目录

1 安全博弈论 1

1.1 研究背景 1

1.2 安全博弈模型与均衡 3

1.2.1 Stackelberg 均衡 4

1.2.2 均衡求解 6

1.2.3 Stackelberg 安全博弈模型及求解 8

1.2.4 安全博弈实例 10

1.3 复杂环境下的安全博弈 12

1.3.1 信息不完全与不确定性 12

1.3.2 复杂策略空间的处理 16

1.3.3 动态安全博弈 18

1.4 实际应用与成功案例 21

1.4.1 重要基础设施保护 21

1.4.2 交通系统安保调度 23

1.4.3 打击环境资源犯罪与城市犯罪 24

1.4.4 打击犯罪网络 25

1.4.5 其他应用 25

1.5 热点与未来方向 26

1.5.1 研究热点 26

1.5.2 挑战和未来研究方向 28
1.5.3 未来应用领域 30

参考文献 **33**

1 安全博弈论

安全是二十一世纪全人类面临的重要问题。伴随近年来国际局势的深刻变化，全球各国安全事件频发，安全形势日趋严峻。安全场景往往极具策略性，潜在的攻击者精心策划并采取高度优化的攻击方式，破坏性极强。防范与应对这些问题需要充分考虑到这些策略性的行为，并在此基础上提出高效的安保资源分配方案。在此背景下，安全博弈论应运而生，借助经典博弈论模型与前沿的算法、优化技术对安全对策的求解与优化问题加以解决。这些技术的成功应用使得安全博弈论迅速成为博弈论在实际应用上最具影响力的成功案例之一。基于安全博弈模型的算法及自动化决策系统被陆续布署于重要基础设施、交通系统、以及自然资源的保护任务中，并发挥重要作用。本章将回顾安全博弈论在实际应用上的成功案例，并在此基础上对安全博弈论基本模型与求解技术进行介绍，内容涉及均衡概念、策略优化算法设计以及现有文献中丰富的安全博弈场景与模型。同时本章也将对安全博弈研究面临的机遇和挑战进行讨论。

1.1 研究背景

保护关键公共基础设施和目标，如机场、港口、历史名胜、电力设施、大型公共活动，甚至珍稀动物和自然资源等，是各国安保部门面临的共同挑战。有限的安保资源往往意味着安全机构难以为所有保护对象提供全时段的保护，在资源调度上捉襟见肘。此外，潜在的攻击者（如恐怖分子、罪犯、偷猎者等）也可以通过事先侦察去发现安全机构的安保策略的模式和弱点，据此选择最优的攻击策略。在这些情况下，在安保策略中引入随机性是一种平衡安保资源分布、降低对手预判能力的有效手段，如随机选择巡逻路线、保护对象、检测目标等。然而，如何有效地引入这些随机性并非易

事，有诸多问题需要解决。关键问题之一在于不同的随机策略对攻击者的攻击行为会产生不同的影响，而攻击者会根据其对安保部门随机策略的观察和认知优化其攻击策略，从而又影响到安保策略的效果。在这个问题上，博弈论提供了一个恰当的数学模型来建模攻击者的策略性反馈，基于 Stackelberg 博弈模型的安全博弈模型应运而生。尽管 Stackelberg 博弈模型早在上世纪三十年代就被提出 [53, 54]，从计算的角度对该模型进行研究仍然是相对新兴的课题，是在 Conitzer 和 Sandholm 于 2006 年的经典论文 [10] 发表后迅速发展起来的。而安全博弈论初期研究的主要参与者包括南加利福尼亚大学 Milind Tambe 教授领导的 TEAMCORE 研究小组以及杜克大学 Vicent Conitzer 教授领导的研究小组。在他们的影响下越来越多的学者正参与到这项研究，使其成为当前人工智能和多智能体系统领域的研究热点之一。基于安全博弈论的算法和系统也陆续被开发、布署到众多的现实安保场景，包括洛杉矶国际机场安检人员调度、美联邦空中警察调度、美国海岸警卫队巡逻路线制定等极具影响力的应用。本章稍后将对这些代表性的应用进行更为详细的介绍。

经典的 Stackelberg 博弈在一个*领导者*（leader）和一个*跟随者*（follower）之间展开。这些参与者可以是个人或团体 (如警察部队)。每位参与者有一个可以执行的行动集合，即*纯策略*（pure strategy）集。而*混合策略*（mixed strategy）则允许参与者以一定的概率选择不同纯策略。参与者使用的策略的组合决定了所有参与者的收益。通常，博弈模型的实例给定了参与者在不同纯策略组合下的收益。而在混合策略下需要考虑的是参与者在纯策略组合概率分布上的期望收益。在 Stackelberg 博弈模型中，领导者会优先行动，选定并执行一个混合策略。跟随者随后观察到领导者的策略，然后针对该混合策略反馈一个最优的策略以最大化自身的收益。通常的假设是跟随者只能观察到领导者的混合策略，而非从混合策略中采样出的纯策略。在安全领域，安保部门和潜在的攻击者之间的互动能很自然地被抽象为一个 Stackelberg 博弈。安保部门作为博弈的领导者事先选定并执行一套安保策略，而攻击者作为博弈的跟随者在观察到安保策略之后选择最优的攻击方案。例如，安保部门随机地将安保人员分配到机场的检查点或者将空中警察随机地安排到航班上。攻击者则根据这些目标（检查点或航班）被保护的概率选择期望收益最高的目标进行攻击。

表1.1展示了一个简单的安全博弈实例的收益矩阵。该实例包含两个目标，攻击者可以攻击其中任意一个目标，而安保部门拥有一个安保资源，可以将其安排到任意一个目标进行保护。假如安全部门采用纯策略，选择一个确定性的目标进行保护，那么攻击者在观测到该策略以后总是能准确地选择未被保护的目标进行攻击，攻击成功并取得一个正的收益。而如果安保部门采用混合策略对目标进行保护（例如以 50% 的概率分别保护两个目标），其总能以一定的概率挫败攻击者的攻击。如何计算安保部门的最优混合策略是安全博弈论关注的核心问题。

表 1.1 安全博弈收益矩阵实例（两个目标，一个资源）。矩阵中的每一对数值分别是安保部门和攻击者在对应的纯策略组合下的收益。

		攻击者	
		攻击目标 1	攻击目标 2
安保部门	保护目标 1	$5, -3$	$-1, 1$
	保护目标 2	$-5, 5$	$2, -1$

1.2 安全博弈模型与均衡

本节我们从 Stackelberg 博弈模型出发介绍安全博弈论涉及到的基本概念、模型及求解算法。Stackelberg 博弈模型源于上世纪三十年代由同名数理经济学家海因里希 · 弗赖尔 · 冯 · 施塔克尔贝格（Heinrich Freiherr von Stackelberg ）提出的模型 [53, 54]。Stackelberg 博弈是在一个领导者和一个跟随者展开的序贯博弈。博弈中领导者事先固定其策略，而跟随者在知悉领导者策略的基础上其进行反馈。假设领导者和跟随者的纯策略集分别包含 m 和 n 个纯策略。[1] 那么博弈双方的收益可通过两个 m 行 n 列的矩阵 $u^{\mathsf{L}} \in \mathbb{R}^{m\times n}$ 及 $u^{\mathsf{F}} \in \mathbb{R}^{m\times n}$ 给定。其中 u^{L} 和 u^{F} 的上标分别对应领导者（leader）和跟随者（follower）的首字母。这种以收益矩阵形式给出的博弈模型常常也被称作*双矩阵博弈*（bi-matrix games）。如果领导者使用其纯策略集中的第 i 个策略而跟随者使用其第 j 个策略，其结果是领导者和跟随者获得位于各自收益矩阵第 i 行第 j 列的收益值，即 $u^{\mathsf{L}}(i,j)$ 和 $u^{\mathsf{F}}(i,j)$。换言之，领导者和跟随者在该双矩阵博弈中分别是*行玩家*（row player）和*列玩家*（column player）。

我们考虑博弈双方可以使用更为一般的混合策略的情形。混合策略是一个在纯策略集上的概率分布。[2] 以领导者的混合策略为例，我们将其表示为一个 m 维向量 $\mathbf{x} = (x_1, \ldots, x_m) \in \Delta^{m-1}$。[3] 方便起见，我们用

$$u^{\mathsf{L}}(\mathbf{x}, j) := \mathbb{E}_{i\sim\mathbf{x}} u^{\mathsf{L}}(i,j) = \sum_{i=1}^{m} x_i \cdot u^{\mathsf{L}}(i,j)$$

表示领导者在使用混合策略 $\mathbf{x}$ 时跟随者的第 j 个纯策略给领导者带来的期望收益。类似地，我们用 $u^{\mathsf{F}}(\mathbf{x}, j) = \mathbb{E}_{i\sim\mathbf{x}} u^{\mathsf{F}}(i,j)$ 表示同样情况下跟随者的期望收益。由于每个纯策略 j 同时也是一个特殊的混合策略，我们时常将 j 视为一个等价的向量 $\mathbf{y}$，其中

[1] 我们仅考虑 m 和 n 有穷的情况。

[2] 当我们未指明地提及“策略”时我们一般指“混合策略”。

[3] Δ^{m-1} 表示 $m-1$ 维的单纯形，即 $\Delta^{m-1} := \{\mathbf{x} \in \mathbb{R}^m : \sum_{i=1}^m x_i = 1$，且 $x_i \geq 0\ \forall i = 1, \ldots, m\}$。$\Delta^{m-1}$ 包含了所有 $\{1, \ldots, m\}$ 上的概率分布。

$y_j = 1$ 且 $y_\ell = 0\ \forall \ell \in \{1,\dots,n\} \setminus \{j\}$。

对任一领导者纯策略 $\mathbf{x} \in \Delta^{m-1}$，我们将跟随者的第 j 个纯策略称作其对 $\mathbf{x}$ 的最优反馈，当且仅当 $u^{\mathsf{F}}(\mathbf{x}, j) = \max_{\ell \in \{1,\dots,n\}} u^{\mathsf{F}}(\mathbf{x}, \ell)$。由于最优反馈可能存在多个，我们同时也定义以下对 $\mathbf{x}$ 的最优（纯策略）反馈集BR（Best Response）：

$$\mathsf{BR}(\mathbf{x}) := \arg \max_{\ell \in \{1,\dots,n\}} u^{\mathsf{F}}(\mathbf{x}, \ell)。$$

同样地，以上定义可扩展到跟随者也使用混合策略情形。给定任意跟随者混合策略 $\mathbf{y} \in \Delta^{n-1}$，我们令 $u^{\mathsf{L}}(\mathbf{x}, \mathbf{y}) := \mathbb{E}_{j\sim\mathbf{y}} u^{\mathsf{L}}(\mathbf{x}, j) = \sum_{j=1}^{n} y_j \cdot u^{\mathsf{L}}(\mathbf{x}, j)$ 表示跟随者的期望收益，令 $\Delta\mathsf{BR}(\mathbf{x}) := \arg\max_{\mathbf{y}\in\Delta^{n-1}} u^{\mathsf{F}}(\mathbf{x}, \mathbf{y})$ 表示跟随者对 $\mathbf{x}$ 的最优混合策略反馈集。事实上，该定义扩展对基本的 Stackelberg 博弈模型是非必要的：不失一般性，我们可以假设跟随者总是使用纯策略。其原因在我们接下来定义均衡概念后将得以明晰。

1.2.1 Stackelberg 均衡

Stackelberg 均衡描述了博弈双方均使用最优策略的情形，是 Stackelberg 博弈的基本概念。这里的最优性考虑到博弈双方行动的先后性：均衡中跟随者使用的策略 $\mathbf{y}$ 是对领导者策略 $\mathbf{x}$ 直接的最优反馈，而领导者策略 $\mathbf{x}$ 的最优性考虑到跟随者总是会对 $\mathbf{x}$ 进行最优反馈的事实。严格定义 Stackelberg 均衡需要我们明确当跟随者具有多个最优反馈时（即 $\Delta\mathsf{BR}(\mathbf{x})$ 包含多个元素时）如何“打破平局”的问题。悲观假设和乐观假设是两种最为自然的“平局决胜规则”。悲观假设认为跟随者总是会从最有反馈集 $\Delta\mathsf{BR}(\mathbf{x})$ 中选择对于领导者来说最差的反馈；反之，乐观假设认为跟随者总是作出对于领导者最有利的选择。采用这两种不同的假设，我们可以定义弱、强两种 Stackelberg 均衡。强 *Stackelberg 均衡*（strong Stackelberg equilibrium）采用乐观平局决胜规则，而弱 *Stackelberg 均衡*（weak Stackelberg equilibrium）采用悲观平局决胜规则。严格定义如下。

定义 1.2.1 (强/弱 Stackelberg 均衡)**.** 假设 $\mathbf{x} \in \Delta^{m-1}$ 与 $\mathbf{y} \in \Delta^{n-1}$ 分别是领导者与跟随者的混合策略，且 $\mathbf{y} \in \Delta\mathsf{BR}(\mathbf{x})$。当以下条件满足时，策略组合 $(\mathbf{x}, \mathbf{y})$ 被称作是一个强 Stackelberg 均衡：

$$u^{\mathsf{L}}(\mathbf{x}, \mathbf{y}) = \max_{\mathbf{x}'\in\Delta^{m-1}} \max_{\mathbf{y}'\in\Delta\mathsf{BR}(\mathbf{x}')} u^{\mathsf{L}}(\mathbf{x}', \mathbf{y}')。$$

当以下条件满足时，策略组合 $(\mathbf{x}, \mathbf{y})$ 被称作是一个弱 Stackelberg 均衡：

$$u^{\mathsf{L}}(\mathbf{x}, \mathbf{y}) = \max_{\mathbf{x}'\in\Delta^{m-1}} \min_{\mathbf{y}'\in\Delta\mathsf{BR}(\mathbf{x}')} u^{\mathsf{L}}(\mathbf{x}', \mathbf{y}')。$$

由上述定义不难看出，所有的强 Stackelberg 均衡对应相等的领导者收益；[4] 所有的弱 Stackelberg 均衡也具备相同的性质。更进一步，我们可将均衡中跟随者的反馈限制为纯策略。下列定理表明该假设不失一般性。

定理 1.2.1. *假设策略组合 $(\mathbf{x}, \mathbf{y}) \in \Delta^{m-1} \times \Delta^{n-1}$ 是一个强/弱 Stackelberg 均衡，那么总是存在一个跟随者纯策略 $j \in \{1, \dots, n\}$ 使得策略组合 $(\mathbf{x}, j)$ 也是一个强/弱 Stackelberg 均衡。*

证明. 我们仅证明强 Stackelberg 均衡的情况。弱均衡情况下的结论可用相同方法证得。由于 $(\mathbf{x}, \mathbf{y}) \in \Delta^{m-1} \times \Delta^{n-1}$ 是强 Stackelberg 均衡，根据定义1.2.1，

$$\mathbf{y} \in \arg \max_{\mathbf{y}' \in \Delta\mathsf{BR}(\mathbf{x})} u^{\mathsf{L}}(\mathbf{x}, \mathbf{y}')。 \tag{1.1}$$

因此 $\mathbf{y} \in \Delta\mathsf{BR}(\mathbf{x})$。令 $\mathrm{supp}(\mathbf{y}) := \{j \in \mathsf{BR}(\mathbf{x}) : y_j > 0\}$ 为 $\mathbf{y}$ 的支持集。根据(1.1)以及 $\mathrm{supp}(\mathbf{y}) \subseteq \mathsf{BR}(\mathbf{x}) \subseteq \Delta\mathsf{BR}(\mathbf{x})$，下列不等式对所有 $j \in \mathrm{supp}(\mathbf{y})$ 均成立：

$$u^{\mathsf{L}}(\mathbf{x}, \mathbf{y}) \geq u^{\mathsf{L}}(\mathbf{x}, j)。 \tag{1.2}$$

同时，下列不等式也必须成立

$$u^{\mathsf{L}}(\mathbf{x}, \mathbf{y}) \leq u^{\mathsf{L}}(\mathbf{x}, j)\,, \tag{1.3}$$

否则由 $u^{\mathsf{L}}(\mathbf{x}, \mathbf{y}) > u^{\mathsf{L}}(\mathbf{x}, j)$ 我们可构造一个新的跟随者策略 $\mathbf{y}' \in \Delta\mathsf{BR}(\mathbf{x})$ 使得 $u^{\mathsf{L}}(\mathbf{x}, \mathbf{y}') > u^{\mathsf{L}}(\mathbf{x}, \mathbf{y})$；该不等式同(1.1)矛盾。$\mathbf{y}'$ 的具体构造方法如下。令

$$y'_\ell = \begin{cases} 0, & \text{若 } \ell = j \text{ 或} \ell \notin \mathrm{supp}(\mathbf{y})\text{；} \\ \frac{1}{1-y_j} \cdot y_\ell\,, & \text{若 } \ell \in \mathrm{supp}(\mathbf{y}) \setminus \{j\}。 \end{cases}$$

（由假设 $u^{\mathsf{L}}(\mathbf{x}, \mathbf{y}) > u^{\mathsf{L}}(\mathbf{x}, j)$ 可知 $y_j \neq 1$。）显然，因为 $\mathbf{y}'$ 的支持集是 $\mathrm{supp}(\mathbf{y})$ 的子集，我们有 $\mathbf{y}' \in \Delta\mathsf{BR}(\mathbf{x})$；同时，

$$\sum_{\ell=1}^{n} y'_\ell = \frac{1}{1-y_j} \cdot \sum_{\ell \in \mathrm{supp}(\mathbf{y}) \setminus \{j\}} y_\ell = 1$$

(注意对所有 $\ell \notin \mathrm{supp}(\mathbf{y})$ 我们有 $y'_\ell = y_\ell = 0$)。从而保证按此构造的 $\mathbf{y}'$ 是一个概率分

[4] 但跟随者的收益不一定总相等。例如，当领导者的收益矩阵的所有元素都相等时，任何 $\mathbf{x} \in \Delta^{m-1}$ 都是领导者的最优策略，进而对于任意的 $j \in \mathsf{BR}(\mathbf{x})$，策略组合 $(\mathbf{x}, j)$ 都是强 Stackelberg 均衡。然而，对于不同的领导者策略 $\mathbf{x}$ 和 $\mathbf{x}'$，跟随者的最优收益 $\max_{\mathbf{y} \in \Delta\mathsf{BR}(\mathbf{x})} u^{\mathsf{F}}(\mathbf{x}, \mathbf{y})$ 和 $\max_{\mathbf{y} \in \Delta\mathsf{BR}(\mathbf{x}')} u^{\mathsf{F}}(\mathbf{x}', \mathbf{y})$ 不一定相等。

布。更重要的是，我们可以得到以下同(1.1)矛盾的不等式

$$\begin{aligned} u^{\mathsf{L}}(\mathbf{x},\mathbf{y}') &= \frac{1}{1-y_j}\cdot\big(u^{\mathsf{L}}(\mathbf{x},\mathbf{y}) - y_j\cdot u^{\mathsf{L}}(\mathbf{x},j)\big) \\ &= u^{\mathsf{L}}(\mathbf{x},\mathbf{y}) + \frac{y_j}{1-y_j}\cdot\big(u^{\mathsf{L}}(\mathbf{x},\mathbf{y}) - u^{\mathsf{L}}(\mathbf{x},j)\big) > u^{\mathsf{L}}(\mathbf{x},\mathbf{y})。\end{aligned}$$

综合(1.2)和(1.3)可得，对所有 $j\in\mathrm{supp}(\mathbf{y})$ 均有

$$u^{\mathsf{L}}(\mathbf{x},j) = u^{\mathsf{L}}(\mathbf{x},\mathbf{y}) = \max_{\mathbf{x}'\in\Delta^{m-1}}\max_{\mathbf{y}'\in\Delta\mathsf{BR}(\mathbf{x}')} u^{\mathsf{L}}(\mathbf{x}',\mathbf{y}')。 \qquad \square$$

我们给出以下简化的均衡定义。

定义 1.2.2 (简化均衡定义)**.** 假设 $\mathbf{x}\in\Delta^{m-1}$、$j\in\mathsf{BR}(\mathbf{x})$。当以下条件满足时，策略组合 $(\mathbf{x},j)$ 被称作是一个强 Stackelberg 均衡：

$$u^{\mathsf{L}}(\mathbf{x},j) = \max_{\mathbf{x}'\in\Delta^{m-1}}\max_{j'\in\mathsf{BR}(\mathbf{x}')} u^{\mathsf{L}}(\mathbf{x}',j')。$$

当以下条件满足时，策略组合 $(\mathbf{x},j)$ 被称作是一个弱 Stackelberg 均衡：

$$u^{\mathsf{L}}(\mathbf{x},j) = \max_{\mathbf{x}'\in\Delta^{m-1}}\min_{j'\in\mathsf{BR}(\mathbf{x}')} u^{\mathsf{L}}(\mathbf{x}',j')。$$

此外，以上强弱两种均衡定义中，强 Stackelberg 均衡在文献和相关研究中更为广泛采用。在许多场景下，特别是安全领域的场景下，这似乎是一个违背常理的假设。然而，基于两大原因，强 Stackelberg 均衡显得更为合理。其一，对于一个强 Stackelberg 均衡 $(\mathbf{x},j)$，通常情况下，我们可以对 $\mathbf{x}$ 施加一个无穷小的扰动，使得扰动后 $\mathsf{BR}(\mathbf{x})$ 仅包含 j 一个元素。例如，在安全博弈场景中，当多个目标都是攻击者的最优选择时，领导者只需将其对某个对象的保护概率稍微降低（任意小的一个量），该对象就会成为攻击者的唯一最优攻击对象。其二，强 Stackelberg 均衡总是存在（我们接下来将会讲解如何找到一个强 Stackelberg 均衡），而弱 Stackelberg 均衡则不一定存在（见本节末尾实例）。

1.2.2 均衡求解

强 Stackelberg 均衡的求解可被转化为线性规划问题。由于计算过程中需要求解多次线性规划问题，该方法被称作多线性规划（multiple-linear program）方法，其最早由 Conitzer 等人提出 [10]。根据定义1.2.2，我们可将强 Stackelberg 均衡的搜索范围限制

算法 1.1 求解强 Stackelberg 均衡。

初始化：$\hat{u} \leftarrow -\infty$，$\mathbf{z} \leftarrow null$，$\ell \leftarrow 0$；

for $j = 1, \ldots, n$ **do**

  求解线性规划(1.4)，令 $\mathbf{x}^*$ 为其最优解；

  if $u^{\mathsf{L}}(\mathbf{x}^*, j) > \hat{u}$ **then**

    $\hat{u} \leftarrow u^{\mathsf{L}}(\mathbf{x}^*, j)$，$\mathbf{z} \leftarrow \mathbf{x}^*$，$\ell \leftarrow j$；

输出 $(\mathbf{z}, \ell)$。

在形如 $(\mathbf{x}, j)$ 的策略组合上，其中 $j \in \{1, \ldots, n\}$。因此我们可以枚举跟随者纯策略 j，对于每个 j 我们使用以下线性规划去寻找满足强 Stackelberg 均衡的策略组合。

$$\max_{\mathbf{x}} \quad u^{\mathsf{L}}(\mathbf{x}, j) \tag{1.4}$$

$$\text{s.t.} \quad u^{\mathsf{F}}(\mathbf{x}, j) \geq u^{\mathsf{F}}(\mathbf{x}, \ell) \qquad \forall \ell = 1, \ldots, n \tag{1.4-a}$$

$$\mathbf{x} \in \Delta^{m-1} \tag{1.4-b}$$

换言之，我们将领导者的策略空间 Δ^{m-1} 划分为 n 个区域，每个区域对应一个跟随者反馈 $j \in \{1, \ldots, n\}$ 并包含所有使得 $j \in \mathsf{BR}(\mathbf{x})$ 的领导者策略 $\mathbf{x}$。在这些策略下，j 是跟随者的最优反馈之一。上述线性规划的最优解进而对应了单个区域内领导者的最优策略，其中的第一行约束条件等价于 $j \in \mathsf{BR}(\mathbf{x})$。求解对应于所有 $j = 1, \ldots, n$ 的上述线性规划问题，得到的解中目标函数值最大者 $\mathbf{x}^*$ 即领导者在整个策略空间 Δ^{m-1} 上的最优策略。相应的 j 与 $\mathbf{x}^*$ 组成的策略组合 $(\mathbf{x}^*, j)$ 即强 Stacklberg 均衡之一。算法1.1总结了以上步骤。我们知道，线性规划的最优解可在多项式时间内求出 [5]，在实际应用中也有诸多现成的求解器（如 CPLEX、Gurobi 等）可高效求解较大规模的问题。

强 Stackelberg 均衡的性质

至此，我们可将 Stackelberg 均衡同纳什均衡做一个对比。与 Stackelberg 均衡不同，纳什均衡描述了博弈双方同时出牌场景下的稳态。对比纳什均衡，我们可以看到 Stackelberg 均衡在 Stackelberg 博弈场景下的几点优越性。其一，就双矩阵形式给出的博弈模型而言，强 Stackelberg 均衡的求解可在多项式时间内完成，而纳什均衡的求解已知是 PPAD-难的问题 [12]，对于这类问题尚不知晓是否存在多项式时间的算法对其求解。其二，所有的强 Stackelberg 均衡对应相同的领导者收益，而纳什均衡不具备该性质，因而当多个均衡存在时后者面临一个均衡选择问题。其三，强 Stackelberg 均衡中领导者的收益总是不小于其在任一纳什均衡中的收益，我们将该性质的证明留给

读者。

1.2.3 Stackelberg 安全博弈模型及求解

安全博弈模型构建于上述 Skackelberg 博弈模型之上。明确了 Skackelberg 博弈模型及相关基本概念后，我们下面介绍安全博弈模型及其求解方法。

在安全博弈场景中，领导者调度手中的 k 个安保资源保护 n 目标（通常假设 $k \ll n$），而跟随者意图攻击其中的一个目标。因而领导者和跟随者往往也被称作*防御者*（defender）和*攻击者*（attacker）。令 $T = \{t_1, \ldots, t_m\}$ 表示目标集合。在最基本的模型中，我们假设所有安保资源都是完全同质的，所有资源均可被分配给 T 中的任意一个目标。而目标是异质的，每个目标 $t \in T$ 对应一组奖励值 $r^{\mathsf{L}}(t)$ 和 $r^{\mathsf{F}}(t)$，及一组惩罚值 $p^{\mathsf{L}}(t)$ 和 $p^{\mathsf{F}}(t)$。奖励值总是大于相应的惩罚值：$r^{\mathsf{L}}(t) > p^{\mathsf{L}}(t)$ 且 $r^{\mathsf{F}}(t) > p^{\mathsf{F}}(t)$。如果攻击者选择攻击某目标 $t \in T$，同时防御者分配了至少一个资源保护目标 t，那么攻击不会成功，其结果是防御者获得奖励值 $r^{\mathsf{L}}(t)$，而攻击者获得惩罚值 $p^{\mathsf{F}}(t)$；反之，如果目标 t 上未分配任何资源，则攻击成功，攻击者获得奖励值 $r^{\mathsf{F}}(t)$ 而防御者获得惩罚值 $p^{\mathsf{L}}(t)$。

以上便是纯策略组合下博弈双方收益的定义。我们称一个目标处于*被保护*状态，当且仅当一个或多个资源被分配到该目标；反之，我们称该目标处于*未被保护*状态。我们可将防御者的一个纯策略表示为一个 n 维 0/1 向量 $\mathbf{s} = (s_t)_{t \in T} \in \{0,1\}^n$。其中，$s_t = 1$ 表示目标 t 处于被保护状态，而 $s_t = 0$ 表示目标未被保护。给定 k 个资源，防御者可行的纯策略集合为 $\mathcal{T}_k := \left\{\mathbf{s} \in \{0,1\}^n : \sum_{t \in T} s_t \leq k\right\}$。换言之，防御者最多将 k 个目标置于被保护状态下。

在混合策略下，防御者以一定的概率选择执行 $\mathcal{T}_k$ 中的纯策略。令 $\mathbf{x} = (x_{\mathbf{s}})_{\mathbf{s} \in \mathcal{T}_k} \in \Delta(\mathcal{T}_k)$ 表示一个混合策略，其中 $x_{\mathbf{s}}$ 表示防御者选择纯策略 $\mathbf{s}$ 的概率。[5] 从攻击者的角度来说，攻击者关心的是混合策略 $\mathbf{x}$ 下每个目标被保护的概率，简称*保护率*。我们引入一个*保护率向量* $\mathbf{c} = (c_t)_{t \in T} \in [0,1]^n$，其中每个元素 c_t 表示目标 t 的保护率。方便起见，有时我们也将保护率向量写作 $\mathbf{c} = (c_1, \ldots, c_m)$，其中 c_i 对应 T 中的第 i 个目标 t_i 的保护率。给定混合策略 $\mathbf{x}$，对任一目标 $t \in T$ 我们有 $c_t = \sum_{\mathbf{s} \in \mathcal{T}_k} x_{\mathbf{s}} \cdot s_t$。通过该保护率向量 $\mathbf{c}$ 和攻击者的目标选择 t，我们即可给出博弈双方在策略组合 $(\mathbf{x}, t)$ 下的期望收益，分别为：

$$
\begin{aligned}
u^{\mathsf{L}}(\mathbf{c}, t) &:= c_t \cdot r^{\mathsf{L}}(t) + (1 - c_t) \cdot p^{\mathsf{L}}(t), \\
\text{与} \quad u^{\mathsf{F}}(\mathbf{c}, t) &:= c_t \cdot p^{\mathsf{F}}(t) + (1 - c_t) \cdot r^{\mathsf{F}}(t)。
\end{aligned}
$$

[5] 给定任意有限集合 S，我们用 $\Delta(S) := \left\{\mathbf{x} \in \mathbb{R}^{|S|} : \sum_{i \in S} x_i = 1, \text{且} x_i \geq 0 \, \forall i \in S\right\}$ 表示 S 上的概率分布的集合。

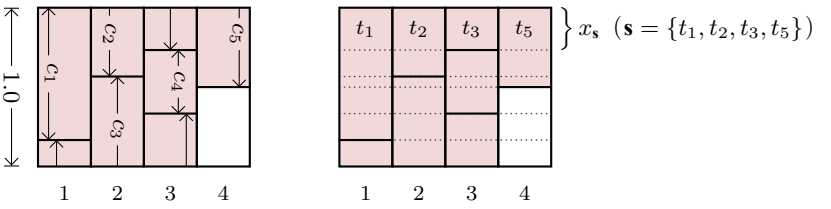

图 1.1 梳形采样算法：示例中领导者有四个资源（$k = 4$），有五个目标需要保护（$n = 5$）。如左图所示，算法首先创建 k 个高为 1 的矩形带，并将它们并排放置。每个矩形带对应于一个资源。给定保护率向量 $\mathbf{c}$，已知 $\mathbf{c}$ 满足 $\sum_{i=1}^{n} c_i \le k$，我们可将保护率 $c_1, \dots, c_n$ 依次填入算法创建的矩形带内。填完之后相邻保护率的连接点对应的水平线将所有矩形带分割成至多 $n+1$ 行，如右图所示。我们将每一行看作一个纯策略 $\mathbf{s}$，令 $s_i = 1$ 当且仅当该行包含 c_i 的一部分。同时我们取该行的高度作为所要构造的混合策略中 $\mathbf{s}$ 的概率 $x_{\mathbf{s}}$。如此一来，我们得到至多 $n+1$ 个纯策略 $\mathbf{s}_1, \dots, \mathbf{s}_{n+1}$，以及这些纯策略上的概率分布 $\mathbf{x} = (x_{\mathbf{s}_1}, \dots, x_{\mathbf{s}_{n+1}})$。显然，每个 $\mathbf{s}_j$ 至多覆盖 k 个目标，且 $\sum_{j=1}^{n+1} x_{\mathbf{s}_j} = 1$。

因此，我们通常直接用保护率向量来表示防御者的混合策略。其优点在于保护率向量 $\mathbf{c}$ 比混合策略的直接表达 $\mathbf{x}$ 更为紧凑；前者仅为 n 维，而后者需要枚举 $\mathcal{T}_k$ 内的所有纯策略，其数量级为 $O(n^k)$。更进一步，我们可以定义以下可行保护率向量集合

$$\mathcal{C}_k = \left\{ \mathbf{c} \in \mathbb{R}^n : 0 \le c_t \le 1 \ \forall t \in T, \ \text{且} \sum\nolimits_{t \in T} c_t \le k \right\} 。 \tag{1.5}$$

不难发现，任意一个混合策略 $\mathbf{x} \in \Delta(\mathcal{T}_k)$ 对应的保护率向量 $\mathbf{c} = \sum_{\mathbf{s} \in \mathcal{T}_k} x_{\mathbf{s}} \cdot \mathbf{s}$ 总是在上述集合 $\mathcal{C}_k$ 中。同时可以证明，对于 $\mathcal{C}_k$ 中的任意向量 $\mathbf{c}$ 都被某个混合策略 $\mathbf{x} \in \Delta(\mathcal{T}_k)$ 实现，使得 $c_t = \sum_{\mathbf{s} \in \mathcal{T}_k} x_{\mathbf{s}} \cdot s_t$ 对于所有目标 $t \in T$ 成立。从 $\mathbf{c}$ 到 $\mathbf{x}$ 的转化可以通过图1.1中展示的*梳形采样算法* (comb-sampling algorithm) 在多项式时间内求得 [58]。

同前述 Stackelberg 博弈模型相同，在安全博弈场景下防御方先行动，选取并执行一个混合策略 $\mathbf{c}$。攻击者观察到 $\mathbf{c}$ 后选择一目标进行攻击。根据定义1.2.1，满足以下条件的策略组合 $(\mathbf{c}, t)$ 构成强 Stackelberg 均衡：

$$(\mathbf{c}, t) \in \arg \max_{\mathbf{z} \in \mathcal{C}, \, \ell \in \mathsf{BR}(\mathbf{z})} u^{\mathsf{L}}(\mathbf{z}, \ell) \quad 。$$

其中对于任一混合策略 $\mathbf{x}$，我们沿用最优反馈集 $\mathsf{BR}(\mathbf{x})$ 的定义。而对于任一保护率向量 $\mathbf{c}$，我们令 $\mathsf{BR}(\mathbf{c})$ 等同于与 $\mathbf{c}$ 对应的（任意）混合策略 $\mathbf{x}$ 的最优反馈集 $\mathsf{BR}(\mathbf{x})$。

前述求解强 Stackelberg 均衡的多线性规划方法同样适用于安全博弈场景。不同之处在于此处我们将保护率向量 $\mathbf{c}$ 作为线性规划的优化对象以缩减问题的规模。对每个

目标 $t \in T$，我们构造以下线性规划。

$$\max_{\mathbf{c}} \quad u^{\mathsf{L}}(\mathbf{c}, t) \tag{1.6}$$

$$\text{s.t.} \quad u^{\mathsf{F}}(\mathbf{c}, t) \geq u^{\mathsf{F}}(\mathbf{c}, \ell) \qquad \forall\, \ell \in T \tag{1.6-a}$$

$$\mathbf{c} \in \mathcal{C}_k \tag{1.6-b}$$

其中第一行的约束条件将策略 $\mathbf{c}$ 限制为使得 $t \in \mathsf{BR}(\mathbf{c})$ 的策略。求解所有 n 个线性规划，得到的解中目标函数值最大者 $\mathbf{c}^*$ 即防御者在整个策略空间 $\mathcal{C}_k$ 上的最优策略。相应的目标 t 与 $\mathbf{c}^*$ 组成的策略组合即强 Stackelberg 均衡之一。

1.2.4 安全博弈实例

在本小节的最后我们回顾表1.1中给出的实例。下表给出了该实例中的奖惩参数值。

表 1.2 表1.1中实例的奖惩参数值。

	目标 t_1	目标 t_2
奖励值（r^{L}）	5	2
惩罚值（p^{L}）	-5	-1

防御者

	目标 t_1	目标 t_2
奖励值（r^{F}）	5	1
惩罚值（p^{F}）	-3	-1

攻击者

该实例也可被表示为双矩阵形式的 Stackelberg 博弈，相应的收益矩阵 u^{L} 和 u^{F} 如下。矩阵的行对应防御者的三个纯策略：保护 t_1、保护 t_2，以及不保护任何目标。[6] 矩阵的列对应攻击者的两个纯策略：攻击 t_1 和攻击 t_2。

$$u^{\mathsf{L}} = \begin{pmatrix} 5 & -1 \\ -5 & 2 \\ -5 & -1 \end{pmatrix} \qquad u^{\mathsf{F}} = \begin{pmatrix} -3 & 1 \\ 5 & -1 \\ 5 & 1 \end{pmatrix}$$

我们可以利用算法1.1求解该博弈的强 Stackelberg 均衡。在该博弈的强 Stackelberg 均衡为 $(\mathbf{c}^*, t_2)$，其中防御者施加的保护率向量为 $\mathbf{c}^* = \left(\frac{3}{5}, \frac{2}{5}\right)$，对应于双矩阵博弈下的混合策略 $\mathbf{x}^* = \left(\frac{3}{5}, \frac{2}{5}, 0\right)$。当防御者使用该策略时，对于攻击者来说攻击目标各个目标

[6] 在该实例下不保护任何目标的纯策略显得没有意义，但在某些情况下最优策略的实现必须借助这样的纯策略。

的收益分别为：

$$u^{\mathsf{F}}(\mathbf{c}^*, t_1) = 5 \times \left(1 - \frac{3}{5}\right) + (-3) \times \frac{3}{5} = \frac{1}{5}$$
$$\text{及} \quad u^{\mathsf{F}}(\mathbf{c}^*, t_2) = 1 \times \left(1 - \frac{2}{5}\right) + (-1) \times \frac{2}{5} = \frac{1}{5}。$$

二者相等，故 $\mathsf{BR}(\mathbf{c}^*) = \{t_1, t_2\}$。在强 Stackelberg 均衡的假设下，攻击者选择攻击有利于防御者的目标 t_1，领导者获得收益 $u^{\mathsf{L}}(\mathbf{c}^*, t_1) = 1$。领导者可以通过降低对 t_1 的保护率一个微小的量引导攻击者的这种选择。

该实例不存在弱 Stackelberg 均衡，可以通过反证法见得。假设存在一个弱 Stackelberg 均衡 $(\mathbf{c}^*, t)$。考虑以下三种情形。

- 情形一：$\mathsf{BR}(\mathbf{c}^*) = \{t_1\}$。根据定义我们有 $u^{\mathsf{F}}(\mathbf{c}^*, t_1) > u^{\mathsf{F}}(\mathbf{c}^*, t_2)$，展开并结合资源约束条件 $c_1^* + c_2^* \leq 1$ 可得 $c_1^* < 3/5$。考虑另一领导者策略 $\mathbf{c}' = (c_1', c_2')$，其中

  $$c_1' = \frac{1}{2} \cdot \left(c_1^* + \frac{3}{5}\right), \quad c_2' = \frac{2}{5}。$$

  因此，$c_1' < 3/5 \Longrightarrow u^{\mathsf{F}}(\mathbf{c}', t_1) > u^{\mathsf{F}}(\mathbf{c}', t_2) \Longrightarrow \mathsf{BR}(\mathbf{c}') = \{t_1\}$。同时，$c_1' > c_1^*$，从而根据防御者收益的单调性可得 $u^{\mathsf{L}}(\mathbf{c}', t_1) > u^{\mathsf{L}}(\mathbf{c}^*, t_1)$。我们有

  $$\min_{t \in \mathsf{BR}(\mathbf{c}')} u^{\mathsf{L}}(\mathbf{c}', t) = u^{\mathsf{L}}(\mathbf{c}', t_1) > u^{\mathsf{L}}(\mathbf{c}^*, t_1) = \min_{t \in \mathsf{BR}(\mathbf{c}^*)} u^{\mathsf{L}}(\mathbf{c}^*, t)。$$

  根据定义1.2.1，$(\mathbf{c}^*, t)$ 不可能是弱 Stackelberg 均衡。

- 情形二：$\mathsf{BR}(\mathbf{c}^*) = \{t_2\}$。同理可证明，在这种情况下 $(\mathbf{c}^*, t)$ 依然不可能是弱 Stackelberg 均衡。

- 情形三：$\mathsf{BR}(\mathbf{c}^*) = \{t_1, t_2\}$。我们有 $u^{\mathsf{F}}(\mathbf{c}^*, t_1) = u^{\mathsf{F}}(\mathbf{c}^*, t_2)$，展开可得

  $$c_2^* = 4 \cdot c_1^* - 2。$$

  结合资源约束条件 $c_1^* + c_2^* \leq 1$ 可得 $c_2^* \leq 2/5$。同时将以上不等式带入领导者的收益函数可得不等式 $u^{\mathsf{L}}(\mathbf{c}^*, t_1) > u^{\mathsf{L}}(\mathbf{c}^*, t_2)$ 对于所有 $c_2^* \leq 2/5$ 恒成立。因此

  $$\min_{t \in \mathsf{BR}(\mathbf{c}^*)} u^{\mathsf{L}}(\mathbf{c}^*, t) = u^{\mathsf{L}}(\mathbf{c}^*, t_2) \leq u^{\mathsf{L}}\left(\frac{2}{5}, t_2\right) = \frac{1}{5}。$$

  此时，考虑另一领导者策略 $\mathbf{c}' = (c_1', c_2')$，其中 $c_1' = 3/5 - 1/100$，$c_2' = 2/5$。不

难验证，$\mathsf{BR}(\mathbf{c}') = \{t_1\}$，且 $u^{\mathsf{L}}(\mathbf{c}', t_1) > 1/5$。因此

$$\min_{t \in \mathsf{BR}(\mathbf{c}')} u^{\mathsf{L}}(\mathbf{c}', t) = u^{\mathsf{L}}(\mathbf{c}', t_1) > u^{\mathsf{L}}(\mathbf{c}^*, t_2) = \min_{t \in \mathsf{BR}(\mathbf{c}^*)} u^{\mathsf{L}}(\mathbf{c}^*, t)。$$

根据定义1.2.1，$(\mathbf{c}^*, t)$ 不可能是弱 Stackelberg 均衡。

综上可得该实例不存在弱 Stackelberg 均衡。

1.3 复杂环境下的安全博弈

前面介绍的模型为建模和求解一个基本的安全博弈场景提供了的思路和方法，为更为复杂的场景下的安全博弈问题提供了模型基础。由于现实世界的诸多复杂性，安全博弈论的实际应用要求研究者在模型中还需要充分考虑到其他因素并解决相应的挑战。本节介绍安全博弈论在信息不完全与不确定性、复杂策略空间两个方面的面临的问题及解决方法。

1.3.1 信息不完全与不确定性

同博弈论领域的其他问题类似，安全博弈论的应用面临的一大挑战是信息不完全的问题。我们前面讲到的模型和算法都建立在完全信息的基础，作为博弈领导者，我们明确地知道博弈中双方的收益矩阵。不幸的是，现实世界的场景往往是信息不完全的。博弈的领导者常常不知道跟随者确切的收益矩阵。在这种情况下，我们无法通过前面介绍的方法计算博弈的均衡。针对该问题，贝叶斯安全博弈模型（Bayesian Stackelberg security game）被提出并用于建模信息不完全场景下的安全博弈。我们在本小节介绍该模型及其求解的问题。

贝叶斯 Stackelberg 博弈

在贝叶斯 Stackelberg 博弈模型中，领导者面临一系列可能的跟随者类型，每种类型对应于一个收益矩阵。令 Θ 为所有可能类型的集合（假设 Θ 为有穷集），$u_\theta^{\mathsf{F}} \in \mathbb{R}^{m \times n}$ 为每个类型 $\theta \in \Theta$ 跟随者的收益矩阵（不失一般性我们假设所有跟随者的纯策略数量相等）。领导者不能确定跟随者的具体类型，但知道每种类型出现的概率 μ_θ。该模型也适用于领导者面对大量跟随者的情形，每种类型 θ 的跟随者占据总人数的比例为 μ_θ。自然地，领导者关心的是他在分布 μ 上的期望收益。据此，我们可定义贝叶斯 Stackelberg 均衡如下，其中我们扩张最有反馈集的定义到每个跟随者类型 $\theta \in \Theta$，令 $\mathsf{BR}_\theta(\mathbf{x}) = \arg\max_{j \in [n]} u_\theta^{\mathsf{F}}(\mathbf{x}, j)$ 为跟随者类型 θ 的最优反馈集。

定义 1.3.1 (贝叶斯 Stackelberg 均衡)**.** 给定领导者混合策略 $\mathbf{x} \in \Delta^{m-1}$，以及每个跟随者 $\theta \in \Theta$ 的纯策略 $j_\theta \in \mathsf{BR}_\theta(\mathbf{x})$。$\mathbf{x}$ 和所有 j_θ 构成一个（强）贝叶斯 Stackelberg 均衡当且仅当以下条件成立：

$$\mathbb{E}_{\theta\sim\mu} u^{\mathsf{L}}(\mathbf{x}, j_\theta) = \max_{\mathbf{y}\in\Delta^{m-1}} \mathbb{E}_{\theta\sim\mu} \max_{\ell\in\mathsf{BR}_\theta(\mathbf{y})} u^{\mathsf{L}}(\mathbf{y}, \ell)。$$

使用 Harsanyi 变形法 [25]，贝叶斯 Stackelberg 博弈可以被转化为一个双矩阵 Stakcelberg 博弈加以求解。然而这样做的代价指数大规模的收益矩阵。事实上，求解贝叶斯 Stackelberg 均衡已被证明是 NP-难的问题 [10]，因此在 P≠NP 的假设下不存在高效的算法解决这个问题。除使用 Harsanyi 变形法外，我们也可以将问题的求解表示为整数线性规划问题。整数线性规划是经典的优化问题之一，尽管其求解也是 NP-难的问题，因其应用的广泛性已有许多启发式算法以及成熟的求解器可对其求解（如前面提到的 CPLEX、Gurobi 等）。我们采用以下整数线性规划求解贝叶斯 Stackelberg 均衡。

$$\begin{aligned}
\max_{\mathbf{x},(y_{\theta,j})_{\theta,j},(v_\theta)_\theta} \quad & \sum_{\theta\in\Theta} \mu_\theta \cdot v_\theta \\
\text{s.t.} \quad & v_\theta \le u^{\mathsf{L}}_\theta(\mathbf{x}, j) + M\cdot(1-y_{\theta,j}) && \forall\,\theta\in\Theta,\ j=1,\dots,n \\
& u^{\mathsf{F}}_\theta(\mathbf{x}, j) + M\cdot(1-y_{\theta,j}) \ge u^{\mathsf{F}}_\theta(\mathbf{x}, j') && \forall\,\theta\in\Theta,\ j,j'=1,\dots,n \\
& \sum_{j=1}^{n} y_{\theta,j} = 1 && \forall\,\theta\in\Theta \\
& y_{\theta,j} \in \{0,1\} && \forall\,\theta\in\Theta,\ j=1,\dots,n \\
& \mathbf{x} \in \Delta^{m-1}
\end{aligned}$$

具体地说，上述规划包含一个连续变量 $\mathbf{x} \in \Delta^{m-1}$ 作为领导者的混合策略，此外，对于每一对 θ、j 我们设置一个整数变量 $y_{\theta,j} \in \{0,1\}$ 表示第 j 个纯策略是否为跟随者类型 θ 选定的最优反馈：$y_{\theta,j}=1$ 表示肯定，$y_{\theta,j}=0$ 表示否定。第三行的约束条件 $\sum_{j=1}^{n} y_{\theta,j}=1$ 保证了有且仅有一个 j 为跟随者选定的反馈策略。此外，对于每个跟随者类型我们还设置一个连续变量 v_θ 表示领导者在这个跟随者类型上取得的期望收益。规划中 M 是一个足够大的常数（任意大于博弈双方收益矩阵中所有元素的常数）。

因此，当 $y_{\theta,j}=1$ 时，j 是类型 θ 的最优反馈，上述第一个约束等价于 $v_\theta \le u^{\mathsf{L}}_\theta(\mathbf{x}, j)$，迫使 v_θ 不超过 $u^{\mathsf{L}}_\theta(\mathbf{x}, j)$；事实上，我们最终会得到 $v_\theta = u^{\mathsf{L}}_\theta(\mathbf{x}, j)$，因为上述规划的最大化目标总是会选取满足条件的最大的 v_θ 值。而当 $y_{\theta,j}=0$ 时，因为 M 足够大，该约束条件自然成立，因而这些不是最优反馈的 j 对领导者的收益不产生任何影响。类

似地，第二行的约束条件保证了 j 是跟随者的最优反馈。

具体到安全博弈场景，我们可以很容易地将上述整数线性规划改写成建立在保护率向量上的规划（类似于1.6）。已经有许多相关的工作对贝叶斯安全博弈进行了研究。在这方面的开创性工作属于 Paruchuri 等人首次提出的贝叶斯安全博弈模型 [39, 41]。这些工作为贝叶斯安全博弈模型在 ARMOR 系统的开发及其在洛杉矶国际机场的布署奠定了重要基础 [28, 44, 45]。基于贝叶斯安全博弈模型的其他应用在社交网络信息污染控制、对抗环境下的路径规划、海上巡逻以及关键基础设施的保护伞发挥了重要作用 [3, 20, 26, 38, 56]。

鲁棒 Stackelberg 均衡

除了收益函数上的不确定性，跟随者的行为常常也存在一定的不确定性。例如，Stackelberg 博弈模型假设跟随者能观察到领导者使用的混合策略。在现实情况下，跟随者无法直接观察到混合策略 $\mathbf{x}$，而只能从观察到的一系列纯策略中得到对 $\mathbf{x}$ 的一个估计 $\tilde{\mathbf{x}}$，两者之间总是存在一定的实际误差。这即可能是统计学上的误差，也可能是实际观察中的偏差。由于跟随者对 $\mathbf{x}$ 和 $\tilde{\mathbf{x}}$ 的反馈可能截然不同，忽视这种误差将会导致严重的后果。例如，在表1.2展示的实例中，我们已经知道领导者在强 Stackelberg 均衡假设下的最优策略为分别以 $3/5$ 和 $2/5$ 的概率保护两个目标。如果攻击者的观察到 100 天中领导者有 43 天保护了目标 t_1，57 天保护了 t_2，攻击者可能会认为领导者在以一个略大于 $3/5$ 的概率保护 t_1，从而严格偏向于攻击 t_2。这将不同于强 Stackelberg 均衡假设的跟随者反馈，使得领导者的实际收益远低于其在强 Stackelberg 均衡中的收益。除了跟随者对领导者策略观测上的误差外，跟随者本身也可能因为收益评估上的精度和误差表现出反馈上的偏差。

针对以上情况，一个自然的改进方法是在计算最优策略是设置一定的冗余。就线性规划求解法(1.4)而言，我们可在约束条件(1.4)-a 中设置一个 ϵ 的冗余，显式地要求领导者策略 $\mathbf{x}$ 引导的最优跟随者反馈明显优于其他反馈：

$$u^{\mathsf{F}}(\mathbf{x}, j) \geq u^{\mathsf{F}}(\mathbf{x}, \ell) + \epsilon \quad 。$$

更为细致的一种做法是定义一个近似最优反馈集

$$\epsilon\text{-}\mathsf{BR}(\mathbf{x}) := \left\{ j : \ u^{\mathsf{F}}(\mathbf{x}, j) > \max_{\ell \in \{1,\ldots,n\}} u^{\mathsf{F}}(\mathbf{x}, \ell) - \epsilon \right\} 。$$

集合中包含了近似最优的跟随者反馈。在此基础上我们可以定义鲁棒 Stackelberg 均衡（定义1.3.2）。其基本假设是当领导者使用策略 $\mathbf{x}$ 时，跟随者会选择 ϵ-$\mathsf{BR}(\mathbf{x})$ 中对领导

者最差的策略。

定义 1.3.2 (鲁棒 Stackelberg 均衡)**.** 当以下条件满足时，策略组合 $(\mathbf{x}, j) \in \Delta^{m-1} \times \{1, \ldots, n\}$ 构成一个ϵ-鲁棒 Stackelberg 均衡：

$$u^{\mathsf{L}}(\mathbf{x}, j) = \max_{\mathbf{x}' \in \Delta^{m-1}} \max_{j' \in \epsilon\text{-}\mathsf{BR}(\mathbf{x})} u^{\mathsf{L}}(\mathbf{x}', j')。$$

鲁棒 Stackelberg 均衡的定义保证了领导者在使用 $\mathbf{x}$ 时不会获得比均衡中更差的收益，即便跟随者对收益的认知存在 ϵ 大的偏差。该定义与上述直接在约束条件(1.4)-a 中设置 ϵ 冗余的方法的不同之处在于后者过于保守地排除掉了一切可能导致跟随者不同反馈的收益太过相近的领导者策略。例如，当跟随者的收益为常数时（收益矩阵所有元素都相同），后者会导致最优策略问题无解。鲁棒 Stackelberg 均衡的求解可以被表示成一个整数线性规划问题。Pita 等人的工作对该问题进行了详尽的研究 [46]，感兴趣的读者可以自行阅读。

非完全理性跟随者模型

另一种建模跟随者行为不确定性的经典模型是*质反应模型*，也称作 QR（quantal response）模型。该模型最早源于经济学领域对有限理性博弈参与者的研究工作 [36]。模型的基本假设是参与者不会确定性地选择最优的纯策略；任何纯策略无论收益大小都有一定的机率被选中，而机率的大小同该策略的收益成正比。具体地到 Stackelberg 博弈场景而言，若跟随者的行为模型遵从 QR 模型，那么当领导者使用策略 $\mathbf{x}$ 时，跟随者会以概率

$$q_j(\mathbf{x}) := \frac{e^{\lambda \cdot u^{\mathsf{F}}(\mathbf{x}, j)}}{\sum_{j'} e^{\lambda \cdot u^{\mathsf{F}}(\mathbf{x}, j')}}$$

选择使用其第 j 个纯策略作为对 $\mathbf{x}$ 的反馈。其中 e 为自然对数，而 $\lambda \geq 0$ 是一个表示跟随者理性程度的参数：当 $\lambda = 0$ 时，跟随者的行为表现出完全随机性，以均等的概率选择每一个反馈，全然不顾反馈对应的收益；当 λ 趋于正无穷时，跟随者的行为表现趋于完全理性，以趋于 1 的概率选择收益最高的反馈。

一系列安全博弈论应用借助 QR 模型建模现实世界中攻击者的不完全理性行为 [1, 2, 46, 50, 64–66]。这些应用的一大难点在于最优领导者策略的计算。感兴趣的读者可参见 Pita 等人以及 Yang 等人在这方面的研究工作 [46, 64, 66] 了解相关求解算法。

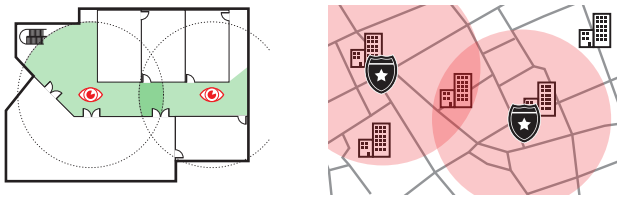

图 1.2　具有保护外部性的安全场景。左图：监控摄像头的可视范围。右图：城市警力的响应范围。

1.3.2　复杂策略空间的处理

上文介绍的基本模型中对于资源分配方案的唯一约束是安保资源数量的上限。在其他复杂场景中，领导者往往面临更多资源分配上的约束条件。例如，在空中警察调度问题中 [57]，警务人员需要被分配到不同的航班上。通常因为航班旅程较短，单个警务人员在一次调度中可以被分配到多个航班执行任务。我们把每个航班看作是一个需要保护的目标，而每个警务人员可被看作多个安保资源，被分配到多个航班上。从这个角度看，问题中有更多的约束条件需要考虑。很显然，单个警务人员虽然能被分配到多个航班，但这些航班在时间上不能有重叠。同时，在空间上，它们必须首尾相连，前一个航班的终点同时也是下一个航班的起点。更为细节的一些问题包括警务人员在航班间的休息时间等。引入这些约束条件，其直接结果是保护率向量的可行空间往往不再能像(1.5)一样被简洁地刻画。类似地，在一些现实场景中资源的保护具有一定的外部性，每个资源能为某一半径范围内的所有目标提供保护（如图1.2）[14, 15]，这同样为刻画保护率向量的可行空间带来困难。如果放弃保护率向量转而以原始的纯策略空间上的概率分布的形式表达混合策略，我们又不得不面对呈指数增长的纯策略数量。采用前面介绍的基于线性规划的方法（算法1.1）求解这些问题意味着我们面对一个指数规模的线性规划（确切地说，变量数 m 呈指数增长）。事实上，对于这些带有额外约束条件的安全博弈模型，均衡的求解往往被证明是 NP-难的问题 [如 14, 15, 29]。随着问题规模的增长，我们需要新的方法去处理均衡求解的问题。最为常用且有效的手段是借助处理大规模线性规划的经典方法*列生成法*（column generation）。

列生成法

将列生成法应用于求解安全博弈问题最早可见于 Jain 等人的工作 [27]。随后更多的工作也借鉴了这套办法 [如 14, 15, 23, 24]。列生成法背后的原理同 Carathéodory 定理 [11] 有一定的联系。根据 Carathéodory 定理，假设 P 是一个 d 维空间中的点集合，那么 P 的凸包（convex hull）中的每一个点都能被表示成 P 中至多 $d+1$ 个点的凸组

算法 1.2 采用列生产法求解(1.4)。

随机选择 $\mathcal{S}$ 的一个较小子集 $\mathcal{S}'$；

repeat

// 主问题

1 求解以下定义在纯策略空间 $\mathcal{S}'$ 上的线性规划(1.4)，其中变量 $\mathbf{x}$ 为一个 $|\mathcal{S}'|$ 维向量：

$$\begin{aligned} \max_{\mathbf{x}} \quad & u^{\mathsf{L}}(\mathbf{x}, j) \\ \text{s.t.} \quad & u^{\mathsf{F}}(\mathbf{x}, j) \geq u^{\mathsf{F}}(\mathbf{x}, \ell) && \forall \ell = 1, \ldots, n \\ & \sum_{s \in \mathcal{S}'} x_s = 1 \\ & x_s \geq 0 && \forall s \in \mathcal{S}' \end{aligned}$$

2 令 $\mathbf{x}^*$ 为上述问题最优解，$\alpha_1, \ldots, \alpha_n$、$\beta$ 为最优解对应的对偶变量值，分别对应于上述第一、二行的约束；

// 从问题

3 求解以下优化问题以获取一个能提高主问题解质量的新纯策略：

$$\min_{s \in \mathcal{S}} \quad \sum_{\ell=1}^{n} \left(u^{\mathsf{F}}(s, \ell) - u^{\mathsf{F}}(s, j)\right) \cdot \alpha_\ell + \beta - u^{\mathsf{L}}(s, j)$$

4 令 s^* 为上述优化问题的最优解，y 为 s^* 对应的目标函数值；

5 **if** $y < 0$ **then**

6 将 s^* 加入 $\mathcal{S}'$；

7 **else**

8 输出 $\mathbf{x}^*$ 为线性规划(1.4)（定义在 $\mathcal{S}$ 上）的最优解，结束算法。

合（convex combination）。

对应到安全博弈模型，假设 $\mathcal{S}$ 是领导者的纯策略集，函数 $\mathsf{cov} : \mathcal{S} \to \mathbb{R}^n$ 将每个纯策略映射到该纯策略提供的保护率向量。那么一个混合策略 $\mathbf{x} \in \Delta(\mathcal{S})$ 提供的保护率向量为 $\mathbf{c} = \sum_{s \in \mathcal{S}} x_s \cdot \mathsf{cov}(s)$。令 $P = \{\mathsf{cov}(s) : s \in \mathcal{S}\}$ 表示所有纯策略集对应的保护率向量的集合。那么 $\mathbf{c}$ 就是 P 的凸包中的一个点，从而根据 Carathéodory 定理，$\mathbf{c}$ 总是能被表示成 $\mathsf{cov}(\mathcal{S})$ 中不多于 $n+1$ 个元素的凸组合。换言之，任何一个可行的保护率向量总是能被不多于 $n+1$ 个纯策略实现，即便纯策略空间 $\mathcal{S}$ 保护的纯策略个数可能远大于 $n+1$。更进一步，当我们面临指数大的纯策略空间 $\mathcal{S}$ 时，实现最优的混合策略并不需要用到 $\mathcal{S}$ 中的所有纯策略。基于此结果，列生成法背后的思路是从 $\mathcal{S}$ 的一个较小的子集出发寻找最优策略。

如算法1.2所示，列生成法在一个主问题（master problem）和从问题（slave problem）之间轮回。主问题是原线性规划问题(1.4)定义在原纯策略空间 $\mathcal{S}$ 的一个子集 $\mathcal{S}'$ 上的问题。从一个较小 $\mathcal{S}'$ 出发使我们免于显式地写出规模较大的原问题。然而，因为 $\mathcal{S}'$ 可能漏掉实现原问题最优解的必要的纯策略，我们得到的这个小规模问题的最优解可能并不是离原问题的最优解。为此，算法的从问题的任务是从 $\mathcal{S}$ 中寻找一个新的纯策略加入 $\mathcal{S}'$ 去提升主问题的解质量。加入新的纯策略后主问题的线性规划会增加一列参数，因此该方法被称作列生成法。

可以证明，当不存在新的纯策略 s 使得算法第3行优化问题的目标函数为负数时，当前主问题的最优解对于原问题而言也已是最优。这从主问题的对偶问题的角度可能更易于理解。我们知道原问题是一个变量数目较多的线性规划问题，其对应的对偶线性规划问题则是一个约束数目较多的问题。从对偶问题的角度看，算法1.2从一个较小的约束集出发求解问题。求得的主问题的最优解因而可能违反了一些该约束集以外的约束条件。为此，从问题需要寻找一个当前未被满足的新约束加入主问题，再重新求解。而算法第3行的目标函数恰恰就是对偶问题中对应于原问题变量 s 的约束 $\sum_{\ell=1}^{n}\left(u^{\mathsf{F}}(s,\ell)-u^{\mathsf{F}}(s,j)\right)\cdot\alpha_\ell+\beta-u^{\mathsf{L}}(s,j)\geq 0$。当不存在未被满足的约束后，主问题和原问题的最优解达成一致。

需要指出的是，算法第3行的从问题在很多应用中依然是 NP-难的问题，但往往可以通过近似算法或启发式算法求解。列生成法并不是理论上的高效算法，其优势在于避免了显式地写出指数规模的原问题，并在诸多实际应用所需的问题规模上取得了较好的实际求解效率。

1.3.3 动态安全博弈

除以上挑战外，现实的安全场景甚至还可能是动态的。一个典型的例子是城市马拉松比赛为代表的大型公共活动的安保问题。2013 年 4 月 15 日发生的波士顿马拉松爆炸案，造成了 3 人死亡，近 200 人受伤，原计划在其后举行的多项赛事、演出被取消。自该事件发生以来，这类活动的安全问题引起了安全部门更高的重视。同静态的安全博弈场景不同，马拉松赛是一个随时间变动的过程。活动的主要参与者 —— 选手与沿途观众 —— 是首要的保护对象，而这些参与者的位置随时间不停变化，由比赛起点向终点逐步推移。这类活动场地复杂，袭击可能造成的结果与时间、地点密切相关，因此高效的安保工作也格外重要，也极具挑战。利用安全博弈理论对活动流程、警方策略以及可能的攻击者行为进行建模，能够帮助警方将有限的警力资源发挥出最大的作用。本节以针对马拉松赛设计的安全博弈模型为例 [67, 68] 介绍安全博弈模型在这类动态问题上的扩展，及相应的策略优化算法。

公共活动安全博弈模型

马拉松赛中，攻击者的目标可被看作是路段或比赛沿途的城市区域。类似地，我们用集合 T 来表示可能被攻击的目标集合。假设防御者共有 m 个安全资源，比赛在时间区间 $[0, t_e]$ 内进行。目标的重要程度呈现动态性。例如，马拉松赛中，起点附近目标的重要程度在比赛开始时较高，而随着比赛的进行，运动员和观众逐渐向终点推移，起点周围目标的重要程度随之降低，而终点附近目标的重要程度逐渐升高。因此，防御者需要在活动进行中动态调度资源，将其在目标间进行转移。我们用一个二元组 $S = \langle Q^0, C\rangle$ 来描述一个防守者策略。其中 $Q^0 = (q_i^0)_i$ 表示在 0 时刻时安全资源的分配情况：q_i^0 即时刻 0 时分配给目标 i 的资源数目。$C = (C_k)_k$，其中 $C_k = (c_{ij}^k)_{i,j\in T}$ 表示第 k 次转移：c_{ij}^k 即在第 k 次转移中从目标 i 转移到目标 j 的资源数。令 τ_k 表示第 k 次转移发生的时间。任意两个目标间转移资源所需的时间 d_{ij} 为给定参数。根据 d_{ij} 即可求得任意时刻 $t \in [0, t_e]$ 任意目标 i 被分配到的资源个数 $q_i^t(S)$：

$$q_i^t(S) = q_i^0 + \sum_{C_k\in C, \tau_k \le t-d_{ji}, j\in T} c_{ji}^k - \sum_{C_k\in C, \tau_k\le t, j\in T} c_{ij}^k。$$

攻击者的纯策略可用二元组 (i, t) 表示，即在时刻 t 攻击目标 i。令 $p(r)$ 表示当有 r 个资源在保护某个目标时攻击者攻击这个目标的成功概率。假设 $p(r) = \frac{1}{e^{\lambda r}}$，其中 $\lambda \ge 1$ 是一个表示每增加一个资源对攻击者的成功率影响的参数。若目标没有被任何资源保护，攻击该目标的成功率为 $p(0) = 1$。若目标被无穷多个资源保护，攻击该目标的成功率为 $p(\infty) = 0$。

每个目标 i 的重要性 $v_i(t)$ 是一个给定的关于时间 t 的连续函数。因而当有 r 个资源保护某个目标 i 时，攻击者攻击该目标获得的收益也是一个连续函数。给定一个防守者策略 S，如果攻击者的策略是 (i, t)，那么攻击者的收益可以被表示为

$$U^a(i, t, S) = p(q_i^t(S)) \cdot v_i(t)。$$

同时我们假设博弈为零和博弈，从而防御者的收益为

$$U^d(i, t, S) = -U^a(i, t, S)。$$

攻击者的最优攻击策略 S 和防御者的最优防御策略 $f(S) = \{f_{tg}(S) : S \to i, f_{tm}(S) : S \to t\}$ 构成一个最大最小均衡，满足以下条件：

$$U^a(f_{tg}(S), f_{tm}(S), S) \ge U^a(i, t, S) \qquad \forall i \in T, t \in [0, t_e]$$

及 $U^d(f_{tg}(S), f_{tm}(S), S) \geq U^d(f_{tg}(S'), f_{tm}(S'), S') \quad \forall S'$。

求解最优动态防御策略

以上动态模型的求解难点在于对防御者连续策略空间的处理。Yin 等人提出的求解思路 [67] 是考虑防御者只能在某些固定的时间点上转移目标的简化模型，并证明该简化模型不失一般性：总是存在一个最大最小均衡，其中防御者的资源转移发生在一个有限的时间点集合内。因此只要求出这个点集即可用求解离散策略空间博弈的方法来求解该连续策略空间的博弈问题。Yin 等人进一步提出了基于混合整数线性规划的算法 SCOUT-D。假设资源转移只能发生在集合 $\phi = \{t_k\}$ 中，那么资源到达某目标的时间只能发生在集合 $\varphi = \{t_\delta : t_\delta = t_k + d_{ij}, \forall t_k \in \phi, \forall i, j \in T\}$ 中。定义点集 $\Psi = \{t_\eta : t_\eta \in \varphi\}$，并令 $H = |\Psi|$。令 $\sigma^e ta = \left(\sigma_{ij}^\eta\right)_{i,j \in T}$，其中 σ_{ij}^η 表示如果一个资源从目标 i 被转移到目标 j，且应在时间 t_η 到达，那么转移开始的时间应该是 $t_{\sigma_{ij}^\eta}$。令 $a_i^{t_\eta}$ 表示在时间 t_η 尚未有任何资源从目标 i 转走时，目标 i 被分配的资源数。令 $b_i^{t_\eta}$ 表示在时间 t_η 所有应从 i 转走的资源都被转走后所余下的资源数目。SCOUT-D 给出下列混合整数线性规划：

$$\begin{aligned}
\min \quad & U && && (1.7)\\
\text{s.t.} \quad & \sum_{i \in T} a_i^0 = m && && (1.7\text{-a})\\
& a_i^{t_{k+1}} = b_i^{t_k} + \sum_{j \in T} c_{ji}^{\sigma_{ji}^{k+1}} && \forall\, k \in \{1, \ldots, H-1\} && (1.7\text{-b})\\
& b_i^{t_k} = a_i^{t_k} - \sum_{j \in T} c_{ji}^k && \forall\, k \in \{1, \ldots, H\} && (1.7\text{-c})\\
& c_{ij}^k \in \{0, 1, \ldots\} && \forall\, k \in \{1, \ldots, H\} && (1.7\text{-d})\\
& \sum_{ij} c_{ij}^{t_\eta} = 0 && \forall\, t_\eta \in \varphi, t_\eta \notin \phi && (1.7\text{-e})\\
& b_i^{t_k} \geq 0 && \forall\, i \in T, k \in \{1, \ldots, H\} && (1.7\text{-f})\\
& U \geq \max_{t \in [t_k, t_{k+1}]} p(q_i^t(S)) \cdot v_i(t) && \forall\, i \in T, k \in \{1, \ldots, H-1\} && (1.7\text{-g})
\end{aligned}$$

其约束条件(1.7-a)限制了初始资源配置的可行性。约束(1.7-b)和(1.7-c)限制了资源转移的可行性（类似网络流问题）。约束(1.7-d)要求转移的资源数为整数。约束(1.7-e)要求资源转移仅发生在允许转移的点集。约束(1.7-f)限制了每个目标留下的资源为非负数。约束(1.7-g)则要求攻击者做出全局最优的反应。

剩下需要考虑的是如何处理连续策略空间的问题。首先，每个目标的价值 $v_i(t)$

可被看做一系列单调曲线段的集合。对每个目标 i 我们可定义一个顺序的时间点集合 $\Xi = \{\xi_\rho^i : \rho \in \{0, \ldots, R_i\}\}$，其中 R_i 为 $v_i(t)$ 的单调段数，使得当 $t \in [\xi_\rho^i, \xi_{\rho+1}^i]$ 时，$v_i(t)$ 是单调的。令 $\Xi = \{\Xi^i : i \in T\}$。假设在一个策略空间连续的博弈中，$S$ 是防守方的最大最小均衡策略。可证明了以下的命题（详细证明见 [67]）。

命题 1.3.1. *假如在 S 中，某资源在 t_1 时刻被从目标 i 转移到目标 j，随后又从目标 j 被转移至目标 k，并在时刻 t_2 到达 k。如果 $t_1 + d_{ij} \in \left[\xi_\rho^j, \xi_{\rho+1}^j\right)$ 且 $t_2 \in \left[t_1 + d_{ij}, \xi_{\rho+1}^j\right)$，那么直接在将该资源从 i 转移到 k 不会降低防御者的最优收益。*

基于以上命题可知，存在一个最优的防守者策略 S^1，其中不存在上述命题描述的转移形式。令 $Tr = (i, j, a_i, a_j, \rho, \rho')$ 表示 S^1 中的一次资源转移（当目标 i 被 a_i 个资源保护，目标 j 被 a_j 个资源保护时，某资源在 $\left[\xi_\rho^i, \xi_{\rho+1}^i\right]$ 中被从目标 i 转走，并在 $\left[\xi_\rho^j, \xi_{\rho+1}^j\right]$ 到达目标 j）。定义 $\theta(Tr) = \arg\min_{t \in E}(\max_{t' \in F}\{W_i^{a_j - 1}(t'), W_j^{a_j}(t')\})(E = \left[\xi_\rho^i, \min\{\xi_{\rho+1}^i, \xi_{\rho'+1}^j - d_{ij}\}\right]$，$F = [t, t + d_{ij}])$。

命题 1.3.2. *如果转移 Tr 的发生时间被改变为 $\theta(Tr)$，防御者的最优收益不会降低。*

定义集合 $\Theta = \{\theta(Tr) : Tr = (i, j, a_i, a_j, \rho, \rho'), \forall i, j \in T, \forall a_i, a_j \in \{0, \ldots, m\}, \rho \in \{0, \ldots, R_i\}, \rho' \in \{0, \ldots, R_j\}\}$。以上命题说明存在一个防御者的最大最小均衡策略，其中所有的转移开始于此集合内的时间点。基于此，Yin 等人提出了 SCOUT-C 算法，首先获得该集合，再调用 SCOUT-D 求解博弈中防守方的最优策略。

1.4 实际应用与成功案例

基于 Stackelberg 模型的安全博弈论已经被不同领域的安全机构所应用。本节选取一些最具代表性的应用进行介绍。

1.4.1 重要基础设施保护

机场安检设置及巡逻

洛杉矶国际机场（LAX）是美国最大的目的地机场，每年的旅客流量在 7000 万左右。洛杉矶警方采取不同的措施来保护机场，包括设置车辆检查站、警察部队（警犬）在航站楼巡逻，安全筛选和检查乘客行李。安全博弈论应用主要考虑两方面的保护措施：（1）在进入机场的道路上设置车辆检查站，确定检查站的地点和检查时间；（2）制定警犬在洛杉矶国际机场 8 个航站楼之间的巡逻路线。这 8 个航站楼有不同的特

性，如大小、载客量、客流量、国际与国内航班数量。这些因素导致 8 个航站楼有不同的风险评估结果。由于有限的资源约束，可设置的车辆检查站不足以覆盖所有机场入口，警犬队伍的数量也不足以覆盖所有的航站楼。因此，采取最佳方案分配资源提高效率才能避免固定的部署模式的不足。

图 1.3　警犬在洛杉矶国际机场巡逻

基于贝叶斯 Stackelberg 博弈论的 ARMOR 系统用于规划洛杉矶国际机场检查点的设置以及警犬的巡逻路线 [43, 44]。以设置检查点为例，假设洛杉矶国际机场有 n 条进入机场的道路，警方在这 n 条道路上设置 m（$m < n$）个检查站，其中 m 是设置的最大检查点数量。恐怖分子可以从任意一个入口进行攻击。ARMOR 系统考虑不同类型的攻击者具有不同的收益函数，不同类型代表各种具有不同能力和偏好的恐怖分子。ARMOR 系统采用 DOBSS 算法计算安全部门的最佳资源分配战略 [40]，于 2007 年 8 月成功地部署在洛杉矶国际机场，并一直使用至今。

海岸警卫队巡逻

美国海岸警卫队（USCG）的任务包括维持海上安全、港口安全以及内河航道的安全。由于恐怖主义和毒品走私的威胁，这些地方面临的风险日益增加。美国海岸警卫队通过巡逻的方式来保护港口的基础设施。然而，有限的安全资源使海岸警卫队无法随时随地保护所有重要设施，攻击者有了可乘之机。为了协助美国海岸警卫队的资源分配，TEAMCORE 研究小组设计了基于 Stackelberg 博弈模型的 PROTECT 系统 [2]。

图 1.4　PROTECT 系统从 2011 年起应用于波士顿港

开发 PROTECT 系统的目的是帮助美国海岸警卫队在执行保护港口、水路、和海

岸安全（合称 PWCS）时提高效率。对 PWCS 的巡逻着眼于保护重点设施，由于资源所限，任何设施都无法获得全天候的保护，因此对资源配置的优化就变得至关重要。PROTECT 系统同时考虑攻击者的观测能力和不同设施的价值，输出美国海岸警卫队巡逻的日程表，包括什么时候开始巡逻，每次巡逻经过哪些目标区域，以及在每个目标区域里执行的巡逻活动。PROTECT 系统有很多创新点。首先，它不像以前的系统那样假设攻击者是完全理性的；第二，为了提高效率，系统在寻找均衡和最优解时采取了更加紧凑的方式来表示攻击者的策略空间；第三，PROTECT 系统通过真实的数据来评价其性能。PROTECT 模型正被拓展到纽约的港口，并且可能被更多的美国港口采用。

1.4.2 交通系统安保调度

空中警察调度

美国联邦空中警察署（FAMS）负责分配空中警察到始发地为保护美国的航班，以阻止潜在的攻击。空中警察的分配问题比 ARMOR 系统更具挑战：他们每天需要将有限数量的空中警察分配到成千上万的商业航班，空中警察的分配必须遵守各种类型的限制条件，如每一名空中警察需要飞回其基地，并满足起飞、降落、休息等很多时间上的约束。找出满足所有限制条件的最优随机调度策略是一项非常困难的任务。在此背景下，TEAMCORE 研究小组开发了 IRIS 系统 [57]，并于 2009 年 10 月开始为所有国际航班的空中警察进行调度。由于纯策略的资源分配数量随航班数量以及空中警察数量呈指数增长，DOBSS 算法无法求解最优空中警察调度策略。IRIS 系统使用更快的 ASPEN 算法产生出每天数千架商业航班的空中警察调度方案 [27]。IRIS 系统同时使用基于属性的偏好启发方法来确定 Stackelberg 博弈模型的收益函数。

运输安全管理处的机场安保

美国运输安全管理处（TSA）负责保卫全国超过 400 个机场的安全。为了协助 TSA 对资源进行有效配置，TEAMCORE 研究小组开发了 GUARDS 系统。与前文介绍的 ARMOR，IRIS 一样，GUARDS 也是基于 Stackelberg 博弈模型的，但它还能应对三项新的挑战：1. 调度上百种异质的安保活动。2. 考虑多种潜在的安全威胁。3. 开发面向上百个终端用户的系统。为了应对这些挑战，GUARDS 设计了一个新的博弈框架。这个框架能够处理异构的安保活动，并能对大量的潜在威胁进行简洁的建模。GUARDS 还在通用的求解 Stackelberg 博弈的算法的基础上提出了一种高效的求解算法。GUARDS 目前正在一个机场进行秘密测试，其表现值得期待 [47]。

城市运输系统安全

一些城市的交通系统要求乘客购票乘车，却没有采取强制措施。以洛杉矶地铁为例，它每天运送约 30 万乘客，逃票带来的损失预计每年为 560 万美元。洛杉矶警察局（LASD）雇佣一些工作人员在列车上或者站台上检票。由于巡逻检票的工作人员数量较少，不可能覆盖所有的列车和站台，因此洛杉矶警察局需要一些机制来设计检票人员的巡逻路线。如果巡逻检票的调度策略有比较固定的模式，那么逃票者可能会观察到这个模式并且利用它来逃票。目前洛杉矶警察局依赖人工制定巡逻日程。但是由于人工制定的调度策略通常有固定模式，而且日程的制定需要考虑很多复杂的因素，比如列车运行时间、发车间隔、日程长度等，制定调度策略对人来讲负担很重。TRUST 系统将地铁系统巡逻问题抽象成领导者-跟随者的 Stackelberg 博弈 [69]。领导者（洛杉矶警察局）采用混合策略，跟随者（可能逃票的乘客）观察到这个策略并决定是否买票。由于运输系统的复杂性使得可能的巡逻策略数量呈指数级增长，这给计算最优巡逻策略提出了很大挑战。为了解决此问题，TRUST 使用了紧凑的表达方式同时考虑到了时间和空间结构。洛杉矶警察局目前正在洛杉矶地铁上测试 TRUST 系统，计划根据系统产生的日程来安排巡逻，并检测收入是否增长以确定逃票者是否减少。

1.4.3 打击环境资源犯罪与城市犯罪

安全博弈理论同样被应用于打击环境资源犯罪与城市犯罪。比如，用于保护大片的森林不被乱砍滥伐。由于犯罪分子的行为受到空间限制，警方在制定策略时也需要考虑这些限制。安全博弈理论同样被应用于保护濒危动物。非法猎杀使得珍稀物种的数量大减，例如，全球老虎的数量从二十世界初到现在减少了 95%，9 种虎中的 3 种已经灭绝。为了打击资源和环境犯罪，保护濒危物种，许多国家建立了自然保护区和防卫保护区的特定机构。由于自然保护区往往面积较大，加之防卫机构的资源有限，保护自然保护区有很大难度。防卫机构通常采用巡逻的方式，但他们在设计巡逻路线时，必须考虑到多方面因素，如，有限的巡逻车如何分配，距离岗哨不同距离的区域如何巡逻等。TEAMCORE 研究小组设计的 PAWS 系统旨在帮助防卫机构设计巡逻路线。这个系统能够预测犯罪者可能攻击的区域，并据此设计最佳的巡逻方式 [13]。除此之外，新兴的智能交通系统（例如智能红绿灯）也逐渐成为潜在的攻击目标。对智能交通系统的保护因而也成为安全博弈论的研究课题 [71]。

另一个逐渐被重视的问题是鱼类的保护。水产业是许多国家的支柱产业。但是，据世界自然基金会统计，过度捕捞使得美国、加拿大和其他近大西洋国家的鳕鱼产量锐减。全球鳕鱼产量在过去的 30 年里下降了 70%。如果这个趋势持续下去，世界鳕鱼储备将在 15 年内消失。据美国国家海洋和大气局统计，非法的、未经报道的、不合

管制的捕鱼者每年生产出 1100 万至 2600 万吨海产品，这占到了一些国家的渔业的百分之四十，是对渔业可持续发展的巨大威胁。然而，全天候的监管难以实现，如何利用有限的资源监管渔业生产是一些国家必须面对的挑战。

图 1.5 安全博弈在保护野生动物和森林上发挥作用

1.4.4 打击犯罪网络

大规模恐怖袭击可能由多个犯罪集团或恐怖组织在其形成的犯罪网络上协作完成。2015 年震惊世界的巴黎恐袭便是一个典型的例子。针对这类犯罪活动，监控潜在犯罪分子间的联络是破除犯罪集团间协作的一个有效手段。及时的信息获取能帮助安全部门挫败犯罪企图。然而，即便安全部门精确掌握潜在犯罪分子的信息且拥有足够的技术手段，安全部门也常常缺乏足够的资源对其进行全天候以及一对一的监控。犯罪活动的策划者通常也能根据安全部门的监控情况随时调整其计划。在这种情况下，安全部门需要有效利用有限的安保资源选择监控策略。而安全部门与犯罪集团间的博弈很自然地对应到一个 Stackelberg 博弈，因而能从安全博弈的角度加以解决。

同本章介绍的标准安全博弈模型不同，在攻击者相互协作的场景下，我们所要考虑的攻击者策略不再是一个简单的目标，而是一个具体的协作方式。例如，犯罪策划者需要在潜在犯罪分子的通信网络上选择一个满足某些性质的子图（如连通子图）[60]。又如，当安全部门屏蔽掉部分犯罪分子间的联络通道后，犯罪分子们会在剩余的联络网络上组织若干犯罪活动；这些组织策划行为类似于合作博弈中的结盟行为（coalition formation），因而在对应的安全博弈过程中，攻击者的反馈实际上是一个合作博弈 [23]。

1.4.5 其他应用

上述横跨众多领域的案例仅仅是安全博弈研究成果以及潜在应用的冰山一角。近年来，安全博弈论的研究人员还在打击犯罪网络流、打击海上犯罪、以及网络安全方面推进了诸多的前沿工作 [8, 9, 22, 24, 31, 35, 59, 62, 70, 72, 73]。他们正致力于进一步挖掘该理论的潜力，为更多的现实问题设计出解决方案。感兴趣的读者也可参见

Tambe 以及 Sinha 等人的综述了解更多其他的安全博弈论的应用 [51, 55])。

1.5 热点与未来方向

上述成功的案例激发了安全博弈论领域更多的研究兴趣。本节选取若干当前研究热点进行简要介绍。

1.5.1 研究热点

基于机器学习的安全博弈

近年来机器学习的发展无疑也推动了基于机器学习的安全博弈论。博弈论中均衡的计算的依赖与对参与者的收益函数或矩阵的了解，在这些信息缺失或不确定的情况下均衡计算无法进行。就安全博弈而言，博弈中防御方要计算其最优策略就必须知道攻击者的收益函数。在收益函数缺失的情况下，一系列工作提出了采用机器学习的方法主动获取这些缺失的信息。在该框架下，防御者试图通过同攻击者进行交互而学习攻击者类型或针对攻击者的最优策略 [4, 7, 30, 42, 49]。防御者不断尝试不同的防御策略并观察攻击者对这些策略的最优反馈，学习的过程逐渐收敛到完全信息下的 Stackelberg 均衡。

除此之外也基于机器学习的安全博弈论也被应用于保障网络基础设施安全等场景，如城市网络、交通网络和信息网络的安全。在网络中通过部署数量有限的安全资源（由防御者控制）以防止攻击者的问题可以被建模为网络安全游戏（NSG）。NSG 的目标是为防御者找到一个纳什均衡（NE）策略。通常防御者的策略是通过基于规划的 NE 解决技术来计算的，例如，增量策略生成算法，该算法从一个受限的游戏开始，迭代扩展，直到收敛。然而，在大规模的 NSG 中，例如现实世界中的道路网络，由于攻击路径的数量十分大，基于规划的 NE 解决方法往往会失去效力。最近，人们越来越关注将深度学习（DL）与博弈论结合起来寻找 NE。它们通常以抽样方式执行，并能利用 DNN 的强大表示能力来捕捉底层巨大状态空间的结构，使它们有可能解决大规模和复杂的现实问题。然而，由于行动空间的复杂性，现有的基于 DL 的方法无法解决大规模的 NSG。

Li 等人的工作提出了一种新的学习范式 NSG-NFSP[32]，用于逼近大规模广义形式 NSG 中的 NE 策略。该方法是基于神经虚构自我游戏（NFSP），这使它具备了理论上的收敛性。算法在可扩展性和解决方案的质量方面都明显优于现有算法。

多防御者安全博弈

以往研究考虑的安全博弈场景多在单个防御者与单个攻击者之间展开。而现实世界中某些场景可能包含多个防御者。例如国际公海或边境上打击违法犯罪活动的行动往往由周边多个国家作为防御者参与其中。据报道，在马六甲海峡附近针对大型油船的海盗事件频发，为保证过往船只的安全，周边的新加坡、马来西亚、印度尼西亚等国均有防御力量投入打击海盗的行动中。在我国延绵的陆上边境线上，我方国防力量也时常与邻国安防力量同时开展巡逻，打击跨境违法犯罪活动，维护领土安全。在本土安防活动中也不乏不同安全部门保护区域重叠的情况。传统的单防御者模型无法对这些场景进行建模求解，因此出现了具有多防御者的安全博弈模型。多防御者安全博弈模型的难点在于防御者的异质性：不同防御者具有不同的防御侧重点，相同的目标对不同防御者来说可能具有不同的重要性，因此多防御者模型不能简单地通过将多个防御者看作一个等价的单防御者实现。多防御者的引入给均衡分析带来了新的复杂性。

异质多防御者安全博弈的研究可追溯到 Smith 等人的工作 [52]。后续 Lou 等人 [33, 34] 以及 Gan 等 [16, 17] 人开展了更多工作，在均衡分析和计算开展了基础性的研究。Gan 等人还研究了多防御者模型下的协同机制设计问题 [16]，考虑了如何通过机制设计的方法提升防御效率的问题。Mutzari 等人最近的工作还从非合作博弈的角度考虑了防御者之间如何通过结盟（coalition formation）的方式进行协作的问题 [37]。

多方安全博弈的另一难度在于策略空间的增长以及信息不完全问题。例如当多辆警车合作追捕逃犯时，警方的联合动作空间将随着参与抓捕的警车数呈指数增长。该问题也是一个现实的扩展形式博弈问题（extensive-form game），相比正则形式的博弈更难求解。现有算法，如 CFR，在求解该问题上效果不佳。为了解决这个难题，Xue 等人提出了一个全新的基于 CFR 的框架：CFR-MIX 以解决联合动作空间大所导致的策略空间大的问题 [63]。该方法在 Goofspiel 扑克博弈和追击问题上表现出高效的性能。

欺骗与反欺骗

欺骗与反欺骗是当前的另一研究热点，并以多种形式体现在安全博弈模型中。前面我们提到基于机器学习方法主动获取攻击者收益信息或学习安全博弈均衡的研究工作。这些方法虽然行之有效，但依赖于一个关键的假设，即在博弈的先动方（防御者）通过机器学习算法与后动方（攻击者）交互的过程中，后动方总是按照其真实的效用函数作最优的反馈。一旦该假设不成立，学习算法就存在被操纵的风险，后动方可通过假的最优反馈误导先动方的学习算法。Gan 等人首次指出了这种操纵在 Stackelberg 博弈中的可能性，并提出了*模仿欺骗*（imitative deception）的概念 [19]。采用这种欺骗方式的后动方通过模仿一个虚假的效用函数上的最优反馈误导学习算法，

使其输出基于该虚假效用函数的均衡而从中获益。模仿欺骗极具危害，在安全博弈中，这种欺骗往往导致博弈最终退化为零和博弈 [18]。对于欺骗者来说通过模仿欺骗对博弈进行操纵的余地也相当大：在一般的 Stackelberg 博弈中已被证明后动方可通过模仿欺骗将博弈操纵至任何的假均衡，当且仅当该均衡保证先动方在该博弈中的最大最小值（maximin value），亦即先动方在零和博弈中的收益；此外，欺骗者还可在多项式时间内计算这样的欺骗策略 [6]。而对于后动方而言，设计对抗欺骗的机制往往是 NP-难且难以近似的计算问题 [19]。

对安全资源的伪装和隐藏是安全博弈中的另一种欺骗，不同的是这种欺骗有利于博弈中的防御方。便衣警察的使用便是这样的实际例子。Guo 等人考虑了安全博弈论中安全资源的伪装问题，并使用完美贝叶斯均衡（perfect Bayesian equilibirum）作为模型的求解概念 [21]。该工作比较了完美贝叶斯均衡与 Stackelberg 均衡在资源可隐藏安全博弈中的解质量，并探讨了如何在资源的隐秘性与策略允诺之间寻求平衡。Rabinovich 等人以及 Xu 等人也探讨了如何通过释放信号的方法策略性地向攻击者公开信息从而诱导对安全问题更佳的攻击者行为的问题 [48, 61]。

1.5.2 挑战和未来研究方向

协同优化

协同优化，是使用者（如空中警察署）和计算机协作制定的安全资源分配策略。在安全资源调度问题中通常存在很多限制条件。防御者通常受到资源数量的限制。另外，当面临一些特殊情况或者需要额外的知识时，使用者可能需要对防御者的行为设置限制以影响结果。例如，在 IRIS 系统中，有时需要强制在某些航班上安排空中警察（例如当政府官员需要乘坐航班时）。在现有的安全博弈论应用中，通常只计算出符合所有限制条件的最优解。但由于使用者的有限理性以及对限制条件影响资源调度结果性能的有限了解，用户定义的限制条件可能会产生很差的资源分配方案，甚至导致不存在满足所有限制条件的分配方案。如果放开一些限制条件，那么分配方案的性能就会得到很大提升。放开一些限制条件有无数种方法，而计算机软件并不知道哪些限制条件可以放开，放开多少，以及放开限制条件对分配方案性能的影响。因此需要用户和计算机通过协同优化来共同制定安全资源分配策略。协同优化研究面临的挑战：第一，安全博弈和限制条件的规模使得不能使用穷尽的搜索算法去测试所有的限制条件组合；第二，用户并不完全了解放开限制条件可能引起的后果，这就需要用户偏好发掘技术（preference elicitation）的支持；第三，在用户和计算机之间关于控制权转移的决策也很具有挑战性；第四，很难评价协同优化方法的性能；第五，给计算机设计一个能够解释限制，如何影响资源分配方案性能的用户接口也是一个有挑战性的问题。

多目标优化

在现有的安全博弈论应用中，防御者总是试图最大化某个单一的目标。然而，有些领域的防御者必须同时考虑多个目标。例如，洛杉矶警察局对地铁系统的保护需要考虑逃票、普通犯罪以及恐怖袭击等多种行为。从洛杉矶警察局的视角看，每种攻击类型都具有威胁（收入减少、财产被盗、生命威胁等）。因为这些多元威胁的存在，通常没有一种单一策略可以使所有攻击类型的威胁最小化。由于针对某种特定攻击的保护可能增加其他攻击的威胁，因此必须要做权衡和折衷。多目标安全博弈可以用来应对有多个相互矛盾的优化目标的安全博弈问题。在多目标安全博弈中，不同攻击类型的威胁用不同的博弈矩阵来表示，并且不需要对攻击类型的概率分布。与只有单个最优解的单目标贝叶斯安全博弈不同，多目标安全博弈通常有一个帕累托边界。通过将帕累托边界展示给终端用户，他们能够更好地理解问题的结构和不同安全目标间的平衡。因此，终端用户在选择策略时能够做出更明智的选择。还有一些专家为了更好地建立和评估博弈模型进行了偏好选择等研究，以获取特定领域的更多数据。

其他待解决问题

在安全场景中一个基本的问题是欺骗。神话中的特洛伊木马就是欺骗的一个典型例子。计算机恶意软件的类别被称为木马，象征着恶意软件固有的欺骗行为。在 SSG 文献中已经对防御者的欺骗进行了研究，尽管是在简单的一次性交互的设定中，使用了由防卫者获得额外信息的优势所支持的信号。更广泛地说，在博弈论中的欺骗来源于信息的不对称性。以 SSG 为基础的网络安全方法中的有一些提议着眼于花费资源来为防御者提供额外的信息优势。这就是蜜罐的形式，即引诱敌人来攻击它们的虚拟系统。蜜罐技术，除了能够减少当前攻击的损失外，还能糊弄敌人来揭示他们的秘密。欺骗可能会是非常复杂的，尤其是当防御者和对手同时都在使用的时候。通过信号或设计博弈在连续博弈中提供信息优势的欺骗研究可以使 SSG 应用于高度复杂的防御-对手交互场景。

SSG 模型大多将防御者的行动指定为防御目标，但也有一些例外 (如计划封锁博弈和联盟博弈)。更广泛地来说，防御可以是防御行动、进攻行动和寻找信息的行动的组合。在一个高度使用战术的环境中，防御者必须使用所有的选项并决定采取或不采取哪些行动。值得注意的是，防御者所采取的行动并不一定总是能够改善安全状况，因为一个善于观察的对手，可能会发现防御中的弱点，例如缺乏防御者正在寻求获得的信息。因此，防御者的行为可能需要隐蔽，而且任何防御策略本身必须是安全的且不会受到对抗性攻击。

1.5.3 未来应用领域

我们期待未来安全博弈论能启发和驱动更多的研究与应用。在此我们引述 Sinha 等人在最近的综述中总结的以下未来研究方向 [51])。

新恐怖主义威胁

恐怖主义威胁在过去数十年一直在不断地演变。在反恐力量的压制下，恐怖主义的主要威胁已经逐渐从精心策划的袭击转变为独狼袭击。这些新的威胁从建模上就提出了新的问题：恐怖分子表现得更加机会主义同时又非常坚决地实施攻击。模型需要考虑到主动获取潜在独狼信息的步骤，并且这些步骤需要在有限的资源下进行。

绿色安全博弈新挑战

当前绿色安全博弈的应用已经解决了某些特定场景下的野生动物和自然资源保护问题。然而不同类型的自然资源往往具有很多独特的性质，对其的保护需要充分考虑到这些特质。这为现有解决方案的迁移带来挑战。不同自然保护区的规模、多样性，以及巡逻人员的需求和针对该区域的犯罪类型都不尽相同。因此，如何设计一个灵活多变和相对普适的绿色安全博弈框架是当前研究的一大挑战。此外，实时信息的纳入对巡逻策略也具有重要意义，特别是对于运用无人机等设备进行巡逻的场景。针对自然资源的犯罪行为的另一特点还在于其目标的多样性，偷猎、非法砍伐、侵占自然保护区土地等行为均属于该范畴并可同时同地出现，因此绿色安全博弈常常还是一个多目标优化的问题。

网络安全及隐私保护

在网络安全问题往往包含若干子问题，涉及防御者、攻击者、用户之间的相互作用。相比物理安全，网络安全的一些特性使得这个问题更为复杂，包括多变的环境状态、超大的问题围蜜、攻击的隐蔽性、多参与者等。采用安全博弈论框架研究网络安全问题之一是分配有限的人力资源来筛查自动安防系统产生的大量警报信息。防范社会工程学攻击、调整安全软件阈值、对抗网络钓鱼攻击等都是网络安全博弈关注的问题。

隐私保护是另一重要的安全问题。安全博弈在隐私审计问题上具有很大的应用前景。一般来说，软件系统的隐私问题涉及到在隐私与系统效用之间的平衡，例如社交网络中的位置信息这类隐私。要实现对这种平衡进行有意义的调节需要我们知道潜在

隐私侵犯者的行为模式，而安全博弈论为此提供了一个恰当的模型。隐私问题具有和网络安全相似的上述几个问题和挑战。

对抗电商平台欺诈行为

电子商务平台的主要功能是引导消费者对商家产品的点击量。点击量通过一个排名系统进行分配，该系统根据转化率显示销售者的产品，而转化率指消费者点击该产品后购买该产品的概率。这样做的目的是提升交易总数。通常商家者不能控制其产品的转化率，因此他们会花大量的精力获取更多的点击。商家可以通过广告等正当手段获取点击量，但由于广告费用昂贵，许多商家会转而采用一些非正当手段，例如通过虚假交易提高转化率，通过大量虚假消费者账户购买自己的产品。这些欺诈性的行为极大地降低了消费者点击量分配的有效性，并危及正常的商业环境。

当前电子商务平台主要依靠虚假检测技术来打击电子商务欺诈行为。然而欺诈手段和技术也会“与时俱进”，从而导致检测率下降。比较有效的方法是为电商平台（领导者）设计最优机制来阻止跟随者的欺诈行为。这其中的重要研究挑战包括：1）从交易数据中学习异质销售者的行为模型；2）求解数百万销售者和连续策略空间下的最优策略；3）设计针对市场演化和不确定性的鲁棒政策；4）在阻止欺诈行为的同时平衡平台上的其他目标。

参考文献

[1] An, B., Ordóñez, F., Tambe, M., Shieh, E., Yang, R., Baldwin, C., DiRenzo III, J., Moretti, K., Maule, B., Meyer, G.: A deployed quantal response-based patrol planning system for the US Coast Guard. Interfaces **43**(5), 400–420 (2013)

[2] An, B., Shieh, E., Tambe, M., Yang, R., Baldwin, C., DiRenzo, J., Maule, B., Meyer, G.: PROTECT–a deployed game theoretic system for strategic security allocation for the United States Coast Guard. AI Magazine **33**(4), 96–96 (2012)

[3] Bai, Y., Chen, L., Song, L., Xu, J.: Bayesian Stackelberg game for risk-aware edge computation offloading. In: Proceedings of the 6th ACM Workshop on Moving Target Defense, pp. 25–35 (2019)

[4] Balcan, M., Blum, A., Haghtalab, N., Procaccia, A.D.: Commitment without regrets: Online learning in Stackelberg security games. In: Proceedings of the 16th ACM Conference on Economics and Computation (EC'15), pp. 61–78 (2015)

[5] Bertsimas, D., Tsitsiklis, J.N.: Introduction to linear optimization, vol. 6. Athena Scientific Belmont, MA (1997)

[6] Birmpas, G., Gan, J., Hollender, A., Marmolejo-Cossío, F.J., Rajgopal, N., Voudouris, A.A.: Optimally deceiving a learning leader in Stackelberg games. In: Advances in Neural Information Processing Systems, vol. 33, pp. 20,624–20,635 (2020)

[7] Blum, A., Haghtalab, N., Procaccia, A.D.: Learning optimal commitment to overcome insecurity. In: Proceedings of the 28th Conference on Neural Information Processing Systems (NIPS'14), pp. 1826–1834 (2014)

[8] Cerný, J., Bosanský, B., An, B.: Finite state machines play extensive-form games. In: P. Biró, J.D. Hartline, M. Ostrovsky, A.D. Procaccia (eds.) EC '20: The 21st ACM Conference on Economics and Computation, pp. 509–533. ACM (2020)

[9] Cerný, J., Lisý, V., Bosanský, B., An, B.: Dinkelbach-type algorithm for computing

quantal stackelberg equilibrium. In: C. Bessiere (ed.) Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI'20), pp. 246–253 (2020)

[10] Conitzer, V., Sandholm, T.: Computing the optimal strategy to commit to. In: Proceedings of the 7th ACM Conference on Electronic Commerce (EC'06), pp. 82–90 (2006)

[11] Danninger-Uchida, G.E.: Carathéodory theorem. In: C.A. Floudas, P.M. Pardalos (eds.) Encyclopedia of Optimization, pp. 236–237. Springer US (2001)

[12] Daskalakis, C., Goldberg, P.W., Papadimitriou, C.H.: The complexity of computing a Nash equilibrium. SIAM Journal on Computing **39**(1), 195–259 (2009)

[13] Ford, B., Kar, D., Delle Fave, F.M., Yang, R., Tambe, M.: Paws: Adaptive game-theoretic patrolling for wildlife protection. In: Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems (AAMA'14), p. 1641–1642 (2014)

[14] Gan, J., An, B., Vorobeychik, Y.: Security games with protection externalities. In: Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI'15), p. 914–920 (2015)

[15] Gan, J., An, B., Vorobeychik, Y., Gauch, B.: Security games on a plane. In: Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI'17), p. 530–536 (2017)

[16] Gan, J., Elkind, E., Kraus, S., Wooldridge, M.: Mechanism design for defense coordination in security games. In: Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'20), p. 402–410 (2020)

[17] Gan, J., Elkind, E., Wooldridge, M.: Stackelberg security games with multiple uncoordinated defenders. In: Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'18), p. 703–711 (2018)

[18] Gan, J., Guo, Q., Tran-Thanh, L., An, B., Wooldridge, M.: Manipulating a learning defender and ways to counteract. In: Advances in Neural Information Processing Systems (NeurIPS'19), pp. 8274–8283 (2019)

[19] Gan, J., Xu, H., Guo, Q., Tran-Thanh, L., Rabinovich, Z., Wooldridge, M.: Imitative follower deception in Stackelberg games. In: Proceedings of the 2019 ACM Conference on Economics and Computation (EC'19), p. 639–657 (2019)

[20] Gu, X., Zeng, C., Xiang, F.: Applying a Bayesian Stackelberg game to secure infrastructure system: From a complex network perspective. In: Proceedings of the 2019 4th International Conference on Automation, Control and Robotics Engineering (CACRE'19). Association for Computing Machinery (2019)

[21] Guo, Q., An, B., Bosansky, B., Kiekintveld, C.: Comparing strategic secrecy and Stackelberg commitment in security games. In: Proceedings of the 26th International

Joint Conference on Artificial Intelligence (IJCAI'17), pp. 3691–3699 (2017)

[22] Guo, Q., An, B., Tran-Thanh, L.: Playing repeated network interdiction games with semi-bandit feedback. In: Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI'17), p. 3682–3690 (2017)

[23] Guo, Q., An, B., Vorobeychik, Y., Tran-Thanh, L., Gan, J., Miao, C.: Coalitional security games. In: Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'16), pp. 159–167 (2016)

[24] Guo, Q., An, B., Zick, Y., Miao, C.: Optimal interdiction of illegal network flow. In: Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI'16), pp. 2507–2513 (2016)

[25] Harsanyi, J.C., Selten, R.: A generalized Nash solution for two-person bargaining games with incomplete information. Management Science **18**(5-Part-2), 80–106 (1972)

[26] Hochbaum, D.S., Lyu, C., Ordóñez, F.: Security routing games with multivehicle Chinese postman problem. Networks **64**(3), 181–191 (2014)

[27] Jain, M., Kardes, E., Kiekintveld, C., Ordónez, F., Tambe, M.: Security games with arbitrary schedules: A branch and price approach. In: Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI'10), pp. 792–797 (2010)

[28] Jain, M., Pita, J., Tambe, M., Ordóñez, F., Paruchuri, P., Kraus, S.: Bayesian Stackelberg games and their application for security at Los Angeles International Airport. SIGecom Exchanges **7**(2), 10:1–10:3 (2008)

[29] Korzhyk, D., Conitzer, V., Parr, R.: Complexity of computing optimal Stackelberg strategies in security resource allocation games. In: Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI'10), p. 805–810 (2010)

[30] Letchford, J., Conitzer, V., Munagala, K.: Learning and approximating the optimal strategy to commit to. In: International Symposium on Algorithmic Game Theory (SAGT'09), pp. 250–262 (2009)

[31] Li, S., Zhang, Y., Wang, X., Xue, W., An, B.: CFR-MIX: solving imperfect information extensive-form games with combinatorial action space. In: Z. Zhou (ed.) Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI'21), pp. 3663–3669 (2021)

[32] Li, S., Zhang, Y., Wang, X., Xue, W., An, B.: Cfr-mix: Solving imperfect information extensive-form games with combinatorial action space (2021)

[33] Lou, J., Smith, A.M., Vorobeychik, Y.: Multidefender security games. IEEE Intelligent Systems **32**(1), 50–60 (2017)

[34] Lou, J., Vorobeychik, Y.: Equilibrium analysis of multi-defender security games. In:

Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI'15), pp. 596–602 (2015)

[35] Ma, X., An, B., Zhao, M., Luo, X., Xue, L., Li, Z., Miu, T.T.N., Guan, X.: Randomized security patrolling for link flooding attack detection. IEEE Trans. Dependable Secur. Comput. **17**(4), 795–812 (2020)

[36] McKelvey, R.D., Palfrey, T.R.: Quantal response equilibria for normal form games. Games and Economic Behavior **10**(1), 6 – 38 (1995)

[37] Mutzari, D., Gan, J., Kraus, S.: Coalition formation in multi-defender security games. In: Proceedings of the 35rd AAAI Conference on Artificial Intelligence (AAAI'21), to appear (2021)

[38] Oliva, G., Setola, R., Tesei, M.: A Stackelberg game-theoretical approach to maritime counter-piracy. IEEE Systems Journal **13**(1), 982–993 (2018)

[39] Paruchuri, P., Pearce, J.P., Marecki, J., Tambe, M., Ordóñez, F., Kraus, S.: Efficient algorithms to solve Bayesian Stackelberg games for security applications. In: Proceedings of the 23rd AAAI Conference on Artificial Intelligence (AAAI'08), pp. 1559–1562 (2008)

[40] Paruchuri, P., Pearce, J.P., Marecki, J., Tambe, M., Ordóñez, F., Kraus, S.: Playing games for security: An efficient exact algorithm for solving Bayesian Stackelberg games. In: Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'08), pp. 895–902 (2008)

[41] Paruchuri, P., Pearce, J.P., Tambe, M., Ordonez, F., Kraus, S.: An efficient heuristic approach for security against multiple adversaries. In: Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'07), pp. 311–318 (2007)

[42] Peng, B., Shen, W., Tang, P., Zuo, S.: Learning optimal strategies to commit to. In: Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI'19), pp. 2149–2156 (2019)

[43] Pita, J., Jain, M., Marecki, J., Ordóñez, F., Portway, C., Tambe, M., Western, C., Paruchuri, P., Kraus, S.: Deployed ARMOR protection: The application of a game theoretic model for security at the los angeles international airport. In: Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'08), p. 125–132 (2008)

[44] Pita, J., Jain, M., Ordónez, F., Portway, C., Tambe, M., Western, C., Paruchuri, P., Kraus, S.: ARMOR security for Los Angeles International Airport. In: Proceedings of the 23rd AAAI Conference on Artificial Intelligence (AAAI'08), pp. 1884–1885 (2008)

[45] Pita, J., Jain, M., Ordóñez, F., Portway, C., Tambe, M., Western, C., Paruchuri, P., Kraus,

S.: Using game theory for Los Angeles Airport security. AI Magazine **30**(1), 43 (2009)

[46] Pita, J., Jain, M., Tambe, M., Ordóñez, F., Kraus, S.: Robust solutions to Stackelberg games: Addressing bounded rationality and limited observations in human cognition. Artificial Intelligence **174**(15), 1142–1171 (2010)

[47] Pita, J., Tambe, M., Kiekintveld, C., Cullen, S., Steigerwald, E.: GUARDS: Game theoretic security allocation on a national scale. In: Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'11), p. 37–44 (2011)

[48] Rabinovich, Z., Jiang, A.X., Jain, M., Xu, H.: Information disclosure as a means to security. In: Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'15), pp. 645–653 (2015)

[49] Roth, A., Ullman, J., Wu, Z.S.: Watch and learn: Optimizing from revealed preferences feedback. In: Proceedings of the 48th Annual ACM Symposium on Theory of Computing (STOC'16), pp. 949–962 (2016)

[50] Shieh, E., An, B., Yang, R., Tambe, M., Baldwin, C., DiRenzo, J., Maule, B., Meyer, G., Moretti, K.: Protect in the ports of Boston, New York and beyond: experiences in deploying stackelberg security games with quantal response. In: Handbook of computational approaches to counterterrorism, pp. 441–463. Springer (2013)

[51] Sinha, A., Fang, F., An, B., Kiekintveld, C., Tambe, M.: Stackelberg security games: looking beyond a decade of success. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI'18), pp. 5494–5501 (2018)

[52] Smith, A., Vorobeychik, Y., Letchford, J.: Multi-defender security games on networks. ACM SIGMETRICS Performance Evaluation Review **41**(4), 4–7 (2014)

[53] von Stackelberg, H.: Marktform und Gleichgewicht. J. Springer (1934)

[54] von Stackelberg, H.: Market structure and equilibrium. Springer Science & Business Media (2010)

[55] Tambe, M.: Security and Game theory: Algorithms, Deployed Systems, Lessons Learned. Cambridge University Press (2011)

[56] Tsai, J., Qian, Y., Vorobeychik, Y., Kiekintveld, C., Tambe, M.: Bayesian security games for controlling contagion. In: Proceedings of the ASE/IEEE International Conference on Social Computing(SocialCom), pp. 33–38. IEEE (2013)

[57] Tsai, J., Rathi, S., Kiekintveld, C., Ordóñez, F., Tambe, M.: IRIS-a tool for strategic security allocation in transportation networks. In: Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'09), pp. 37–44 (2009)

[58] Tsai, J., Yin, Z., Kwak, J.y., Kempe, D., Kiekintveld, C., Tambe, M.: Urban security: game-theoretic resource allocation in networked physical domains. In: Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI'10), pp. 881–886 (2010)

[59] Wang, X., An, B., Strobel, M., Kong, F.: Catching Captain Jack: Efficient time and space dependent patrols to combat oil-siphoning in international waters. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI'18), pp. 208–215 (2018)

[60] Wang, Z., Yin, Y., An, B.: Computing optimal monitoring strategy for detecting terrorist plots. In: Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI'16), pp. 637–643 (2016)

[61] Xu, H., Rabinovich, Z., Dughmi, S., Tambe, M.: Exploring information asymmetry in two-stage security games. In: Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI'15), pp. 1057–1063 (2015)

[62] Xue, W., Zhang, Y., Li, S., Wang, X., An, B., Yeo, C.K.: Solving large-scale extensive-form network security games via neural fictitious self-play. In: Z. Zhou (ed.) Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI'21), pp. 3713–3720 (2021)

[63] Xue, W., Zhang, Y., Li, S., Wang, X., An, B., Yeo, C.K.: Solving large-scale extensive-form network security games via neural fictitious self-play (2021)

[64] Yang, R., Jiang, A.X., Tambe, M., Ordóñez, F.: Scaling-up security games with boundedly rational adversaries: A cutting-plane approach. In: Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI'13), pp. 404–410 (2013)

[65] Yang, R., Kiekintveld, C., Ordóñez, F., Tambe, M., John, R.: Improving resource allocation strategy against human adversaries in security games. In: Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI'11) (2011)

[66] Yang, R., Ordonez, F., Tambe, M.: Computing optimal strategy against quantal response in security games. In: Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'12), pp. 847–854 (2012)

[67] Yin, Y., An, B., Jain, M.: Game-theoretic resource allocation for protecting large public events. In: Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI'14), pp. 826–833 (2014)

[68] Yin, Y., Xu, H., Gan, J., An, B., Jiang, A.X.: Computing optimal mixed strategies for security games with dynamic payoffs. In: Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI'15), pp. 681–687 (2015)

[69] Yin, Z., Jiang, A.X., Tambe, M., Kiekintveld, C., Leyton-Brown, K., Sandholm, T., Sullivan, J.P.: TRUSTS: Scheduling randomized patrols for fare inspection in transit systems using game theory. AI Magazine **33**(4), 59 (2012)

[70] Zhang, Y., An, B.: Converging to team-maxmin equilibria in zero-sum multiplayer games. In: Proceedings of the 37th International Conference on Machine Learning (ICML'20), *Proceedings of Machine Learning Research*, vol. 119, pp. 11,033–11,043. PMLR (2020)

[71] Zhang, Y., An, B., Tran-Thanh, L., Wang, Z., Gan, J., Jennings, N.R.: Optimal escape interdiction on transportation networks. In: Proceedings of the 26th International Joint Conference on Artificial Intelligence (AAAI'17), p. 3936–3944 (2017)

[72] Zhang, Y., Guo, Q., An, B., Tran-Thanh, L., Jennings, N.R.: Optimal interdiction of urban criminals with the aid of real-time information. In: The 33rd AAAI Conference on Artificial Intelligence (AAAI'19), pp. 1262–1269 (2019)

[73] Zhao, M., An, B., Kiekintveld, C.: Optimizing personalized email filtering thresholds to mitigate sequential spear phishing attacks. In: D. Schuurmans, M.P. Wellman (eds.) Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI'16), pp. 658–665 (2016)