基于Stackelberg博弈的海事安全问题研究

巩绪福 蔚承建 钱 震 车宝真 沈 航
(南京工业大学计算机科学与技术系 南京 210009)

摘 要 近年来,安全问题在全世界范围内得到了越来越多的重视,如何利用有限的安全资源最大限度地部署防御策略保护重要的设施以及目标是许多安全部门所面临的一项艰巨的挑战。针对海事安全巡逻问题,提出基于Stackelberg博弈的安全模型进行安全资源调度,在安全资源有限的情况下,对现实世界中出现的时空限制和人类行为不完全理性因素进行了综合考虑,放宽了经典安全博弈模型对攻击者是完美理性这一假设。在随机最优反应均衡理论的基础上考虑了攻击者的行为偏好,对非完全理性情况下的最佳策略和完全理性情况下的最佳策略进行对比和分析,实验结果表明,非完全理性下的新模型在现实问题中可以获得更高的收益,可更有效地用于海事安全巡逻问题。

关键词 安全资源分配,人类行为模型,海事安全巡逻,Stackelberg博弈
中图法分类号 TP301 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.05.027

Maritime Security Research Based on Stackelberg Game

GONG Xu-fu WEI Cheng-jian QIAN Zhen CHE Bao-zhen SHEN Hang
(Department of Computer Science and Technology, Nanjing Tech University, Nanjing 210009, China)

Abstract In recent years, the security problem is becoming more and more popular all over the world. How to use the limited security resources to maximize the deployment of defensive strategies to protect critical facilities and goals is a critical challenge that a lot of security department have to face. For maritime safety patrol, security model based on Stackelberg game was proposed to carry out the security resource scheduling. In the case of limited security resources, the constraints of time and space appear in real world and factors that human behavior are not completely rational were comprehensive considered, and the security game's assumption that attack is rational perfectly was relaxed. Based on the theory of quantal response equilibrium, the attacker's behavior preference was considered, then the optimal strategy under the condition of imperfect rationality and the optimal strategy under the condition of complete rationality were compared and analyzed. Experimental results show that the new model under the condition of imperfect rationality can get higher returns in real-world problems, and can be more effectively used in maritime security patrol.

Keywords Security resource allocation, Human behavior model, Maritime security patrols, Stackelberg game

1 引言

近年来,安全问题在全世界范围内受到了越来越多的重视,如何利用有限的资源最大限度地部署防御策略保护重要的设施以及目标是许多安全部门所面临的一项艰巨的挑战。博弈论为研究有限安全资源的最优分配策略提供了一个合适的数学模型,以最大限度地提高资源分配的有效性。Vincent Conitzer等[9]于2006年奠定了Stackelberg博弈在安全部门中的应用基础,使其迅速发展起来。Tambe M等人利用安全博弈论建立洛杉矶机场的ARMOR安全系统[19],由于当地旅客流量很大,警方没有足够的警力来保护整个机场,搜查所有旅客不符合现实,当地警方在考虑了不同检查站位置的收益不同,以及对手信息未知和随机策略容易被针对的情况下,根据收益合理地分配机场各处检查关卡和防爆警犬的巡逻,该系统目前仍在使用中。Z Yin等人在洛杉矶地铁中用TURSTS系统来进行安全检查和打击逃票问题[14]。同样由于资源的限制,需要设计高效的巡逻人员路线,且在实际应用中考虑到了随着逃票行为的减少,削减巡逻路线以及应对突发情况时能够实时出现新的巡逻方案。随着开放型经济的增长,港口在现代物流中占有非常重要的地位。同时,随着港口安全事故以及走私分子的增多,海事安全巡逻问题也越来越被重视。在资源有限的情况下,安全部门面临着如何以有限的资源有效维护港口安全,实现效益最大化的问题。

针对上述问题,提出一个基于人类行为模型下的Stac-

到稿日期:2016-04-26 返修日期:2016-09-10 本文受江苏省自然科学基金(BK20150960)资助。
巩绪福(1990—),男,硕士生,主要研究方向为博弈论安全,E-mail:691954891@qq.com;蔚承建(1957—),男,博士,教授,硕士生导师,主要研究方向为云计算、演化计算、人工智能多代理博弈论;钱 震(1990—),男,硕士生,主要研究方向为博弈论安全、多代理;车宝真(1992—),男,硕士生,主要研究方向为移动互联网;沈 航(1984—),男,博士,讲师,主要研究方向为云计算、移动互联网、无线多媒体通信协议等。

kelberg博弈安全应用，可用于海事辖区进行巡逻，保护港口，打击犯罪行为。该应用考虑了现实世界中的地理限制和时间限制，生成符合现实要求的巡逻路线，并且放宽了安全博弈模型对攻击者是完美理性这一假设，在随机最佳反应均衡理论的基础上对攻击者行为策略进行建模，对非完全理性情况下的最佳策略和完全理性情况(DOBSS 算法)下的最佳策略进行对比与分析；然后进一步考虑了人类行为的偏好性，改进模型使之更加符合现实情况，并与原模型进行对比分析，模拟实验结果表明新模型具有更高的收益，更加具有现实意义。

2 数学建模

为了将海事区域巡逻问题转变为一个 Stackelberg 安全博弈问题。需要定义：1)攻击者的策略集合；2)防守者的策略集合；3)收益矩阵；4)对手行为模型。这些策略和收益主要和港口(目标)有关，例如某些港口被列为潜在的攻击对象。因此在文中定义攻击者对单一目标发动攻击的行为作为攻击者的纯策略。防守者策略的定义和攻击者的策略定义不同，安全部门每天有一个固定的持续时间在设定的巡逻路线上巡逻。本节将图 1 中的每个顶点作为一个目标，且每个目标与其他目标之间存在巡逻路径，目标与目标之间的路径会形成一个无向图。文中将这个图形产生的可行巡逻路线作为防守者的策略集合。

2.1 数学模型

为了详细描述 Stackelberg 博弈海事巡逻模型的实际应用，本文在百度地图的海岸线上选取 10 个港口建立模型，如图 1 所示。

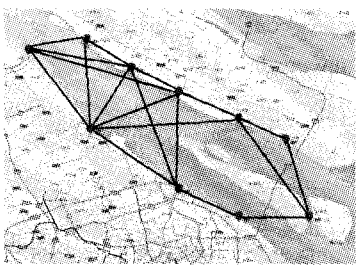

图 1 港口地图

根据百度地图上的港口，从中提取数学模型，从而可以得到如图 2 所示的图形 $G(T,Er)$。

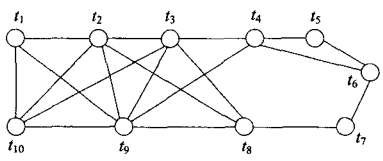

图 2 港口数学模型

在本文的海事安全巡逻模型中，安全部门作为防守者需要对图 2 中的港口进行保护，因为攻击者可能会在这些目标进行攻击和犯罪活动。在模型中安全部门从某个目标出发，然后到达每一个目标 $t\in T$ 时，会在防守活动集合 $\{m_1,m_2,\cdots,m_k\}$ 中选择一个防守活动对目标 t 进行部署，在目标上防守活动所花费的时间越长，防守策略的有效性越高[1]。这会影响最后的巡逻时间表的结果：在一次巡逻路线中巡逻的地点少但花费在巡逻活动上的时间长，在另外一次巡逻路线中巡逻的地点多但是花费在巡逻活动上的时间少。安全部门最终在最大巡逻时间内回到出发地点。而港口图 $G(T,E_r)$ 中的 T 表示需要巡逻的目标，E_r 表示地点之间的路径。为每一条边 $e\in E_r$ 会分配一个时间 $\tau(e)$，表示防守者经过这条边需要的时间。参数符号意义如表 1 所列。

表 1 符号含义表

符号	参数意义
$G(T,E_r)$	巡逻图形
T	保护目标的集合
t_b	基本出发地点
$E(m)$	防守活动的有效性
E_r	边的集合
$\tau(e(t_i,t_j))$	经过边 e 的时间
$\tau(m)$	实施防守活动 m 的时间
Γ	每条巡逻路线的最大巡逻时间
J	防守者巡逻路线集合

2.2 防守者策略和收益计算

攻击者的攻击策略集合是所有可攻击的目标 T。防守者策略集合是在图型 G 上生成的所有可行巡逻路线。对于一条巡逻路线 A_j，将其定义为一个有序排列的三元组 $A_j=[A_j^1,\cdots,A_j^n,\cdots]$[44]。第 n 个三元组 $A_j^n=(t,m,y)$ 表示防守者在目标 t 上实施和完成行动 m 的时间为 y。巡逻路线开始和结束都在同一个基本起始地点 $t_b\in T$。不同的防守活动需要不同的时间，且对目标的保护效果也不同。文中用 $E(m)$ 表示防守活动 m 的有效性，其范围为 0%～100%，表示在目标 t 上实施防守活动 m 后可以阻止目标被攻击的可能性。

以一个具有 5 个目标的简化场景(见图 3)为例进行防守者策略分析和收益计算。用 τ_{ij} 来表示从目标 t_i 到目标 t_j 所用的时间，如 t_1 到 t_2 的时间 τ_{12} 为 2。假设基本出发地点 $t_b=t_1$，防守者现在有两个可以在目标上实施的防守活动 $\{m_1,m_2\}$，其中 m_2 表示保护目标进行深度搜索和巡逻，m_1 表示只是通过观察来保护目标。表 2 列出了防守活动所需的时间和有效性。

表 2 防守活动类型

防守活动	m_2	m_1
有效性 $E(m_i)$	0.6	0.1
时间	2	0

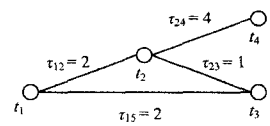

图 3 路线图举例

表 3 列出了图 3 中防守者可行的部分巡逻策略，行表示防守者的策略，列表示攻击者的策略。一个巡逻时间表由巡逻地点的顺序以及在地点中实施的防守活动构成，需要将巡逻时间表中巡逻的地点排序，在巡逻表中的第一个巡逻地点表示防守方出发的基本巡逻地点。在此例中，目标 t_1 是基本巡逻地点，所有巡逻路线的开始和结束都在目标 t_1。例如表 3 中的 A_1 路线，防守方首先去巡逻地点 t_1 实施防守活动 m_1，

接着访问巡逻地点 t_2 实施防守活动 m_1，最终回到区域地点 t_1 实施防守活动 m_1。后面 3 列数值对分别表示防守者和攻击者在目标 1,2,3 上得到的收益。

表 3　简单的巡逻博弈路线示例

巡逻路线	目标 1	目标 2	目标 3
$A_1=\{(t_1:m_1),(t_2:m_1),(t_1:m_1)\}$	80,−80	40,−40	−30,30
$A_2=(t_1:m_2),(t_2:m_1),(t_1:m_1)$	160,−160	40,−40	−30,30
$A_3=(t_1:m_1),(t_2:m_1),(t_1:m_2)$	160,−160	40,−40	−30,30
$A_4=(t_1:m_1),(t_3:m_1),(t_2:m_1),(t_1:m_1)$	80,−80	40,−40	20,−20
$A_5=(t_1:m_1),(t_2:m_1),(t_3:m_1),(t_1:m_1)$	80,−80	40,−40	20,−20

在海事安全巡逻博弈问题中，防守者策略的数量，即巡逻路线的数量的增长对于计算是一个挑战。为了解决这一问题，在文中的防守者巡逻策略中，使用一种紧凑策略的方法，可以合并等价的巡逻路线和删掉可被替代的巡逻路线。

对比巡逻路线 A_2 和 A_3，A_4 和 A_5，其收益值是相同的。从等价方面而言，不同巡逻路线经过相同区域的不同排列组合提供相同的收益结果；此外，如果一个区域在一次巡逻中被多次访问实施不同的防守活动，只考虑防守者提供最高收益的活动。所以，如果表 3 中不同巡逻路线在所有目标上的防守活动是相同的，即使它们的访问顺序不同，这些巡逻路线也是等价的，这些等价的巡逻路线可以合并成一个紧凑的防守者策略，如 A_2 和 A_3 可合为一条紧凑路线 $\{(t_1:m_2),(t_2:m_1)\}$。

对比巡逻路线 A_1 和 A_2，因为防守活动 m_2 给予防守方比 m_1 更高的回报，A_1 可以从防守者的策略集合中删除，因为 A_2 的巡逻路线覆盖 A_1 相同的区域且可以获得更高的收益。

为了生成紧凑的巡逻策略，定义在港口路线图 $G(T,E_r)$ 中，给定基本出发地点 t_b（巡逻开始和结束的地点），在最大可持续的巡逻时间 Γ 内，紧凑策略路线生产可以转变为求解如下线性规划问题，如式(1)所示：

$$\begin{aligned}&\sum_{t\in T}\vec{e}_t\max\{f(m)\cdot E(m)\}\\&\sum_{e\in out(t)}f(e)=\sum_{e\in in(t)}f(e),\forall t\\&f(e)\in\{0,1,2\},\forall e\\&\sum_{e\in out(u)}f(e)\cdot\tau(e)+\sum_{t\in T}\sum_{i=1}^{k}f(m_i)\cdot\tau(m_i)\leqslant\Gamma\end{aligned}\tag{1}$$

其中，$\vec{e}_t$ 表示$[0\cdots t\cdots 0]^{\mathrm{T}}$，第 t 行为 1，e 表示路线图中目标之间的路径；$f(e)$表示经过边 e 的次数；$f(m)$表示活动 m 执行的次数；$E(m)$表示活动 m 的有效性；$in(t)$表示到达目标 t 的边；$out(t)$表示从目标 t 出发的边，每条边限制最多经过两次，具有实际意义并加快了算法速度；$\tau(e)$表示经过这条边所需要的时间；$\tau(m_i)$表示执行活动的时间。所有经过的边所需要的时间加上所有执行活动所需要的时间的总和不能超过最大巡逻时间。

另外，从获得的收益方面而言，如果攻击方选择攻击目标 $t_i\in T$ 但攻击失败，则防守者得到收益 R_i^d，而攻击者会得到一个惩罚 P_i^a，否则防守者会得到一个惩罚 P_i^d，攻击者则得到一个收益 R_i^a。假如防守方选择巡逻路线 j 而攻击者选择攻击目标 t_i，用 G_{ij}^d 来表示防守者此时的收益，可以由防守者在目标 i 上的收益/惩罚以及 E_{ij} 的线性组合来表示，其中 E_{ij} 表示巡逻路线 j 保护目标 i 的有效性，这和布置的防守活动 m_i 有关，其表示在目标 t 上实施防守活动 m 后可以保护目标的可能性。如式(2)所示，如果 E_{ij} 的值为 0，则表示目标 t_i 并不在巡逻路线 j 中。

$$G_{ij}^d=E_{ij}R_i^d+(1-E_{ij})P_i^d\tag{2}$$

2.3　攻击者行为模型

随机最优反应均衡在行为博弈论中是一个重要的理论[2-4]。最近的一项研究表明随机反应模型是一个有效预测人类行为的模型[5]。随机最优反应模型假设人们并不严格实现最大化自己的收益，在有干扰的决策过程中并非是完全理性地进行行为决策。本文在海事巡逻 Stackelberg 博弈中，假设攻击者是有限理性的，在随机最优反应模型的基础上考虑攻击者行为偏好对攻击者行为建模，本文后面的实验结果表明，基于攻击者行为建模的 Stackelberg 安全模型比完美理性模型表现得更优异。

在本文 Stackelberg 海事安全博弈模型中，对攻击者行为进行建模。在行为博弈论中的随机最优反应均衡[6]下，攻击者会根据攻击每个目标取得的收益值选取攻击目标。攻击者攻击某一个目标获得的收益越高，选取该目标的概率就越高，随机最优反应的公式如式(3)所示：

$$q_i(x)=\frac{e^{\lambda U_i^a(x_i)}}{\sum_{k=1}^{T}e^{\lambda U_k^a(x_k)}}\tag{3}$$

文中用该公式对攻击者的非理性因素进行建模。但是随机最优反应均衡没有考虑攻击者对防守覆盖率最低的目标有倾向攻击的行为偏好。本文在考虑该倾向偏好的情况下对攻击者进行建模。防守者在考虑攻击者行为的条件下最大化预期收益。在本文的应用中，计算防守者的最大收益可以看作一个非线性非凸函数优化问题[7]。

假设攻击者为了追求攻击成功率，对防守资源覆盖最少的目标有一定的攻击倾向，其选择攻击目标 t_i 的概率如式(4)所示：

$$q_i(x)=\frac{e^{\lambda_U U_i^a(x_i)}e^{\lambda_F F_i(x)}}{\sum_{k=1}^{T}e^{\lambda_U U_k^a(x_k)}e^{\lambda_F F_k(x)}}\tag{4}$$

其中，$F_i(x)\in\{0,1\}$表示目标 t_i 是否是防守资源覆盖率最低的目标，如式(5)所示：

$$F_i(x)=\begin{cases}1, & x_i\leqslant x_{i'},\forall T_{i'}\in T\\0, & \text{其他}\end{cases}\tag{5}$$

在式(4)中，x_i 是防守者对目标 t_i 的覆盖率，参数 $\lambda_U\in[0,\infty]$代表了行为的理性程度。当 $\lambda_U=0$ 时，表示攻击者没有考虑收益，此时 q_i 表示攻击者攻击目标的概率是随机分布。随着 λ_U 的增加，q_i 会越来越接近一个纯策略，选择使攻击者收益最大的目标。当 $\lambda_U=\infty$时，表示攻击者是完美理性的，选择收益最大的目标进行攻击。在此模型中，考虑攻击者对覆盖率最低的目标有攻击偏好，可以通过增加 λ_F 的值来提升对目标 t_i 的保护覆盖率。而 $e^{\lambda_F F_i(x)}$ 表示攻击对覆盖率最

小的目标的偏向。当目标 t_i 不是防守资源覆盖率最低的，$F_i(x)$ 的值为 0，此时 $e^{\lambda_F F_i(x)}$ 的值为 1，表示攻击者不会对这个目标 t_i 有额外的攻击偏好。如果目标 t_i 是防守资源覆盖率最低的，那么攻击者选择目标 t_i 的可能性会增加。参数 $\lambda_F>0$ 表示攻击者对防守资源覆盖率最低的目标的偏好程度；当 $\lambda_F=0$ 时，攻击者对最小覆盖率目标没有攻击偏好。λ_F 的值越大，攻击者对覆盖率最低的目标的偏好越强。

攻击者根据概率 q_i 选择攻击目标，$U_i^a(x_i)$ 表示攻击者选择目标 $t_i\in T$ 的收益。向量 x 表示防守者对保护目标集合的覆盖率，使用上节描述的符号，同时定义 R_i^a 和 P_i^a 以及 R_i^d 和 P_i^d 分别表示攻击者和防守者选择目标 t_i 的收益和惩罚。

攻击者攻击目标 t_i 的收益为：

$$U_i^a(x_i)=x_iP_i^a+(1-x_i)R_i^a=R_i^a+x_i(P_i^a-R_i^a) \tag{6}$$

同样地，防守者防守目标 t_i 的收益为：

$$U_i^d(x_i)=x_iR_i^d+(1-x_i)P_i^d=P_i^d+x_i(R_i^d-P_i^d) \tag{7}$$

防守者在攻击者攻击目标 t_i 的可能性为 q_i 的前提下，最大化预期收益。给定需要保护的目标集合 T，覆盖率向量 x，基于攻击者随机最佳反应模型下的防守者的预期收益为：

$$U^d(x)=\sum_{i=1}^{T}q_i(x)U_i^d(x_i)=\sum_{i=1}^{T}q_i(x_iR_i^d+(1-x_i)P_i^d) \tag{8}$$

因此，给定防守者紧凑策略集合 J 以及有效性矩阵 E_{ij}，及防守方资源数 M，K 表示一个较大的常数，以使 x_i 最小时，F_i 的值为 1。计算防守者的最优策略可以定义为如下非线性非凸优化问题 P_1：

$$P_1\begin{cases}\max\limits_{x,a}\dfrac{\sum\limits_{i=1}^{T}e^{\lambda_U(R_i^a-(R_i^a-P_i^a)x_i)}e^{\lambda_F F_i}((R_i^d-P_i^d)x_i+P_i^d)}{\sum\limits_{i=1}^{T}e^{\lambda_U}e^{\lambda_F F_i}((R_i^a-(R_i^a-P_i^a)x_i))}\\ \text{s.t. } x_i=\sum\limits_{j=1}^{J}a_jE_{ij},\ \forall i\\ x_i-(1-F_i)K\leqslant x_{\min}\leqslant x_i,\ \forall t_i\in T\\ \sum\limits_{i=1}^{T}F_i=1,F_i\in\{0,1\},\ \forall t_i\in T\\ \sum\limits_{j=1}^{J}a_j=M\\ 0\leqslant a_j\leqslant 1,\ \forall j\end{cases} \tag{9}$$

上述符号的含义如表 4 所列。

表 4 符号含义表

符号	参数意义
T	需要保护目标的集合
J	防守方巡逻策略的集合
E_{ij}	巡逻策略 j 对保护目标 t_i 的有效性
a_j	选择紧凑策略 j 的概率
x_i	保护目标 t_i 的覆盖率
M	防守资源数
R_i^d	成功防守 t_i 时防守方的收益
P_i^d	失败防守 t_i 时防守方的惩罚
R_i^a	成功进攻 t_i 时攻击方的收益
P_i^a	失败进攻 t_i 时攻击方的惩罚
λ_U	攻击者理性程度值
λ_F	攻击者行为偏好程度值

3 海事安全巡逻模型求解

文中需要通过计算问题 P_1 来求解防守方的最优策略。为了克服式(9)中函数的非线性非凸函数计算带来的问题，使用二分法和线性逼近的方法将其转化成为一个混合整数线性规划问题来进行求解。

3.1 二分搜索法

为了简化式(9)，本节定义式(10)－式(12)来简化目标函数，如下所示：

$$\omega_i=e^{\lambda_u R_i^a}>0 \tag{10}$$

$$\beta_i=\lambda_u(R_i^a-P_i^a)>0 \tag{11}$$

$$\gamma_i=R_i^d-P_i^d \tag{12}$$

式(9)中整型变量 F_i 由式(5)中的函数 $F(x_i)$ 决定。限制条件为式(9)的第 3 行，其中 K 表示一个比较大的常数，$x_{\min}$ 表示防守者对目标覆盖率 x_i 中的最小值，如果防守者对目标 t_i 的防守覆盖率最低，则 F_i 的值为 1，其他目标 F_i 的值为 0。首先采用迭代的方法消除该问题中的整形变量 F_i，对于迭代中的每一个子问题，将所有目标集合中的其中一个目标假定为防守覆盖率最低的目标，设置 F_i 为 1，然后求解这个子问题，得到子问题中防守者的最大预期收益和相应的混合策略，所有求解的子问题的解集合中防守者最大预期收益的最大值和相应的混合策略即为该问题的最优解。

用式(10)－式(12)中所定义的符号可以将求解的子问题 P 简化表示为 $N(x)/D(x)$，其中 $N(x)$ 由式(13)表示，$D(x)$ 由式(14)表示：

$$N(x)=\sum_{i=1}^{T}\omega_i\gamma_ix_ie^{-\beta_ix_i}+\sum_{i=1}^{T}\omega_iP_i^de^{-\beta_ix_i} \tag{13}$$

$$D(x)=\sum_{i=1}^{T}\omega_ie^{-\beta_ix_i}>0 \tag{14}$$

假设用 F 来表示需要解决的线性规划问题 P 的可行区域，P^* 表示在 SSE 均衡的最优策略，那么可以将求解最优策略的 P 问题转化为：

$$P^*=\max_{(x,a)\in F}\frac{N(x)}{D(x)} \tag{15}$$

二分搜索法可以有效地求解这种分式目标函数问题[8]。二分搜索法的核心思想是通过解决不含分式目标函数的相关优化问题来迭代计算最优值 P^* 的上下界。给定一个上下界之间的有效函数值 r，可以定义如下优化问题：

$$\Omega:\delta_r^*=\min_{x\in F}rD(x)-N(x) \tag{16}$$

可以证明 $r\leqslant P^*$ 与 $\delta_r^*\leqslant 0$ 是等价的。

假设 $N(x)$ 和 $D(x)$ 均为连续函数，且其定义域为闭合区域集合 F，$D(x)>0,\forall x\in F$。如果 $P^*=\max\limits_{(x,a)\in F}\dfrac{N(x)}{D(x)}$，那么 $r\leqslant P^*$ 是 $\delta_r^*\leqslant 0$ 的充分必要条件。

(充分性)因为 P^* 是在封闭区域集合中连续函数的最优函数值，那么必定存在一个最优解 x^*，即 $P^*=\dfrac{N(x^*)}{D(x^*)}\geqslant r$，那么 $rD(x^*)-N(x^*)\leqslant 0$。

(必要性)因为存在 x 使得 $\delta_r{}^*=rD(\underline{x})-N(\underline{x})\leqslant 0$，转化为 $r\leqslant N(\underline{x})/D(\underline{x})\leqslant P^*$，推出 $r\leqslant P$。

所以，为了解决上述优化问题，可以通过判断 $\delta_r^*\leqslant 0$ 来比较 r 与全局最大值的大小关系。表 5 描述了对于 P_1 问题的优化算法，其中，ε 表示误差范围，P_M 表示收益矩阵，M 表示安全资源的数量。

表 5　算法流程

```
Algorithm process
1. Input:ε,P_M,M
2. (U_0,L_0)←建立上下界(P_M,M)
3. (U,L) ← (U_0,L_0)
4.   P* ←−∞
5.   while 1≤i≤T do
6.              F_i=1
7.         while U−L ≥ ε do
8.              r←(U+L)/2
9.       计算 Ω,使得 x^r,δ_r^* 是最优解和对应的函数值
10.            if δ_r^* ≤ 0 then
11.                   L←r
12.                else
13.                   U←r
14.            end if
15.        end while
16.    if P* < δ_r^*   then
17.          P* ←δ_r^* ,x* ←x^r
18. end while
19. return P* ,x*
```

表 5 在第 2 行中初始化了所求目标函数的上界(U_0)和下界(L_0)。接着在每一次迭代中，r 是上界 U 和下界 L 之间的中位数。通过计算公式 Ω 在第 6 行判断 $r \leqslant P^*$。如果二分搜索法的上下界需要被替换，那么要返回一个可用的策略 x^r，一直迭代到上下界满足 $U-L<\varepsilon$，求出最优值。

初始下界：文中将防守者随机分布时的收益值作为下界。因为随机均匀分布没有考虑优化问题。

初始上界：因为 $P_i^d \leqslant U_i^d \leqslant R_i^d$，所以可以得到 $U_i^d \leqslant \max_{i=1}^{T} R_i^d$，而防御者的收益为 $\sum_{i=1}^{T} q_i U_i^d$，所以可以将最大的 R_i^d 代入作为 U_i^d 的上界。

3.2　分段线性逼近法

在 3.1 节的二分法中，需要计算问题 Ω，这是一个非线性规划问题，文中用逼近的方法来解决这个问题，首先将 Ω 写成式(17)的形式：

$$\sum_{i=1}^{T}\omega_i(r-P_i^d)e^{-\beta_i x_i}e-\sum_{i=1}^{T}\omega_i\gamma_i x_i e^{-\beta_i x_i} \tag{17}$$

为了计算式(17)，定义两个非线性函数 $g_i^{(1)}(x_i)=e^{-\beta_i x_i}$ 和 $g_i^{(2)}(x_i)=x_i e^{-\beta_i x_i}$，其中 $x_i \in [0,1]$，$1 \leqslant i \leqslant T$。

将 x_i 的范围[0,1]均匀地分成 K 份，然后引入一组新的变量 $x_{ik}(k=1,\cdots,K)$来表示 x_i 被分成 K 个部分中的每一段。因此得到 $x_{ik} \in [0,\frac{1}{K}]$，$\forall k=1,\cdots,K$ 和 $x_i=\sum_{k=1}^{K} x_{ik}$。为了确保$\{x_{ik}\}$是 x_i 的有效部分，所有 x_{ik} 必须满足：当 $x_{ik}>0$ 时，$x_{ik'}=\frac{1}{k}$，$\forall k'<k$。以图 4 和图 5 为例演示上述两个函数的分段逼近方法。

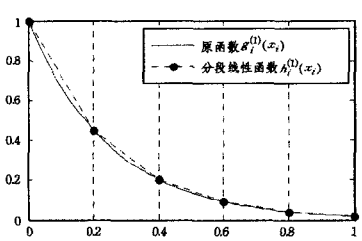

图 4　函数 $g_i^{(1)}(x_i)$的分段逼近

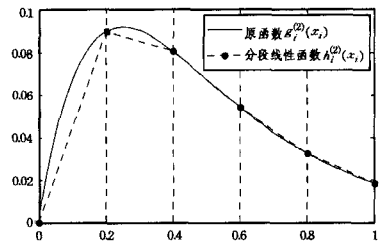

图 5　函数 $g_i^{(2)}(x_i)$的分段逼近

本文用$\{x_{ik}\}$将两个非线性函数 $g_i^{(1)}$ 和 $g_i^{(2)}$ 变为分段线性函数。使$\{(\frac{k}{K},g_i^{(1)}(\frac{k}{K})),k=0,\cdots,K\}$ 为函数 $g_i^{(1)}(x_i)$线性分段的 $K+1$ 个分割点，并且用 $y_{ik}(k=1,\cdots,K)$来表示每个线性部分的斜率。$g_i^{(1)}(x_i)$的线性分段函数可以表示为 $h_i^{(1)}(x_i)$：

$$h_i^{(1)}(x_i)=g_i^1(0)+\sum_{k=1}^{K}y_{ik}x_{ik}=1+\sum_{k=1}^{K}y_{ik}x_{ik} \tag{18}$$

函数 $g_i^{(2)}(x_i)$的线性分段函数可以表示为 $h_i^{(2)}(x_i)$：

$$h_i^{(2)}(x_i)=g_i^2(0)+\sum_{k=1}^{K}u_{ik}x_{ik}=\sum_{k=1}^{K}u_{ik}x_{ik} \tag{19}$$

其中，$u_{ik}(k=1,\cdots,K)$表示 $g_i^{(2)}(x_i)$每个线性部分的斜率，而$(\frac{k}{K},g_i^{(2)}(\frac{k}{K}))(k=0,\cdots,L)$为函数 $g_i^{(2)}(x_i)$线性分段的$L+1$个分割点。为了用分段函数的形式表示目标函数 Ω，将每个变量 x_i 分为 L 个部分，表示为 $x_{ik}(k=1,\cdots,L)$。根据式(18)和式(19)，可以把 3.1 节中所要计算的 Ω 转化为如下形式：

$$\begin{aligned}
&\min_{x,z,a}\sum_{i=1}^{T}\theta_i(r-P_i^d)(1+\sum_{k=1}^{K}y_{ik}x_{ik})-\sum_{i=1}^{T}\theta_i a_i\sum_{k=1}^{K}u_{ik}x_{ik}\\
&\text{s.t. } 0\leqslant x_{ik}\leqslant\frac{1}{K},\forall i,k=1,\cdots,K\\
&\quad z_{ik}\frac{1}{K}\leqslant x_{ik},\forall i,k=1,\cdots,K-1\\
&\quad x_{i(k+1)}\leqslant z_{ik},\forall i,k=1,\cdots,K-1\\
&\quad z_{ik}\in\{0,1\},\forall i,k=1,\cdots,K-1\\
&\quad \sum_{k=1}^{K}x_{ik}=\sum_{E_j\in E}a_j A_{ij},\forall i\\
&\quad \sum_{E_j\in E}a_j=M,0\leqslant a_j\leqslant 1,\forall j
\end{aligned} \tag{20}$$

这样就将原问题变为了一个 MIQP 的问题，这是可以在 IBM 的 CPLEX 框架下实施的最终形式，将会在第 4 节进行讨论。

4　实验设计与分析

本节通过一系列实验对所描述的模型进行性能分析和评价。首先介绍 Stackleberg 海事巡逻模型的参数输入，然后对实验结果进行分析和讨论。

4.1　实验设计

使用 IBM 的 ILOG CPLEX 软件来计算领导者的最优策略，IBM 的 ILOG CPLEX 是一种非常强大的线性规划处理工具，支持各种编程语言，实验使用 JAVA 语言编程实现，将模型中的港口(目标)作为对象，属性包括执行防守活动 m_i 所需的时间、可以到达的港口集合以及对应的航行时间、防守者防守该目标的收益(成功或失败)、攻击者攻击该目标的收益(成功或失败)。SecurityGame(安全博弈)类包含目标数、资源

数、目标集合。Constraint 类把式(20)中的限制条件转换成 CPLEX 可以处理的对象。整个实验结果都运行在一个配置 Inter 酷睿 i5 1.7GHz 处理器、4GB 内存的机器上。

在 Stackelberg 海事安全巡逻模型中,需要 5 种类型的基本输入:1)可用的安全资源的数量(巡逻船);2)在每个巡逻区域可执行的不同防守活动以及防守活动的能力(有效性);3)需要被保护目标的集合;4)每个目标防守成功或者失败的收益回报值;5)不同类型的调度约束(例如时间限制以及地理限制)。

除了上述的基本输入信息,还需要决定人类行为中模型中的理性值和行为偏好值。显然,均匀随机和完美理性都是不符合实际情况的。根据数据的实验结果分析发现,当理性参数 $\lambda_U>4$ 时,攻击者开始趋于完美理性,攻击者攻击目标的可能性开始集中在一个目标上。在给定防守者和攻击者对目标的收益值的情况下经过大量实验,选取 $\lambda_U=0.96$ 作为攻击者的理性值,$\lambda_F=0.8$ 作为攻击者对覆盖率最低的目标的攻击偏好。

图 6 在图 2 的港口数学模型上进一步设定了路程时间,在本次模拟实验中定义:1)可用的巡逻船数量为 2;2)执行防守活动 m_3 表示对保护目标进行深度搜查防守,其有效性是 0.6,所需时间为 60min,执行防守活动 m_2 表示对部分可疑地点人物进行搜查,其有效性是 0.3,所需时间为 30min,执行防守活动 m_1 表示仅路过观察,其有效性为 0.1,所需时间为 10min;3)需要保护的目标的集合为 $T=[t_1,t_2,t_3,t_4,t_5,t_6,t_7,t_8,t_9,t_{10}]$;4)目标成功的收益范围为[0,20],目标失败惩罚的收益范围为[−20,−0],由随机函数产生,具体数值如表 8 所列;5)本文设定的最大巡逻时间为 6h,每个目标之间相连路线的限制如图 6 所示。表 6 和表 7 列出了本文需要在程序中建立图形所要输入的文件。

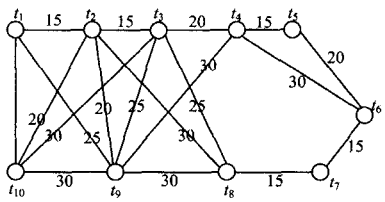

图 6 港口数学模型

表 6 模形输入文件 1

	模型输入文件
(1)	Base(1)
(2)	Vertex(1),(10),(30),(60)
(3)	Vertex(2),(10),(30),(60)
(4)	Vertex(3),(10),(30),(60)
(5)	Vertex(4),(10),(30),(60)
(6)	Vertex(5),(10),(30),(60)
(7)	Vertex(6),(10),(30),(60)
(8)	Vertex(7),(10),(30),(60)
(9)	Vertex(8),(10),(30),(60)
(10)	Vertex(9),(10),(30),(60)
(11)	Vertex(10),(10),(30),(60)

在模型输入文件 1 中,第 1 行代表在模型中设定的基本区域是 t_1,第 2 行—第 11 行是图形中的顶点 t_1-t_{10},即需要保护的目标。后面 3 列分别表示的是防守活动 m_1,m_2 和 m_3 需要的时间。而在表 7 中,每行表示的是图形的地理限制。如从第 1 行—第 3 行表示地点 1 到区域 2,9,10 之间有可行的航道,其中第 3 列表示区域之间航行需要的时间,第 4 行—第 8 行表示从地点 2 到区域 1,3,8,9,10 的可行航道和时间,依次类推,表 7 只列出了部分边的输入。

表 7 模型输入文件 2

	模型输入文件
(1)	Edge(1),(2),(15)
(2)	Edge(1),(9),(25)
(3)	Edge(1),(10),(15)
(4)	Edge(2),(1),(15)
(5)	Edge(2),(10),(20)
(6)	Edge(2),(9),(20)
(7)	Edge(2),(8),(30)
(8)	Edge(2),(3),(15)

表 8 收益表格

保护目标	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
防守者收益	17	9	10	12	18	6	5	3	8	16
防守者惩罚	−17	−9	−10	−12	−18	−6	−5	−3	−8	−16
攻击者收益	17	9	10	12	18	6	5	3	8	16
攻击者惩罚	−17	−9	−10	−12	−18	−6	−5	−3	−8	−16

根据以上输入,程序输出的部分结果如表 9 所列,生产的巡逻路线是在防守者收益最大时,按其混合策略中对每条巡逻路线的概率分布进行选取,每行代表一天的巡逻路线,其中包含了巡逻船从基本出发区域开始巡逻的起始时间、每次巡逻的顺序地点和实施的防守活动。

表 9 巡逻路线

日期	巡逻路线	出发时间
第一天	$(1:m_2)(9:m_2)(4:m_1)(6:m_2)$ $(7:m_2)(8:m_1)(2:m_2)(1:m_1)$	17 点
第二天	$(1:m_2)(10:m_3)(9:m_3)$ $(8:m_1)(2:m_3)(1:m_1)$	9 点
第三天	$(1:m_2)(2:m_1)(8:m_1)(7:m_2)(6:m_1)$ $(5:m_1)(4:m_1)(9:m_2)(1:m_2)$	14 点
第四天	$(1:m_1)(2:m_2)(3:m_1)(4:m_3)$ $(9:m_2)(2:m_1)(10:m_2)(1:m_1)$	6 点
第五天	$(1:m_3)(9:m_1)(3:m_3)$ $(8:m_2)(2:m_3)(1:m_1)$	11 点

4.2 实验结果分析

在本文的第一个实验中,只考虑攻击者的理性因素 λ_U,不考虑攻击者的行为偏好 λ_F 的情况($\lambda_F=0$),此时使用改进后的模型方法与 DOBSS 算法进行比较,DOBSS 算法是一种认为攻击者是完美理性($\lambda_U=\infty$)的算法,而改进后的模型考虑了攻击者的不理性因素,使用 λ_U 来预测攻击者的不理性程度,通常情况下对手 λ_U 的值不是一个固定的数,而是一个范围。根据大量实验数据,本文使用 $\lambda_U=0.96$ 时的模型与 DOBSS 算法进行比较,实验结果如图 7 所示。

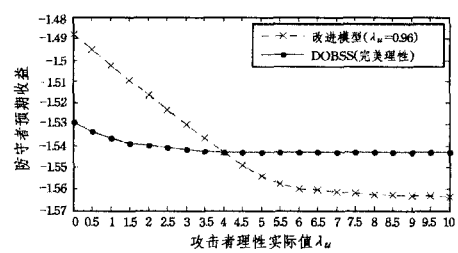

图 7 改进后的模型与 DOBSS 算法的防守者收益对比

图 7 中的纵坐标表示防守者的预期收益,横坐标表示实

际面对攻击者的理性值。从图 7 可以发现，在面对攻击者的理性值为 0～4 时，改进后的模型比传统 DOBSS 算法的收益高。而当攻击者的理性值大于 4 时，攻击者选择攻击目标的行为趋近完美理性，此时 DOBSS 算法的收益较高。所以在面对 $\lambda_U \in [0,4]$ 之间的非完美理性的对手，新的模型更符合实际且可以获得更高的收益。

在本文的第二个实验中，考虑人类对手对防守资源覆盖最少的目标有攻击偏好的情况。选取 $\lambda_U=0.96, \lambda_F=0.8$ 的新模型与随机最优均衡理论模型进行比较，随机最优均衡理论只考虑了攻击者的不理性因素，没有考虑攻击者的行为偏好（$\lambda_U=0.96, \lambda_F=0$）。实验结果如图 8 所示，如果在攻击者行为没有偏好的情况下，随机最优均衡理论的模型值较大，不过随着 λ_F 的增加，防守者的预期收益开始快速降低。当 λ_F 大于 0.5 之后，改进后的模型可以取得较大的收益值，所以在现实情况下，当攻击者考虑成功率，倾向于攻击防守资源覆盖率最小的目标时，本文改进后的模型会比原有模型获得更好的收益，更适用于实际情况。

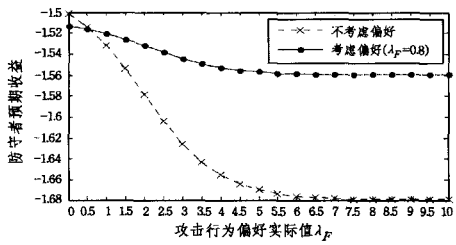

图 8　改进后的模型与随机最优均衡策略的防守者收益对比

结束语　本文针对海事安全巡逻中的辖区的安全资源分配问题，提出了一种基于 Stackelberg 博弈的安全资源调度模型，对江苏海岸线附近的港口进行数学建模并给出实际运行的优化巡逻路线和最佳期望效益。该模型放宽了经典安全博弈模型中攻击者为完美理性的假设，并进一步在随机最佳反应均衡理论的基础上对攻击者的攻击行为偏好进行建模，使模型更加接近实际场景。实验结果证明了改进后的模型的优越性。

参 考 文 献

[1] BASILICO N, GATTI N, AMIGONI F. Leader-follower strategies for robotic patrolling in environments with arbitrary topologies[C]// Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 1. International Foundation for Autonomous Agents and Multiagent Systems, 2009: 57-64.

[2] HAILE P A, HORTAÇSU A, KOSENOK G. On the empirical content of quantal response equilibrium[J]. The American Economic Review, 2008, 98(1): 180-200.

[3] MCKELVEY R D, PALFREY T R. Quantal response equilibria for normal form games[J]. Games and economic behavior, 1995, 10(1): 6-38.

[4] BOXALL P, PURCELL J. Strategy and human resource management[M]. Palgrave Macmillan, 2011.

[5] WRIGHT J R, LEYTON-BROWN K. Beyond Equilibrium: Predicting Human Behavior in Normal-Form Games[C]// AAAI. 2010.

[6] PITA J, JAIN M, TAMBE M, et al. Robust solutions to Stackelberg games: Addressing bounded rationality and limited observations in human cognition[J]. Artificial Intelligence, 2010, 174(15): 1142-1171.

[7] NIKNAM T. A new fuzzy adaptive hybrid particle swarm optimization algorithm for non-linear, non-smooth and non-convex economic dispatch problem[J]. Applied Energy, 2010, 87(1): 327-339.

[8] SINGH S, GUPTA P, BHATIA D. Multiparametric sensitivity analysis in programming problem with linear-plus-linear fractional objective function[J]. European Journal of Operational Research, 2005, 160(1): 232-241.

[9] CONITZER V, SANDHOLM T. Computing the optimal strategy to commit to[C]// Proceedings of the 7th ACM conference on Electronic commerce. ACM, 2006: 82-90.

[10] PARUCHURI P, PEARCE J P, MARECKI J, et al. Playing games for security: An efficient exact algorithm for solving Bayesian Stackelberg games[C]// Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 2. International Foundation for Autonomous Agents and Multiagent Systems, 2008: 895-902.

[11] XIA Z, ZHANG S. A kind of network security behavior model based on game theory[C]// International Conference on Parallel and Distributed Computing, Applications and Technologies. IEEE Xplore, 2003: 950-954.

[12] PITA J, JAIN M, ORDÓNEZ F, et al. ARMOR Security for Los Angeles International Airport[C]// AAAI. 2008: 1884-1885.

[13] JAIN M, TSAI J, PITA J, et al. Software assistants for randomized patrol planning for the lax airport police and the federal air marshal service[J]. Interfaces, 2010, 40(4): 267-290.

[14] YIN Z, JIANG A X, TAMBE M, et al. TRUSTS: Scheduling randomized patrols for fare inspection in transit systems using game theory[J]. AI Magazine, 2012, 33(4): 59.

[15] FANG F, STONE P, TAMBE M. When security games go green: Designing defender strategies to prevent poaching and illegal fishing[C]// International Joint Conference on Artificial Intelligence (IJCAI). 2015.

[16] LETCHFORD J, KORZHYK D, CONITZER V. On the value of commitment[J]. Autonomous Agents and Multi-Agent Systems, 2014, 28(6): 986-1016.

[17] JAIN M. Thwarting Adversaries with Unpredictability: Massive-scale Game-Theoretic Algorithms for Real-world Security Deployments[D]. Los Angeless: University of Southern California, 2013.

[18] SHIEH E. Not a Lone Ranger: Unleashing Defender Teamwork in Security Games[D]. Los Angeless: University of Southern California, 2015.

[19] PITA J, JAIN M, ORDÓNEZ F, et al. ARMOR Security for Los Angeles International Airport[C]// AAAI. 2008: 1884-1885.