_t = 1-(1-x_t)(1-y_t)$ gives

$$y_t \geq \frac{\gamma_t - x_t}{1-x_t} \geq \frac{\gamma_t - s_t^\epsilon}{1-s_t^\epsilon}.$$

Therefore, if the second condition holds, we have $\gamma_{t'} > 1$ for some t' and it follows by the above equation that $y_{t'} > 1$, which contradicts the assumption that $\mathbf{y}$ is a feasible strategy. If the third condition hold, we have $\sum_{t\in T} y_t \geq \sum_{t\in T_\mathbf{x}^+} \frac{\gamma_t - s_t^\epsilon}{1-s_t^\epsilon} > k_R$, which means that $\mathbf{y}$ violates the budget constraint, so it cannot be a feasible strategy, either — a contradiction, too.

As a result, $\mathsf{BR}(\mathbf{c}) \cap T_\mathbf{x}^- = \varnothing$, which implies that $t^* := \mathsf{br}_i(\mathbf{c}) \in T_\mathbf{x}^+$ for every $i \in D$. By definition, $x_{t^*} \geq s_{t^*}^\epsilon$, so we have $c_{t^*} \geq x_{t^*} \geq s_{t^*}^\epsilon \geq s_{it^*}^\epsilon$ and hence, by monotonicity of the utility function and (4), it follows that

$$U_i^d(\mathbf{c}, t^*) \geq U_i^d(\mathbf{s}_i^\epsilon, t^*) \geq U_i^d(\mathcal{CS}) + \epsilon,$$

so $\mathbf{x}$ is ϵ-successful, too.

The "only if" direction. Suppose for the sake of contradiction that there is an ϵ-successful deviation strategy $\mathbf{z}$ for D, but it holds that $v > -\infty$, $\gamma_t \leq 1$ for all $t \in T_\mathbf{x}^+$, and $\sum_{t\in T_\mathbf{x}^+} \frac{\gamma_t - s_t^\epsilon}{1-s_t^\epsilon} \leq k_R$.

We first argue that $z_t \geq x_t$ for all $t \in T$.

Claim 1. *$z_t \geq x_t$ for all $t \in T$.*

Proof. Suppose this is not the case, we let $Q = \{t \in T : z_t < x_t\}$ so $Q \neq \varnothing$; we let $t^* \in T$ be a target that maximizes $U^a(\mathbf{z}, t)$ among all $t \in Q$. We show that the revenger's coalition can take the following strategy $\mathbf{y}$ to thwart the deviation of D: for each $t \in T$, we let

$$y_t = \begin{cases} \frac{\gamma_t - s_t^\epsilon}{1-s_t^\epsilon} & \text{if } t \in T_\mathbf{x}^+ \setminus \{t^*\} \\ 0 & \text{if } t \in T_\mathbf{x}^- \cup \{t^*\}. \end{cases}$$

Indeed, $\mathbf{y}$ is a feasible strategy for R: By assumption, $\gamma_t > 1$ for all $t \in T$, so $\frac{\gamma_t - s_t^\epsilon}{1-s_t^\epsilon} \leq 1$ and $y_t \leq 1$ for all t; By Lemma 5.3 (i), we have $U^a(\mathbf{x}, t) \geq v$ for all t, so expanding this inequality and using Lemma 5.3 (ii) we get $\frac{r^a(t)-v}{r^a(t)-p^a(t)} \geq x_t = s_t^\epsilon$ for all $t \in T_\mathbf{x}^+$, which means that $\gamma_t \geq s_t^\epsilon$ and in turn $y_t \geq 0$ for all t; Finally, the assumption that $\sum_{t\in T_\mathbf{x}^+} \frac{\gamma_t - s_t^\epsilon}{1-s_t^\epsilon} \leq k_R$ means that $\mathbf{y} \in \mathcal{C}_{k_R}$.

When coalition D uses $\mathbf{z}$, the revengers' coalition using $\mathbf{y}$ results in coverage $c_t = 1-(1-z_t)(1-y_t)$ for each target t. To see that $\mathbf{y}$ is effective in thwarting the deviation, we first show that $\mathbf{y}$ results in $t^* \in \mathsf{BR}(\mathbf{c})$, i.e., $U^a(\mathbf{c}, t^*) \geq U^a(\mathbf{c}, t)$ for all $t \in T$. We analyse the value of $U^a(\mathbf{c}, t)$ for different t.

- First of all, since $v > -\infty$, we have $U^a(\mathbf{x}, t^*) \geq v$ by Lemma 5.3 (i); since $y_{t^*} = 0$ and $z_{t^*} < x_{t^*}$, we have

$$c_{t^*} = 1-(1-z_{t^*})(1-y_{t^*}) = z_{t^*} < x_{t^*},$$

which, by monotonicity of the utility function, means that

$$U^a(\mathbf{c}, t^*) = U^a(\mathbf{z}, t^*) > U^a(\mathbf{x}, t^*) \geq v. \tag{9}$$

- For all $t \in T_\mathbf{x}^+ \setminus (Q \cup \{t^*\})$, we have $y_t = \frac{\gamma_t - s_t^\epsilon}{1-s_t^\epsilon}$ and hence,

$$\begin{aligned} c_t &= 1-(1-z_t)(1-y_t) \\ &\geq 1-(1-x_t)(1-y_t) \\ &= 1-(1-s_t^\epsilon)\left(1 - \frac{\gamma_t - s_t^\epsilon}{1-s_t^\epsilon}\right) = \gamma_t, \end{aligned}$$

where we use $z_t \geq x_t$ as $t \notin Q$; this further implies that

$$U^a(\mathbf{c}, t) \leq U^a(\gamma_t, t) = v. \tag{10}$$

- For all $t \in T_\mathbf{x}^- \setminus (Q \cup \{t^*\})$, we have $y_t = 0$, so $c_t = 1-(1-z_t)(1-y_t) = z_t \geq x_t$ and hence,

$$U^a(\mathbf{c}, t) \leq U^a(\mathbf{x}, t) = v, \tag{11}$$

where $U^a(\mathbf{x}, t) = v$ due to Lemma 5.3 (i).

- For all $t \in Q$, by the choice of t^* we have $U^a(\mathbf{z}, t) \leq U^a(\mathbf{z}, t^*)$, so it follows that

$$U^a(\mathbf{c}, t) \leq U^a(\mathbf{z}, t) \leq U^a(\mathbf{z}, t^*) = U^a(\mathbf{c}, t^*), \tag{12}$$

where we also use the fact that $c_t = 1-(1-z_t)(1-y_t) \geq z_t$ and the first transition of (9).

Thus, combining (9)–(12) gives $U^a(\mathbf{c}, t) \leq U^a(\mathbf{c}, t^*)$ for all $t \in T$, as we desire; we have $t^* \in \mathsf{BR}(\mathbf{c})$. Since

$$c_{t^*} = z_{t^*} < x_{t^*} \leq s_{t^*}^\epsilon = s_{it^*}^\epsilon$$

for some $i \in D$, we have

$$\begin{aligned} U_i^d(\mathbf{c}, \mathsf{br}_i(\mathbf{c})) &\leq U_i^d(\mathbf{c}, t^*) \\ &< U_i^d(s_{it^*}^\epsilon, t^*) = U_i^d(\mathcal{CS}) + \epsilon, \end{aligned} \tag{13}$$

where the last equality holds by (4) (note that now we have $s_{it^*}^\epsilon > 0$, so $s_{it}^\epsilon = \frac{U_i^d(\mathcal{CS})+\epsilon-p_i^d(t^*)}{r_i^d(t^*)-p_i^d(t^*)}$). In other words, defender i will not benefit from the deviation by more than ϵ, which contradicts the assumption that $\mathbf{z}$ is an ϵ-successful deviation of D. □

Claim 2. *$z_t > x_t$ for all $t \in T_\mathbf{x}^-$.*

Proof. Similarly to the proof of Claim 1, suppose that this is not the case and let $Q = \{t \in T_\mathbf{x}^- : z_t = x_t\}$; given Claim 1, we have $Q \neq \varnothing$. Pick an arbitrary $t^* \in Q$. We show that the revenger's coalition can take the following strategy $\mathbf{y}$ to thwart the deviation of D: for each $t \in T$, we let

$$y_t = \begin{cases} \frac{\gamma_t - s_t^\epsilon}{1-s_t^\epsilon} & \text{if } t \in T_\mathbf{x}^+ \\ 0 & \text{if } t \in T_\mathbf{x}^-. \end{cases}$$

Let $\mathbf{c}$ be the coverage vector resulting from the revengers' coalition using the above $\mathbf{y}$. For all $t \in T_\mathbf{x}^+$ and all $t \in T_\mathbf{x}^-$, the same arguments for (10) and (11) can be used to show that

$$U^a(\mathbf{c}, t) \leq v \quad \text{for all } t \in T.$$

Moreover, since $t^* \in Q \subseteq T_\mathbf{x}^-$, we have $z_{t^*} = x_{t^*}$ and $y_{t^*} = 0$; hence, $c_{t^*} = 1-(1-z_{t^*})(1-y_{t^*}) = x_{t^*}$ and

$$U^a(\mathbf{c}, t^*) = U^a(\mathbf{x}, t^*) = v$$

by Lemma 5.3 (i).

Therefore, $U^a(\mathbf{c}, t^*) \geq U^a(\mathbf{c}, t)$ for all $t \in T$, which means that $t^* \in \mathsf{BR}(\mathbf{c})$. The same argument for (13) can be used to show that $U_i^d(\mathbf{c}, \mathsf{br}_i(\mathbf{c})) < U_i^d(\mathcal{CS}) + \epsilon$, and hence we obtain a contradiction with the assumption that $\mathbf{z}$ is ϵ-successful for D. □

Now we have shown that $z_t \geq x_t$ for all $t \in T$ and $z_t > x_t$ for all $t \in T_{\mathbf{x}}^-$. Observe that

$$T_{\mathbf{z}}^- := \{t \in T : z_t < s_t^\epsilon\} \subseteq T_{\mathbf{x}}^-,$$

which gives

$$\max_{t \in T_{\mathbf{z}}^-} U^a(\mathbf{z}, t) \leq \max_{t \in T_{\mathbf{x}}^-} U^a(\mathbf{z}, t) \leq \max_{t \in T_{\mathbf{x}}^-} U^a(\mathbf{x}, t),$$

where the second transition is due to the fact that $z_t \geq x_t$ for all $t \in T$. In other words, $\mathbf{z}$ is a strictly better solution to the optimization problem in (5), which contradicts the assumption that $\mathbf{x}$ is an ϵ-safe deviation strategy for D.

A.7 Proof of Theorem 5.6

Let Θ be the set of defender types (so $\lambda = |\Theta|$), let N_θ be the set of defenders of each type $\theta \in \Theta$, and let θ_i be the type of each defender $i \in [n]$. We first argue that if there is an ϵ-successful deviation strategy $\mathbf{x}$ for coalition D, then $\mathbf{x}$ is also ϵ-successful for the following coalition

$$\widehat{D} := \{i \in [n] : \theta_i = \theta_j \text{ for some } j \in D\}.$$

Indeed, since $D \subseteq \widehat{D}$, we have $k_{\widehat{D}} \geq k_D$, which means that $\mathbf{x} \in \mathcal{C}_{k_D} \subseteq \mathcal{C}_{K_{\widehat{D}}}$ is a feasible strategy of coalition $\widehat{D}$. By definition, the fact that $\mathbf{x}$ is ϵ-successful for D means that for any $i \in D$ and any strategy $\mathbf{y} \in \mathcal{C}_{k_R}$ of the revengers' coalition $R = [n] \setminus D$, it holds that

$$U_i^d(\mathcal{CS}') \geq U_i^d(\mathcal{CS}) + \epsilon, \tag{14}$$

where $\mathcal{CS}' = \{\langle D, \mathbf{x}\rangle, \langle R, \mathbf{y}\rangle\}$. Consider an arbitrary defender $i \in \widehat{D}$ and any strategy $\hat{\mathbf{y}} \in \mathcal{C}_{k_{\widehat{R}}}$ of the revengers' coalition $\widehat{R} = [n] \setminus \widehat{D}$; let $\mathcal{CS}'' = \left\{\langle D, \mathbf{x}\rangle, \langle \widehat{R}, \hat{\mathbf{y}}\rangle\right\}$. We have $\widehat{R} \subseteq R$, so $\hat{\mathbf{y}} \in \mathcal{C}_{k_{\widehat{R}}} \subseteq \mathcal{C}_{k_R}$; it then follows by (14) that

$$U_j^d(\mathcal{CS}'') \geq U_j^d(\mathcal{CS}) + \epsilon$$

for all $j \in D$. Now, if we pick a defender $j \in D$ with $\theta_j = \theta_i$, the utility functions U_i^d and U_j^d will be the same, so this immediately gives $U_i^d(\mathcal{CS}'') \geq U_i^d(\mathcal{CS}) + \epsilon$, which implies that $\mathbf{x}$ is ϵ-successful for $\widehat{D}$.

Given the above observation, we only need to consider coalitions D such that for every $\theta \in \Theta$ either $D \cap N_\theta = N_\theta$ or $D \cap N_\theta = \varnothing$, and check if there is an ϵ-successful strategy for them. There are 2^λ such coalitions and for each of them this can be done in time $\mathrm{poly}(m, n)$ according to Theorem 5.5.

B Examples

B.1 Example of an SSG with an empty weak 0^+core

First we introduce a new notation. We say that defender i prefers target t over target t', and write $t' \prec_i t$, if $r_i^d(t') < p_i^d(t)$. In this case, no matter what strategy is taken, defender i will always prefer the attacker to attack target t rather than target t'.

Example B.1. Consider an SSG with 3 targets, t_1, t_2, t_3, and 3 defenders. Defenders $1, 2$ have no security resources, $k_1 = k_2 = 0$, while defender 3 has $k_3 = 1$ security resource. Assume further that $p^a(t_1) = p^a(t_2) = p^a(t_3) = 0$, $r^a(t_1) = r^a(t_2) = 1$, but $r^a(t_3) = 0.5$.

Now, let's say that the defenders' utility functions are as follows: Defender 1 has the preference $t_3 \prec t_2 \prec t_1$. Defender 2 has preference $t_3 \prec t_1 \prec t_2$. For defender 3, the preference is $t_3 \prec t_1 = t_2$. We claim that in this case, the strict core is empty.

Proposition B.2. *Example B.1 is an example of an SGG where the strict 0^+ core is empty. That is, there is no sequence of coalition structures $\{\mathcal{CS}_\ell\}_{\ell=1}^\infty$ in the strict ϵ_ℓ-core, with $\epsilon_\ell \to 0$.*

Proof. Since coalition structures in the strict core must form a maximal coverage by Lemma 4.5, any coalition structure $\mathcal{CS}$ in the core must correspond to the coverage vector $(1/2, 1/2, 0)$. Therefore, the only thing left to decide is the partition in the coalition structure.

We will show that for any coalition structure, there always exist a weak ϵ deviation with $\epsilon \geq \delta_0$ for some fixed δ_0. This in turn means that there cannot be a sequence of coalition structures in the strict ϵ_ℓ Core, where $\epsilon_\ell \to 0$, as this will imply the existence of a coalition structure $\mathcal{CS}_{\ell_0}$ in the strict ϵ_{ℓ_0}-core, where $\epsilon_{\ell_0} < \delta_0$, resulting in a contradiction. We let $\delta_1 = p_1^d(t_1) - r_1^d(t_2)$, $\delta_2 = p_2^d(t_2) - r_1^d(t_1)$ and $\epsilon_3 = p_3^d(t_1) - r_3^d(t_3)$

Indeed, let $\mathcal{CS}$ be some coalition structure. If $t_2 \in \mathsf{BR}(\mathbf{c})$, let the coverage of target t_2 be denoted by c. Note that since the total amount of resource is up to 1, it must be that $c \leq 1/2$. Assume for a moment that $c < 1/2$. The utility of defender 1 is therefore $U_1^d(\mathcal{CS}) = c \cdot r_1^d(t_2) + (1-c) \cdot p_1^d(t_2) \leq r_1^d(t_2)$. The utility of defender 3 is $U_3^d(\mathcal{CS}) = c \cdot r_3^d(t_2) + (1-c) \cdot p_3^d(t_2)$.

Now consider the following deviation of defenders $1, 3$: $\mathbf{c}^D = (c, 1/2)$. Since defender 2 has no resources, his response is to do nothing, that is $\mathbf{c}^R = (0, 0)$. Therefore, the coverage vector after the deviation is $\mathbf{c}' = \mathbf{c}^D$, and thus $\mathsf{BR}' = \mathsf{BR}(\mathbf{c}') = \{t_1\}$. This means that defender 1 will get at-least $p_1^d(t_1) \geq U_1^d(\mathcal{CS}) + \delta_1$. Defender 3 will get the same utility $U_3^d(\mathcal{CS})$ he had before. Therefore, in this case we have a weak δ_1 successful deviation.

Similarly, in the case where $t_1 \in \mathsf{BR}(\mathbf{c})$, we can find a weak δ_2 successful deviation. Therefore, there always exist a weak ϵ successful deviation, with $\epsilon \geq \delta_0 = \min(\delta_1, \delta_2)$, as desired.

Therefore, the only case left to deal with is the coverage vector $(1/2, 1/2, 0)$. Assume in contradiction it is in the strict 0^+-core, and let $\{\mathcal{CS}_\ell\}_{\ell=1}^{\infty}$ where each $\mathcal{CS}_\ell$ is in the strict ϵ_ℓ-core, with $\epsilon_\ell \to 0$ and $\mathcal{CS}_\ell \to \mathcal{CS}$.

For some ℓ_0, we have that for every $\ell \geq \ell_0$, $\epsilon_\ell \leq \delta_0$. In that case, as we have shown, $\mathcal{CS}_\ell$ must form a coverage $(1/2, 1/2, 0)$. However in this case, consider the deviation of defender 3, $(0, 0, 0)$. Since defenders $1, 2$ have no resources, this will be the coverage vector after the deviation as well. In this case, $\mathsf{BR}(\mathbf{c}) = \{1, 2\}$, and therefore since $3 \prec_3 1, 2$, defender 3 will get an improvement of utility, from $r_3^d(t_3)$, to $p_3^d(t_1) = p_3^d(t_2)$. This is an ϵ_3 successful deviation. Let $\ell \geq \ell_0$ be such that $\epsilon_\ell < \epsilon_3$, and we will get a contradiction to the fact that $\mathcal{CS}_\ell$ is in the ϵ_ℓ-core.

Therefore, taking $\delta_0 = \min(\delta_1, \delta_2, \delta_3)$ gives us the desired property.

□