中文引用格式:黄积成,陈文静,贾楠. 关键基础设施安全资源配置优化研究[J]. 中国安全科学学报,2026,36(2):218-226.
英文引用格式:HUANG Jicheng,CHEN Wenjing,JIA Nan. Research on security resource allocation optimization of critical infrastructure[J]. China Safety Science Journal,2026,36(2):218-226.

关键基础设施安全资源配置优化研究*

黄积成, 陈文静** 副教授, 贾 楠
(中国人民公安大学 信息网络安全学院,北京 100038)

中图分类号:X956 **文献标志码**:A **DOI**: 10.16265/j.cnki.issn1003-3033.2026.02.0519
基金项目:中央高校基本科研业务费专项资金资助(2024JKF03);北京市社会科学基金资助(21JCC108)。

【摘 要】 为提高关键基础设施安全资源配置的合理性与有效性,解决传统方法在安全资源分配均衡性与防护效能方面的不足,基于前景理论(PT)、敌手入侵序列图(ASD)和Stackelberg博弈理论,构建关键基础设施安全资源配置优化模型。首先,通过构建ASD-Stackelberg博弈模型求解关键基础设施内部安全资源的最优空间配置;其次,引入PT理论刻画攻防双方的有限理性特征,确定最优安全资源投入阈值;最后,以核电厂安防系统为案例开展实证研究。结果表明:相较于传统经验配置方法,文中所建模型输出的最优资源配置方案,能够使核电厂整体安防效能提升25.9%,且能够准确找出最优安全资源投入量。该结果证实模型在求解最优配置方案方面的有效性与可行性,且其在提升安全防护能力方面具有显著优越性。

【关键词】 关键基础设施; 安全资源配置; 前景理论(PT); 敌手入侵序列图(ASD); Stackelberg博弈模型

Research on security resource allocation optimization of critical infrastructure

HUANG Jicheng, CHEN Wenjing, JIA Nan
(School of Information Network Security, People's Public Security University of China, Beijing 100038, China)

Abstract: In order to improve the rationality and effectiveness of critical infrastructure security resource allocation, and solve the shortcomings of traditional methods in the balance of security resource allocation and protection effectiveness, an optimization model of critical infrastructure security resource allocation was constructed based on PT, ASD and Stackelberg game theory. Firstly, the ASD-Stackelberg game model was constructed to solve the optimal spatial allocation of security resources within critical infrastructure. Secondly, the PT theory was introduced to describe the bounded rationality characteristics of both attack and defense sides, and the optimal security resource investment threshold was determined. Finally, the nuclear power plant security system was taken as a case study. The verification and analysis results show that compared with the traditional empirical configuration method, the optimal resource allocation scheme output by this model has significantly improved the overall security effectiveness of nuclear power plants by

* 文章编号:1003-3033(2026)02-0218-09; 收稿日期:2025-09-15; 修稿日期:2025-12-08
** 通信作者:陈文静(1977—),女,河北张家口人,硕士,副教授,主要从事安全防范系统风险评估与系统规划设计研究。E-mail:chenwenjing@ppsuc.edu.cn。

25.9%, and can accurately identify the optimal amount of security resource investment. This result not only confirms the effectiveness and feasibility of the model in solving the optimal configuration scheme, but also quantitatively establishes its significant superiority in enhancing security protection capabilities.

Keywords: critical infrastructure; security resource allocation; prospect theory (PT); adversary intrusion sequence diagram (ASD); Stackelberg game model

0 引　言

国家的关键基础设施一旦遭受损害或功能丧失,将对国家安全和社会秩序造成毁灭性的打击[1]。大量的公共安全事件和案例表明:国家的关键基础设施通常是恐怖分子或极端刑事犯罪分子攻击的首选目标,如 2016 年,布鲁塞尔机场和地铁爆炸袭击事件[2];2019 年,沙特阿美石油设施袭击事件[3]等。在安全资源有限的情况下,合理调度和配置安全资源,提高关键基础设施的防御能力,使关键基础设施在遭遇恐怖袭击、刑事犯罪等社会公共安全事件时损失最小[4],已成为当前亟待解决的问题。

合理地调度和配置安全资源以抵御恐怖袭击,需要明确考虑攻击者的行为策略和自适应性[5],因而,学者们采用博弈论寻求该问题的解决方法。MAJOR[6]将博弈论理论引入恐怖袭击与关键基础设施安全资源配置研究中,建模分析恐怖主义袭击风险,发现可通过量化攻防互动来优化安全资源配置。GORDON[7]通过建立目标损失概率模型,量化了恐怖分子选择不同目标发动袭击的可能性,并依据目标受袭击的优先级提出安全资源的配置策略。FENG Qilin 等[8]采用贝叶斯博弈方法,结合风险评估理论,研究了城市内多个化工厂面临潜在恐怖袭击时的安全资源分配问题。CUI Yan 等[9]构建贝叶斯网络模型和恶意入侵博弈模型,研究油气管道失效机制,并通过风险评估模型,识别威胁,计算出每段管道所需的安全资源投入。巩绪福等[10]基于 Stackelberg 博弈建立海事安全资源调度模型,通过数学建模优化了江苏海岸线港口的巡逻路线,提升了巡逻资源的利用效率。王东营等[11]引入博弈论思想,提出一种主客观混合赋权方法,对油气管道段进行风险排序,据此将安全资源优先配置于高风险管段,增强了管道系统的整体安全性。随着博弈论在安全领域的发展,众多学者将演化博弈[12]、信号博弈、欺骗博弈[13]等博弈模型引入安全资源配置研究中。然而,当前研究仍存在以下几方面局限:①现有研究大多为静态博弈,缺少攻防双方的动态分析,未能充分发挥防御者的先动优势,导致安全资源配置效率不足[14]。②现有研究主要应用在多保护目标场景中,优化安全资源在多个目标之间的分配;而在单一目标保护场景中又面临潜在攻击者信息未知的困难。③现有的博弈模型假设的攻击者和防御者在决策时都是绝对理性的,但在现实中,攻防双方决策时展示其并非绝对理性,如防御者可能会投入大于其保护目标价值的安全资源去规避风险,这使得模型与真实场景存在不兼容问题。

鉴于此,笔者考虑单一目标关键基础设施的现实情况,结合敌手入侵序列图(Adversary Sequence Diagrams,ASD)和 Stackelberg 博弈理论,构建关键基础设施安全资源配置优化模型,充分发挥防御者先行配置安全资源的先动优势,优化关键基础设施内部的安全资源配置;同时,引入前景理论(Prospect Theory,PT),考虑人的有限理性行为,模拟攻防双方的资源投入,计算最优的安全资源投入量;并将该模型应用于 G 市核电厂,验证模型的可行性和适用性,以期为关键基础设施安全资源配置研究提供新思路。

1 单一目标保护安全资源配置问题描述

核设施、军工厂等关键基础设施构成典型的"高价值单一目标"保护场景,此类场景的核心特征在于目标价值高度集中、物理环境封闭、防护体系纵深化[15],如图 1 所示。

攻击者总想从外部(非限制区)入侵到关键基础设施内部(核心区)进行破坏,而关键基础设施安全管理人员(防御者)投入大量的安全资源,如人力资源、物理设施和技术系统,分别配置于周界围墙,以及车辆、人行通道等,探测攻击者的入侵和抵御破坏。如何科学合理地配置这些安全资源,达到最好的安全防护效果,是文中重点讨论的问题。为了使模型构建更加科学严谨,提出以下基本假设:①攻击者足够了解关键基础设施内部的安全资源配置。研究表明:攻击者会提前了解到 80%以上的信息[4]。②攻击者的攻击获益与防御者保护目标的安全价值

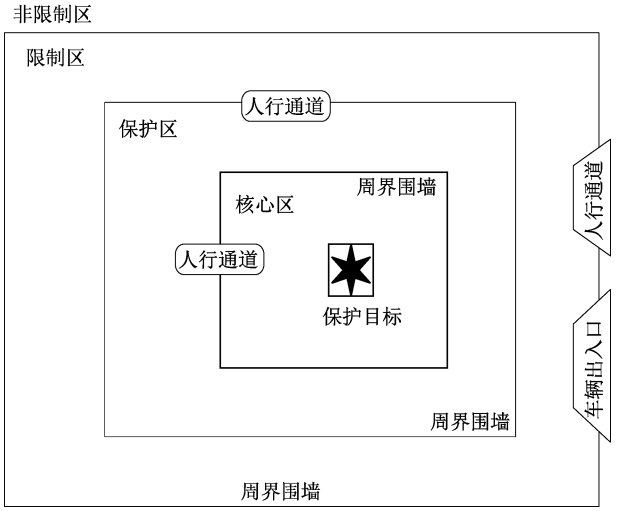

图 1　关键基础设施二维简化平面

Fig. 1　Two-dimensional simplified plan of critical infrastructure

是相等的,为简化模型,通常将其假定为 1 个常数。③文中讨论的有限理性行为并非绝对不理性的行为(缺乏目的导向性的极端行为),一方面这是符合对现实场景中对攻击者与防御者的假设,另一方面绝对非理性行为缺乏可分析性,探讨其在安全资源配置中的应用无实际意义。

2　PT-ASD-Stackelberg 博弈模型构建

2.1　ASD-Stackelberg 博弈模型构建

ASD 模型能够通过分析假设的敌手为了达到某种目的所需要经过的各个障碍,从而找出实体保护系统的不足。该模型是一种定性的分析模型,并没有定量的给出系统中最薄弱的路径,但是它能够有效地抽象出攻击者入侵关键基础设施的入侵路径[16]。

Stackelberg 博弈模型描述的是在动态博弈中,由于博弈双方策略选择顺序的不同,出现优先选择策略的先行者(通常称为领导者)以及观察到领导者选择策略后再做出决策的追随者,领导者在决策时会考虑到追随者对自己的决策做出的反应,而制定己方决策,以最大化己方收益[17]。在 ASD-Stackelberg 博弈模型中,防御者 D 为领导者,追随者为潜在攻击者 A。

根据关键基础设施的二维平面图,抽象建模为 ASD-Stackelberg 安防网络图,如图 2 所示,图 2 中,圆圈部分表示车辆、人行通道和周界围墙等配置安全资源的节点(统称安防节点),T 为核心区破坏保护目标。在实际情况中,防御者先行在各安防节点配置安全资源 $\boldsymbol{x}=[x_1,x_2,\cdots,x_n]$;$A$ 在观察到内部安全资源配置后,从中找到 1 条防御最薄弱的路径入侵;D 的目的是利用有限的安全资源来优化安防网络,使得攻击者从最薄弱的攻击路径入侵所付出的代价最大,即最不容易攻击成功。

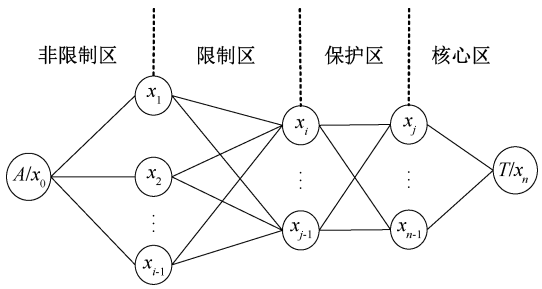

图 2　ASD-Stackelberg 安防网络

Fig. 2　Security resource configuration graph based on ASD-Stackelberg

在 ASD-Stackelberg 博弈模型中,A 从非限制区 x_0 入侵到 T,需要经过若干层防护,每层防护都有对应的若干个安防节点,攻击者在经过每层防护时都会尽可能选择最薄弱的安防节点通过。此外,还需要考虑相邻层级 2 个安防节点之间的距离(图 2 中安防节点间的线段),距离越长,攻击者所花费的时间就越多,入侵被发现的概率越大。攻击者选择安防节点时,需要考虑距离因素,通过下式归一化处理后作为路径权重 $\boldsymbol{J}=[C_{01},C_{02},\cdots,C_{ij},\cdots,C_{jn}]$。

$$C_{ij}=\frac{Y_{ij}-Y_{\min}}{Y_{\max}-Y_{\min}} \tag{1}$$

式中:C_{ij} 为路径权重向量中的元素;Y_{ij} 为节点 x_i 到 x_j 的距离,m;$Y_{\min}$ 为最短距离,m;$Y_{\max}$ 为最长距离,m。

将 A 经过某条路径,通过对应安防节点被发现或中断的值,即路径权重与安防节点安全资源量的乘积定义为路径代价[16]。由此可以定义 A 到重点目标的路径代价为关于 $\boldsymbol{x}$ 的函数,以图 2 为例,其中,$x_0\rightarrow x_1\rightarrow x_i\rightarrow x_j\rightarrow x_n$ 这 1 条路线的路径代价计算公式 $f_1(x)$ 表示为:

$$f_1(\boldsymbol{x})=C_{01}\times x_1+C_{1i}\times x_i+C_{ij}\times x_j+C_{jn}\times x_n \tag{2}$$

假定 A 从外部到 T 有 k 条路径可供选择,那么所有入侵路线的路径代价可以表示为一组路径代价向量 $\boldsymbol{F}(\boldsymbol{x})$:

$$\boldsymbol{F}(\boldsymbol{x})=[f_1(\boldsymbol{x}),f_2(\boldsymbol{x}),\cdots,f_k(\boldsymbol{x})] \tag{3}$$

A 在了解到关键基础设施内部安全资源配置的前提下,总是可以找到路径代价最小的入侵路线,对应最小损失代价,用 A_{ct} 表示攻击者的攻击代价(cost):

$$A_{ct} = \min[f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \cdots, f_k(\boldsymbol{x})] \tag{4}$$

对于 D 而言,调整各安防节点的安全资源配置,尽可能使得攻击者在选择关键基础设施中最薄弱的路径入侵时,攻击者所付出的代价是最大的。最大最小准则[18]来求解,目标函数如下:

$$\max_{x_1,x_2,\cdots,x_n} \{\min[f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \cdots, f_k(\boldsymbol{x})]\}$$

$$\begin{cases} \boldsymbol{Px} \leqslant \boldsymbol{b} \\ \boldsymbol{Qx} = \boldsymbol{e} \\ \delta(\boldsymbol{x}) \leqslant 0 \\ \varphi(\boldsymbol{x}) = 0 \end{cases} \tag{5}$$

式中:$\boldsymbol{x} = [x_1, x_2, \cdots, x_n]$ 表示各安防节点配置的安全资源量,模型约束条件由线性与非线性部分共同构成:其中,$\boldsymbol{Px} \leqslant \boldsymbol{b}$ 和 $\boldsymbol{Qx} = \boldsymbol{e}$ 分别定义了资源分配的线性不等式与等式约束,$\boldsymbol{P}$、$\boldsymbol{Q}$ 为系数矩阵,$\boldsymbol{b}$ 和 $\boldsymbol{e}$ 为对应常数向量;此外,$\delta(\boldsymbol{x}) \leqslant 0$ 和 $\varphi(\boldsymbol{x}) = 0$ 则分别引入了非线性不等式与等式约束[19]。在具体的安防实践中,需要考虑一些实际情况,如关键基础设施的安全资源总量是有限的;部分安全资源不便于转移或转移需要消耗更高昂成本;部分节点需要分配更多安全资源等。

求解出 $\boldsymbol{x}^* = [x_1^*, x_2^*, \cdots, x_n^*]$,就是 D 的最优安全资源配置策略,依据求解结果配置关键基础设施安防节点的安全资源量,使得攻击者入侵路线中最薄弱的路径尽可能配置更多的安全资源。

在确定攻防双方策略集后,若要进一步计算双方的资源投入,首先需要明确其收益函数。假定关键基础设施保护目标的安全价值为 ω,其中,$\omega>0$ 表示遭受攻击后内部的损失后果,具体数值可由关键基础设施内部安全管理人员评估得出。A 的目的是破坏保护目标,即 A 的获益(obtain)用 A_{ob} 表示,其中 $A_{ob}=\omega$,而攻击所需要付出的代价为 A_{ct}。D 配置安全资源的获益包含 2 部分:一部分是关键基础设施内部安全价值不受损害获益 ω;另外一部分安防系统的效能提升,使得攻击方入侵最少要付出代价 A_{ct},即最短代价路径获益,与木桶原理相似,其他路径的获益也都用最短路径代价 A_{ct} 来衡量,整体安防系统包含 k 条路径,安防系统效能提升获益为 kA_{ct}。用 D_{ob} 表示 D 的获益(obtain),防御者获益为:$D_{ob}=\omega+kA_{ct}$。D 付出代价(cost)用 D_{ct} 表示,代表防御成本,即安全资源投入量。

攻防双方收益函数 U_a、U_d 如下:

$$\begin{cases} U_a = A_{ob} - A_{ct} \\ U_d = D_{ob} - D_{ct} \end{cases} \tag{6}$$

在 ASD-Stackelberg 博弈模型中,攻防双方都尽可能使得己方获得最大收益,对于绝对理性的攻击者和防御者来说,当 $U_a<0$ 时,A 是不会发动攻击的;$U_d<0$ 时,D 会适当减少安全资源投入。

2.2　考虑有限理性行为的 PT

然而,在现实场景中,A 与 D 往往表现出有限理性特征。如美国小型机场安检资源过度配置[20]、偏远无人区电信基站的安全资源过度投资等,实践中常出现将过量安全资源配置于价值普通的基础设施情况,这正符合 PT 所揭示的风险决策特征[21]:决策者在面对收益时倾向于风险规避(如防御者过度保护低价值目标),而在面临损失时则转向风险寻求(如攻击者甘冒更大风险)。可见:引入 PT 框架,可以更准确地刻画安全博弈参与者在风险情境下的有限理性决策行为。

基于 PT 的安全资源配置决策过程分为风险编码和策略评估 2 个阶段。在风险编码阶段,将可能出现的安全资源配置结果分为攻防双方的获益和成本或代价(式(6))。在策略评估阶段,决策者通过非对称价值函数评估不同安全资源投资方案下攻防双方的效用价值 V_a、V_d,其中,对安全损失的规避倾向往往大于对同等程度安全增益的追求,最终形成具有风险偏好的最优安全资源投入决策。非对称价值函数如下式:

$$\begin{cases} V_a = v(-A_{ct}) + v(A_{ob}) \\ V_d = v(-D_{ct}) + v(D_{ob}) \end{cases} \tag{7}$$

式中 v 为前景价值函数。$v(\Delta x)$ 表示如下:

$$v(\Delta x) = \begin{cases} (\Delta x)^g \Delta x \geqslant 0 \\ -\lambda(-\Delta x)^l \Delta x < 0 \end{cases} \tag{8}$$

式中:Δx 为相对参考点的得失,g 和 l 为指数参数,并且满足 $g \in (0,1)$ 和 $l \in (0,1)$;g 为收益之上的风险厌恶,其值越大说明在增益域中的风险规避越小,l 为损失之上的风险寻求,其值越小说明在损益域中寻求更小的风险;λ 为损失厌恶系数,通常 $\lambda>1$,因为人们普遍认为,理性人对损失比收益更加敏感[22]。

3　关键基础设施安全资源配置优化

以 G 市核电厂为例验证该模型,利用 ASD 模型

抽象建模,得出核电厂 ASD-Stackelberg 安防网络图,如图 3 所示。

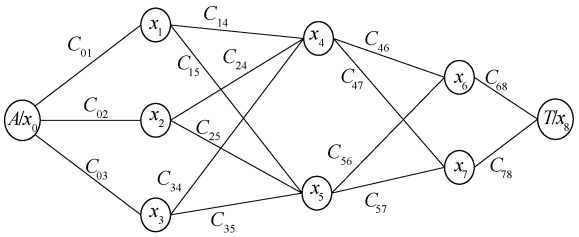

图 3　核电厂的 ASD-Stackelberg 安防网络

Fig. 3　Nuclear power plant safety resource allocation diagram based on ASD-Stackelberg

3.1　确定最优攻击路径

1）核电厂内安全资源量。在安防系统效能评估中,通常采用专家评分法、熵权法,以及主客观博弈论赋权法等量化评估安防节点的安全资源[23]。现假定核电厂内有限安全资源总量为 $Z=800$,包含有形的(如人力、设备、资金)和无形的(如时间、技术)安全资源,将安全资源均分给各安防节点,令 $x_i=100, i=1,2,\cdots,8$。

2）计算路径权重。依据式(1)将核电厂内部各路径标准化,取 $Y_{\min}=0$ 计算。其中,节点 x_0 至 x_1,x_2,x_3 的路径均表示从周界外围进入关键基础设施内部,其距离 $Y_{\max}$ 是远大于关键基础设施内部其他路径的,取 $C_{01}=C_{02}=C_{03}=1$,同时,为了与内部路径区分和计算方便,取 $Y_{\max}=200$ m。通过测量核电厂内部的距离,得出各安防节点之间路径长度和标准化后的权重以及路径花费,见表 1。

表 1　节点间路径长度、路径权重以及路径代价

Table 1　Path length, path weight and path cost between nodes

路径	C_{01}	C_{02}	C_{03}	C_{14}	C_{15}	C_{24}	C_{25}	C_{34}	C_{35}	C_{46}	C_{47}	C_{56}	C_{57}	C_{68}	C_{78}
距离/m	200	200	200	180	80	150	50	80	100	80	60	60	60	40	30
标准化	1.00	1.00	1.00	0.90	0.40	0.75	0.25	0.40	0.50	0.40	0.30	0.30	0.30	0.20	0.15
$C_{ij}\times D_j$	100	100	100	90	40	75	25	40	50	40	30	30	30	20	15

3）最短代价路径计算。根据 ASD-Stackelberg 博弈模型,A 在提前了解到核电厂内部安全资源配置情况时,选择最优攻击路线。在核电厂 ASD-Stackelberg 安防网络图中(图 4),从初始点 x_0 到 T 共有 12 条路径,在博弈中也称为攻击者的攻击策略 $S_{a1},S_{a2},\cdots,S_{a12}$,通过枚举法列出所有路径代价并排序,见表 2。

表 2　安防网络起始点 x_0 到 T 的路径及代价

Table 2　Path and cost from starting point x_0 to point T of security network

策略	路径	路径代价
S_{a1}	$x_0\rightarrow x_2\rightarrow x_5\rightarrow x_7\rightarrow x_8/T$	170
S_{a2}	$x_0\rightarrow x_2\rightarrow x_5\rightarrow x_6\rightarrow x_8/T$	175
S_{a3}	$x_0\rightarrow x_1\rightarrow x_5\rightarrow x_7\rightarrow x_8/T$	185
S_{a4}	$x_0\rightarrow x_3\rightarrow x_4\rightarrow x_7\rightarrow x_8/T$	185
S_{a5}	$x_0\rightarrow x_1\rightarrow x_5\rightarrow x_6\rightarrow x_8/T$	190
S_{a6}	$x_0\rightarrow x_3\rightarrow x_5\rightarrow x_7\rightarrow x_8/T$	195
S_{a7}	$x_0\rightarrow x_3\rightarrow x_4\rightarrow x_6\rightarrow x_8/T$	200
S_{a8}	$x_0\rightarrow x_3\rightarrow x_5\rightarrow x_6\rightarrow x_8/T$	200
S_{a9}	$x_0\rightarrow x_2\rightarrow x_4\rightarrow x_7\rightarrow x_8/T$	220
S_{a10}	$x_0\rightarrow x_1\rightarrow x_4\rightarrow x_7\rightarrow x_8/T$	235
S_{a11}	$x_0\rightarrow x_2\rightarrow x_4\rightarrow x_6\rightarrow x_8/T$	235
S_{a12}	$x_0\rightarrow x_1\rightarrow x_4\rightarrow x_6\rightarrow x_8/T$	250

从表 2 可以看出,攻击者采取策略 S_{a1}:$x_0\rightarrow x_2\rightarrow x_5\rightarrow x_7\rightarrow x_8/T$ 对核电厂入侵破坏所付出的代价最小,即该路径是核电厂安防网络中的最薄弱路径,也是攻击者的最优攻击策略,此时所付出代价为 170。

3.2　安全资源配置优化

根据 ASD-Stackelberg 博弈模型,防御者可以预测攻击者的攻击路径,优化安全资源配置。文中考虑 2 种情形:①不增加安全资源投入时,仅利用原有安全资源配置优化;②额外增加安全资源投入时,寻找最优的安全资源配置和最优安全资源投入量。根据核电厂内部安全管理人员的评估,核电厂内保护目标遭受攻击后损失安全价值为 $\omega=1\ 200$。

3.2.1　不增加安全资源的配置优化

在不增加安全资源投入时,安防节点总资源为 $Z=800$,优化现有安全资源的配置,根据式(5)中的目标函数,这 12 条路径具体优化过程如下:

$$\max_{x_1,x_2,\cdots x_8}\{\min[f_1(\boldsymbol{x}),f_2(\boldsymbol{x}),\cdots,f_{12}(\boldsymbol{x})]\} \quad (9)$$

在实际情况中,由于核电厂的周界围墙过大,需要的安全资源比人行通道和车辆出入口资源多;同时 x_8 作为基础设施的最后一道防御,且是攻击者必须经过的节点,因此,x_8 节点所分配的资源要高于其他节点;为保证每个节点都有一定的安全资源,所有安防节点至少配置平均安全资源的 25%。则约束条件可以设定为:

$$\begin{cases} x_1 + x_2 + \cdots + x_n = Z \\ x_1 \leqslant x_3 \\ x_2 \leqslant x_3 \\ x_8 \geqslant x_i \forall \ \ i = 1,2,\cdots,7 \\ x_i \geqslant Z/4n \forall \ \ i = 1,2,\cdots,8 \end{cases} \tag{10}$$

式中 $n=8$ 对应核设施内部 8 个安防节点。

最后解得各安防节点的最优安全资源分配,见表 3。优化后,最短路径付出代价为 214,相较于未优化前采用传统平均分配方式最短路径代价 170,采用 ASD-Stackelberg 博弈模型进行安全资源配置可以使得安全资源利用率提升 25.9%。

表 3　优化后各安防节点安全资源配置

Table 3　Security resource configuration of each security node after optimization

节点	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
安全资源	157	165	169	25	56	25	32	169

从表 3 可以看出,安全资源更多被分配在第 1 层防线的节点上,这与安防纵深防护原理基本吻合,尽可能把所有攻击者拦在关键基础设施的最外层防御外面,最大程度减小伤亡损失。

优化前后攻击者使用不同策略的路径代价如图 4 所示。D 采用 ASD-Stackelberg 博弈模型优化安全资源配置后,A 在最短代价路径上花费代价更大。此外,优化后的安防网络 12 条路径攻击代价在 220 上下波动,说明每条路径上配置的安全资源更加均衡。与木桶原理相似,安防系统的防护能力往往采用最薄弱路径来衡量,优化后最薄弱路径的攻击代价大幅度提高,说明整体安防系统的防护能力显著增强。

综上,ASD-Stackelberg 博弈模型可以在安全资源总量不变的前提下,通过调整安全资源配置优化安防系统,并将部分安全资源往最短代价路径上倾斜,从而增加最薄弱路径的安全资源,使得安全资源

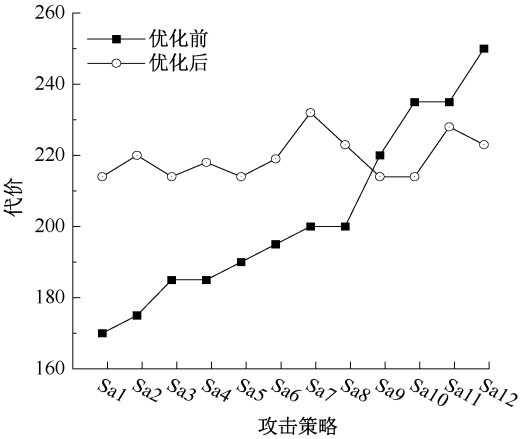

图 4　优化前后攻击者使用不同策略的路径代价

Fig. 4　Path costs for attackers using different strategies before and after optimization

配置更加均衡,提高安全资源利用效率。在原有安防资源不变的前提下,通过优化,将原本最短路径花费代价从 170 提升至 214,整体核电厂安全防护能力提升了 25.9%。

3.2.2　增加安全资源的配置优化与最优投入量

额外增加安全资源时,假定额外投入安全资源量为 z,安防节点的总安全资源量,即 D 付出的代价 $D_{ct}=Z+z$,约束条件为:

$$\begin{cases} x_1 + x_2 + \cdots + x_n = Z + z \\ x_1 \leqslant x_3 \\ x_2 \leqslant x_3 \\ x_8 \geqslant x_i \forall \ \ i = 1,2,\cdots,7 \\ x_i \geqslant (Z+z)/4n \forall \ \ i = 1,2,\cdots,8 \end{cases} \tag{11}$$

不同安全资源投入时的最优安全资源配置及最短路径代价见表 4,随着安全总资源 $Z+z$ 的不断增大,分别求得 $x_1,x_2,\cdots,x_8$ 的对应解,这些解就是增加安全资源 z 投入后的最优安全资源配置,也就是防御者在不同安全资源投入下的最优防御策略,分别对应 $S_{d0},S_{d1},\cdots,S_{d10}$。

表 4　不同安全资源投入时的最优安全资源配置及最短路径代价

Table 4　Optimal security resource allocation and shortest path cost with different security resource inputs

策略	$Z+z$	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	路径代价
S_{d0}	800	157.0	165.0	169.0	25.0	56.0	25.0	32.0	169.0	214.0
S_{d1}	880	183.9	188.0	188.0	27.5	27.5	27.5	49.6	188.0	238.0
S_{d2}	960	200.3	205.0	205.0	30.0	33.2	30.0	50.8	205.2	259.6
S_{d3}	1 040	214.3	222.1	222.1	32.5	52.0	32.5	42.4	222.1	281.1
S_{d4}	1 120	219.8	231.6	237.3	35.0	78.7	35.0	45.3	237.3	300.5
S_{d5}	1 200	247.4	256.3	256.3	37.5	59.0	37.5	49.8	256.3	322.5
S_{d6}	1 280	251.2	264.7	271.2	40.0	90.0	40.0	51.9	271.2	343.4
S_{d7}	1 360	282.0	291.0	291.0	42.5	66.2	42.5	51.0	291.9	367.6

续表 4

策略	$Z+z$	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	路径代价
S_{d8}	1 440	296.9	307.6	307.6	45.0	71.0	45.0	59.3	307.6	389.3
S_{d9}	1 520	313.2	324.6	324.6	47.5	76.0	47.5	62.0	324.6	410.9
S_{d10}	1 600	329.9	341.7	341.7	50.0	78.7	50.0	66.4	341.7	432.5

由表 4 可知:随着投入安全资源量的不断增加,各安防节点分配资源从总体上看也是不断提高的,其中 x_1, x_2, x_3, x_8 的增加速度更快,说明这几个安防节点更重要,需要适当将安全资源向这些节点倾斜。

引入 PT 讨论攻击者和防御者对收益和风险的有限理性决策,进而讨论最优安全资源投入阈值。根据式(6)—式(8),计算攻击者的前景效用值,参数取值为 $g=0.88$、$l=0.9$、$\lambda=2.25$[12]。

表 5　不同安全资源投入下攻击者前景效用价值

Table 5　Attacker prospect utility value under different security resource input

防御策略	S_{d0}	S_{d1}	S_{d2}	S_{d3}	S_{d4}	S_{d5}
效用价值	155.0	126.7	101.0	76.8	54.9	27.6
防御策略	S_{d6}	S_{d7}	S_{d8}	S_{d9}	S_{d10}	—
效用价值	6.0	−22.1	−45.6	−68.9	−94.4	—

由表 5 可知:攻击者的前景效用价值随着防御资源的不断增加而减小,在防御资源投入达到 S_{d6} 时攻击者的前景效用值接近于 0,表示继续增加安全资源投入,攻击者预期收益将小于付出的代价风险,此时攻击者采取攻击的概率极低(几乎不会采取攻击行动)。对于防御者而言,当安全资源增加至 S_{d6} 对应 1 280 时,攻击者采取攻击行动的概率极低,即使继续增加安全资源投入提升安防系统防护能力,也不会再增加获益,因此,防御获益 $D_{ob}=\omega+kA_{ct}$ 中的 k 条路径获益不会继续增加,即没有必要再投入安全资源造成浪费,可直接令之后防御者前景效用值为 0,则对应防御者采取策略 S_{d0}, S_{d1},…, S_{d10} 的前景效用价值见表 6。

表 6　不同安全资源投入下防御者前景效用价值

Table 6　Defender prospect utility value under different security resource input

防御策略	S_{d0}	S_{d1}	S_{d2}	S_{d3}	S_{d4}	S_{d5}
效用价值	414.5	426.4	430.4	430.6	423.2	427.4
防御策略	S_{d6}	S_{d7}	S_{d8}	S_{d9}	S_{d10}	—
效用价值	427.7	0	0	0	0	—

通过模拟试验可以得出,随着防御者安全资源投入的增多,核电厂安防网络中最短代价路径花费也不断上升,攻击者的攻击代价增加,前景效用价值不断下降,如图 5 所示。当防御资源增加到策略 S_{d6} 时,攻击者前景效用价值接近于 0,继续增大安全资源投入,攻击者的前景效用价值降为负数,此时采取攻击行动的概率极低。

对于防御者而言,S_{d0} 到 S_{d3} 段防御者前景价值随着防御资源投入的增加而增大,但增速逐渐变慢,到 S_{d3} 时达到最大。而后继续增加安全资源投入,防御者前景价值反而减小,说明继续投入安全资源量所需要花费的成本高于其带来的安防系统效能增加收益。在 S_{d0} 到 S_{d10} 这 11 组防御策略中,采取 S_{d3} 时,防御者前景价值最大,说明核电厂的最优安全资源投入量为 1 040。

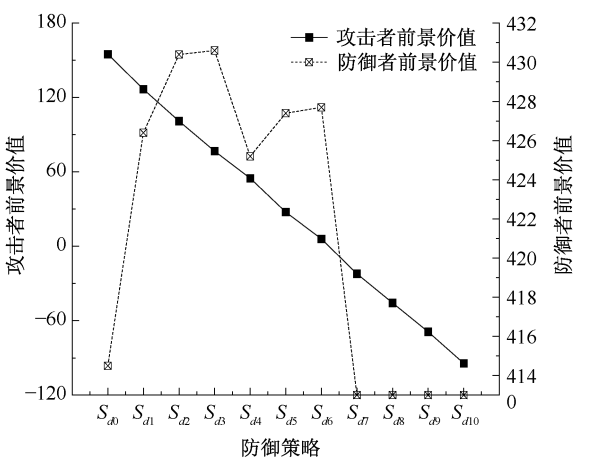

图 5　不同安全资源投入下攻防双方前景价值

Fig. 5　Prospect value of attack and defense under different security resources investment

通过引入 PT,构建核电厂的 PT-ASD-Stackelberg 博弈模型,求解出攻击者和防御者在有限理性博弈下的最优安全资源投入。最后计算得出当防御者采取 S_{d7}—S_{d10} 策略投入安全资源时,攻击者采取攻击行动概率极低。而防御者使用 S_{d3} 防御策略时,前景效用价值最高,此时安全资源投入量 $Z+z$ 为 1 040,之后再增加安全资源投入,防守者的前景效用价值下降,说明投入成本高于获得的回报获益,继续投入安全资源可能造成部分安全资源闲置浪费。因此,该核电厂最优防御策略为安全资源追加至 1 040。

为系统验证所提模型的科学性与有效性,首先,在假设合理性分析层面,文中提出的“完全信息博

弈”“风险可量化”“攻击者理性”是符合关键基础设施安防领域的实际认知和普遍规律的,确保了理论前提的可靠;其次,将上述前提形式化,其核心架构植根于博弈论中的 Stackelberg 安全博弈经典模型,不仅保证了分析框架的数学严谨性,更使模型能够天然地捕捉并模拟攻防双方的战略互动行为;最后,通过核电厂案例进行实证检验,结果表明:与传统经验配置方法相比,模型输出的资源配置方案使整体防护效能提升 25.9%,且能够准确找出最优安全资源投入量。这证明该模型能够将合理的假设与坚实的理论转化为对现实世界决策精准的模拟与优化能力,展现出模型的实用价值。

4　结　论

1) 文中构建的 PT-ASD-Stackelberg 模型,能够将攻击策略映射为入侵路径,避免讨论潜在攻击者未知信息,可将攻防双方的博弈转化为完全信息博弈;发挥关键基础设施安全资源配置的先动优势;且模拟攻击者和防御者在面临未知风险时的有限理性决策,有助于更好地把握最优安全资源投入。

2) 验证分析结果表明:①该模型可以在保持原有安全资源不变时,通过优化安全资源配置,使核电厂整体安防效能提升 25.9%;②在增加安全资源投入时,当投入量达到 1 280,攻击者的前景收益接近于 0,说明攻击者几乎不会采取攻击;防御者的最优安全资源投入总量为 1 040,之后继续增加投入,核电厂总体安防效能增加,但投入成本也随着增加,且成本增加速率高于所带来的防御获益增速。这证明该模型能够有效优化核电厂的安全资源配置,输出最优安全资源配置方案,并且能够准确计算最优安全资源的投入量。

3) 文中在计算最短代价路径时,仅采用路径距离作为边权重的建模方法存在局限性,未能充分考虑路径障碍物分布及巡逻探测概率等安全因素,未来可以进一步采用通过路径所花费的时间来优化改进;仅考虑了充分了解情况攻击者对关键基础设施内部安全信息下的完全信息博弈,未来将进一步研究攻击者掌握部分信息下的不完全信息博弈。

参 考 文 献

[1] 冯治中,王强. 刍议加强我国关键基础设施安全保护工作[J]. 浙江警察学院学报,2022(1):43-50.
FENG Zhizhong, WANG Qiang. On strengthening safeguard work of critical infrastructure in China[J]. Journal of Zhejiang Police College,2022(1):43-50.

[2] 李景山,郭佳芸. 爆炸恐怖袭击情景下地铁车站应急疏散模拟研究[J]. 山西建筑,2024,50(19):127-132.
LI Jingshan,GUO Jiahui. Emergency evacuation simulation study of subway stations under terrorist raid scenarios[J]. Shanxi Architecture,2024,50(19):127-132.

[3] 高寒,杨允锴,赵洪伟,等. 基于数据挖掘的危化品设施遭袭事件特征分析[J]. 中国安全生产科学技术,2025,21(8):214-220.
GAO Han, YANG Yunkai, ZHAO Hongwei, et al. Analysis of attack incident characteristics on hazardous chemical facilities based on data mining[J]. Journal of Safety Science and Technology,2025,21(8):214-220.

[4] ZHANG Laobing, RENIERS G. Applying a Bayesian Stackelberg game for securing a chemical plant[J]. Journal of Loss Prevention in the Process Industries, 2018, 51(1): 72-83.

[5] BIER V M, COX L A, AZAIEZ M N. Game theoretic risk analysis of security threats[M]. Boston: Springer, 2008: 1-12.

[6] MAJOR J A. Advanced techniques for modeling terrorism risk[J]. The Journal of Risk Finance, 2002, 4(1): 15-24.

[7] WOO G. Quantitative terrorism risk assessment[J]. The Journal of Risk Finance, 2002, 4(1): 7-14.

[8] FENG Qilin, CAI Hao, CHEN Zhilong, et al. Using game theory to optimize allocation of defensive resources to protect multiple chemical facilities in a city against terrorist attacks[J]. Journal of Loss Prevention in the Process Industries, 2016, 43: 614-628.

[9] CUI Yan, QUDDUS N, MASHUGA C V. Bayesian network and game theory risk assessment model for third-party damage to oil and gas pipelines[J]. Process Safety and Environmental Protection, 2020, 134: 178-188.

[10] 巩绪福, 蔚承建, 钱震, 等. 基于 Stackelberg 博弈的海事安全问题研究[J]. 计算机科学, 2017, 44(5): 153-159.

GONG Fuxu, WEN Chenjian, QIAN Zhen, et al. Maritime security research based on stackelberg game[J]. Computer Science, 2017, 44(5):153-159.

[11] 王东营,陈小平,刘权,等. 基于改进的云模型-FMEA 的油气管道风险排序[J]. 中国安全科学学报,2024,34(5):61-68.
WANG Dongying, CHEN Xiaoping, LIU Quan, et al. Risk ranking of oil and gas pipeline based on improved cloud model-FMEA [J]. China Safety Science Journal, 2024, 34 (5): 61-68.

[12] 柴瑞瑞,刘德海,陈静锋,等. 考虑防御拓扑特征的暴恐事件演化博弈模型和仿真分析[J]. 运筹与管理, 2017, 26(5): 28-36.
CAI Ruirui, LIU Dehai, CHEN Jingfeng, et al. Evolutionary game model and simulation analysis of terrorist incidents considering defense topological characteristics[J]. Operation Research and Management, 2017, 26(5):28-36.

[13] 林晨. 基于可靠性的系统攻防博弈建模与优化研究[D]. 成都: 西南财经大学, 2021.
LIN Chen. Research on modeling and optimization of system attack-defense game based on reliability[D]. Chengdu: Southwest University of Finance and Economics, 2021.

[14] YAO Yifan, CHEN Wenjing. Security risk assessment of projects in high-risk areas based on attack-defense game model[J]. Scientific Reports, 2023, 13(1), 5-8.

[15] 陈鹤,刘永,赵国海,等. 核安保相关理论研究与实践探索[J]. 中国安全科学学报,2023, 33 (7): 75-81.
CHEN He, LIU Yong, ZHAO Guohai, et al. Theoretical research and practical exploration on nuclear security[J]. China Safety Science Journal, 2023, 33 (7): 75-81.

[16] 郭熹. 基于风险熵模型的安防系统风险与效能评估技术研究[D]. 武汉: 武汉大学, 2011.
GUO Xi. Research on risk and effectiveness evaluation technology on crime prevention system based on risk entropy model[D]. Wuhan: Wuhan University, 2011.

[17] FUDENBERG D, TIROLE J. Game theory[M]. Cambridge: MIT Press, 1991:91-102.

[18] 辛春林,张建文,张艳东. 基于最小最大准则的危险品运输网络优化研究[J]. 中国安全科学学报,2016,26(8):84-89.
XIN Chunlin, ZHANG Jianwen, ZHANG Yandong. Hazardous materials transportation network optimization based on min-max criterion[J]. China Safety Science Journal,2016,26(8):84-89.

[19] 陈宝林. 最优化理论与算法(第 2 版)[M]. 北京: 清华大学出版社, 2005: 80-92.
CHEN Baolin. Optimization theory and algorithms(2nd ed)[M]. Beijing: Tsinghua University Press, 2005: 80-92.

[20] STEWART, M G, MUELLER J. Cost-benefit analysis of airport security: are airports too safe? [J]. Journal of Air Transport Management, 2014, 35(1): 19-28.

[21] TVERSKY A, KAHNEMAN D. Advances in prospect theory: cumulative representation of uncertainty[J]. Journal of Risk and Uncertainty, 1992,5 (4):297-323.

[22] 周慧娟,甕冬阳,李蓓,等. 城市轨道交通运营中断下客流分配方法[J]. 中国安全科学学报,2023, 33(6): 144-151.
ZHOU Huijuan, WENG Dongyang, LI Bei, et al. Passenger flow distribution method under urban rail transit operation[J]. China Safety Science Journal, 2023, 33(6): 144-151.

[23] 任建宇. 面向恐怖袭击的安全防范系统效能评估方法研究[D]. 北京:中国人民公安大学,2022.
REN Jianyu. Research on efficiency evaluation method of security prevention system for terrorist attack[D]. Beijing: People's Public Security University of China, 2022.

作者简介： 黄积成 (2000—),男,江西宜春人,硕士研究生,主要研究方向为安全防范、攻防博弈等。E-mail:1073188638@qq.com。