带有欺骗证据的蜜罐博弈攻防策略优化机制

宋丽华，姜洋洋，邢长友，张国敏
（陆军工程大学指挥控制工程学院，江苏 南京 210007）

摘 要：利用博弈模型实现蜜罐行为策略的优化是提高蜜罐诱捕能力的重要手段。现有研究存在动作空间简单、割裂博弈全过程的问题。基于此，提出了带有欺骗证据的蜜罐博弈机制（HoneyED）。HoneyED 在扩大攻防动作空间的基础上，综合考虑博弈全过程，关注攻击者信念变化及这种变化对攻防策略的影响；然后基于信念求解理论均衡策略；最后基于深度反事实遗憾值最小化（Deep-CFR）设计了攻防混合策略均衡近似求解算法，得到了执行近似混合策略的攻防智能体。理论和实验结果表明，虽然攻击方在信念达到一定阈值后应及时退出博弈以获得最大收益，但所得蜜罐策略在考虑风险的情况下能尽量降低攻击方信念以诱骗其继续攻击，从而获得更大收益，且能针对具有不同欺骗识别能力的攻击方选择最佳响应。

关键词：蜜罐博弈；策略适应性；信念；欺骗证据；深度反事实遗憾值最小化
中图分类号：TP393
文献标志码：A
DOI: 10.11959/j.issn.1000−436x.2022226

Optimization mechanism of attack and defense strategy in honeypot game with evidence for deception

SONG Lihua, JIANG Yangyang, XING Changyou, ZHANG Guomin
Command & Control Engineering College, Army Engineering University, Nanjing 210007, China

Abstract: Using game theory to optimize honeypot behavior is an important method in improving defender's trapping ability. Existing work tends to use over simplified action spaces and consider isolated game stages. A game model named HoneyED with expanded action spaces and covering comprehensively the whole interaction process between a honeypot and its adversary was proposed. The model was focused on the change in the attacker's beliefs about its opponent's real identity. A pure-strategy-equilibrium involving belief was established for the model by theoretical analysis. Then, based on the idea of deep counterfactual regret minimization (Deep-CFR), an optimization algorithm was designed to find an approximate hybrid-strategy-equilibrium. Agents for both sides following hybrid strategies from the approximate equilibrium were obtained. Theoretical and experimental results show that the attacker should quit the game when its belief reaches a certain threshold for maximizing its payoff. But the defender's strategy is able to maximize the honeypot's profit by reducing the attacker's belief to extend its stay as long as possible and by selecting the most suitable response to attackers with different deception recognition abilities.

Keywords: honeypot game, strategy adaptability, belief, evidence for deception, Deep-CFR

0 引言

作为 “价值在于被探查、攻击或泄露的安全资源”[1]，蜜罐被广泛部署于网络中以获取攻击者的相关信息，提升网络系统的防御能力。为充分发挥蜜罐的诱捕能力，以及解决蜜罐策略的静态性导致的易被攻击者检测工具识破并加以利用的问题，现有研究主要关注蜜罐行为策略的适应性增强。运

收稿日期：2022−07−28；**修回日期**：2022−10−31
通信作者：姜洋洋，460734257@qq.com
基金项目：国家自然科学基金项目（No.62172432）
Foundation Item: The National Natural Science Foundation of China (No.62172432)

用博弈论和强化学习的马尔可夫决策过程（MDP, Markov decision process）模型分析防御方与攻击方的交互过程，优化蜜罐对攻击者命令的响应策略是提高其适应性的重要方法。

蜜罐动作策略的优化需要综合考虑以下两点：一是尽可能延长与攻击者的交互，这要求蜜罐在动作选择方面尽可能与正常系统保持一致，从而避免引起攻击者的怀疑；二是有意识地衡量攻击者攻击指令的潜在威胁，以免系统被攻破，这要求蜜罐不能为了降低攻击者怀疑而一味地执行攻击指令，最终造成巨大的损失。因此蜜罐行为策略的适应性增强体现在衡量上述收益和风险，针对攻击者可能采取的策略，选择最佳的动作响应。

然而，现有对蜜罐行为策略的研究存在如下局限性。博弈论方面：一是蜜罐可选动作较少，只考虑蜜罐的常规动作响应，即执行或者不执行，没有将新的蜜罐可采取动作（如伪造输出等）纳入考虑，不符合现实情况；二是没有综合考虑攻防多轮交互的全过程，只关注某个阶段，局部最优不代表全局最优，单轮最优策略可能导致博弈提前结束，损失后续博弈带来的蜜罐收益；三是没有考虑攻击者对防御方类型信念的变化对攻防策略的影响，信念通过直接影响攻击方期望收益计算从而影响攻击者的策略选择，根据贝叶斯法则，蜜罐通过不同的动作选择可以改变后续博弈的信念走向从而延续与攻击方的交互，因此忽略信念的变化也就是忽略了蜜罐策略对后续博弈的影响，也容易导致蜜罐选择局部最优策略，丧失获取更多轮收益的可能；四是传统博弈求解方法难以处理大规模决策问题。强化学习方面：一是传统强化学习方法不适用于大规模决策问题，求解能力有限；二是强化学习的研究只适用于固定策略的恶意软件，面对可动态调整策略的高级攻击者可能不存在最优解，很难学得稳定的防御策略。

面对可动态调整策略的高级攻击者，为增强蜜罐行为策略适应性应该利用博弈模型描述交互过程，结合实际情况细化博弈要素定义，综合博弈全过程考虑单个决策节点的动作选择，并寻求大规模决策问题求解方法。

本文将博弈论和深度强化学习进行结合，对蜜罐博弈模型进行了改进，并根据深度强化学习方法深度反事实遗憾最小化（Deep-CFR, deep counterfactual regret minimization）思想设计了求解博弈近似混合策略均衡的算法。本文的主要贡献如下。

1) 引入攻击方信念，将蜜罐与攻击者交互过程建模为多轮次不完全信息动态博弈模型，允许蜜罐伪造输出欺骗攻击者，而攻击者对欺骗证据具有一定的识别能力。

2) 综合博弈全过程求解了攻防双方纯策略均衡。

3) 为求解混合策略均衡，以最小化动作选择导致的收益遗憾值为目标，设计实现了基于 Deep-CFR 的近似求解算法。

1　相关工作

蜜罐行为策略的适应性增强方法中应用最广的是博弈论和强化学习方法。博弈论方面，Wagener 等[2]以安全外壳（SSH, secure shell）协议攻击为应用背景，利用多轮动态博弈模型建模 SSH 蜜罐与攻击者的交互过程，蜜罐的动作空间包括执行攻击指令（允许）与不执行攻击指令（阻塞），最后通过求解均衡得到了蜜罐的最优阻塞概率。Hayatle 等[3]以僵尸网络为应用背景，利用多轮次不完全信息动态博弈建模蜜罐和僵尸主控机之间的交互过程，均衡结果表明蜜罐不能更新其响应策略，随时间推移一定会被攻击者识别为蜜罐。王鹃等[4]提出了一种博弈论、软件定义网络（SDN, software defined network）和 docker 技术融合的动态蜜罐设计方案，该方案是一个包括低、中、高交互蜜罐的混合蜜罐，通过攻防双方不完全信息动态博弈计算出均衡解，确定选择何种蜜罐以何种行为应对。

上述研究中，文献[2]通过固定攻击方策略求解蜜罐的最优策略，文献[3]只求解了蜜罐在单轮博弈过程的最佳动作选择，并且两者在建模时均只定义了蜜罐的简单动作空间，没有考虑整个交互过程中攻击方信念的变化以及这种变化对攻防双方动作选择的影响；文献[4]虽涉及了蜜罐的具体行为，但主要优化的是蜜罐类型的选择，属于部署策略的优化，同时其按攻击阶段将博弈全过程分成多个单轮博弈，求解单轮博弈均衡。

强化学习方面，Wagener 等[5]设计了 Heliza 蜜罐，此后开发出了很多具有代表性的强化学习蜜罐，如基于深度 Q 网络算法开发的 SSH 高交互蜜罐系统[6-7]、基于逆向强化学习开发的物联网蜜罐[8]、针对自动重复恶意软件开发的蜜罐[9]和结合攻击严重性分析的 Modified-Cowrie[10]。强化学习蜜罐的共同做法是将攻击者建模为环境的一部分，通过不断交

互学得针对攻击方固定策略的最佳响应，这意味着其只适用于固定策略的攻击方，在应对策略可变的攻击方方面不如博弈论方法有效，很可能无法学得稳定的蜜罐策略。但是这些工作在定义攻防动作方面更加细致，其定义了蜜罐制造虚假输出欺骗攻击者的可选动作，若攻击者发出下载攻击工具的指令，蜜罐可以选择替换其中的部分比特，让攻击者相信其执行了指令，同时导致攻击工具不可用，或者输出伪造的更常出现的故障信息（如网页无法找到等），而不是简单地返回下载指令的错误代码，在阻塞攻击指令执行的同时，避免大幅度提升攻击方对防御方类型为密罐的怀疑。这种伪造输出欺骗的行为会带有一定的欺骗证据，攻击方对欺骗证据也具有一定的识别能力，如 Pawlick 等[11-12]基于存在蜜罐的网络中防御方选择暴露每个系统的类型或伪装系统的场景，利用信号博弈建模攻防双方的交互。该工作讨论了 2 种场景：攻击方对防御方发送的虚假信息不具备识别能力以及具备一定的识别能力，最后分别得到了相关均衡结果。然而其目标并非蜜罐行为策略的优化，且只考虑了单轮博弈过程的均衡求解。

无论是博弈论还是强化学习的现有研究，都因为使用传统博弈求解方法和传统强化学习算法而在处理大规模决策问题方面能力有限。现有研究开发出了解决大型不完全信息博弈的深度强化学习算法，其利用深度神经网络的函数近似功能使算法成功地扩展到大型状态动作空间，并能收敛到近似混合策略均衡。因此，将基于深度强化学习的近似混合策略与博弈模型结合，既能弥补传统强化学习方法在应对策略可动态调整攻击方的不足，又能针对大规模博弈问题学得稳定的蜜罐最优策略。

2 带有欺骗证据的蜜罐博弈模型

基于上述问题，本文将攻防动作空间拓展，基于多轮次非合作不完全信息动态博弈模型构建带有欺骗证据的蜜罐博弈模型（HoneyED, honeypot game with evidence for deception）：防御方可以伪造输出信息变相阻止攻击命令的执行，但是这种伪造并非无法识别，攻击方可以对防御方的输出采取相关手段进行验证，并以一定的概率识别出防御方的欺骗行为；攻击方对对手的真实身份（蜜罐或生产系统）有一定的信念，并根据防御方响应实时更新这一信念，信念会影响攻击方的动作选择。

2.1 博弈要素定义

攻击方动作空间。攻击方有两类动作：一是攻击（attack），即发送攻击命令让防御方执行，其中带有攻击工具和攻击目标等信息；二是退出（exit），即断开与蜜罐的连接，中途退出博弈。

防御方动作空间。防御方有四类动作：一是允许（allow），即执行攻击命令，并返回实际输出信息；二是阻塞（block），即不执行攻击命令，并返回常规的错误代码；三是虚假允许（f-allow），即不执行攻击命令，针对攻击命令伪造输出信息造成攻击成功的假象；四是虚假阻塞（f-block），即不执行攻击命令，针对攻击命令伪造最有可能的阻塞信息，以缓解由阻塞导致的攻击方怀疑大幅度增长。

欺骗证据。针对防御方的反馈信息，攻击方能以一定的概率识别出其欺骗行为。对于虚假允许和虚假阻塞，验证后的攻击方分别以 p_{va} 和 p_{vb} 概率识别出欺骗行为，识别出欺骗行为的攻击方将选择退出。HoneyED 博弈模型要素定义如表 1 所示。

表 1 HoneyED 博弈模型要素定义

符号	含义
$N=(N_a,N_d)$	博弈参与人，N_d 表示防御方，N_a 表示攻击方
$T=(T_a,T_d)$	参与人的类型空间，T_a = {网络攻击方}，T_d = {蜜罐，生产系统}
$A=(A_a,A_d)$	参与人的动作空间，$A_a=\{\text{attack，exit}\}$；$A_d=\{\text{allow,block,f-allow,f-block}\}$
$R_a(a_d,a_a,\theta_d)$	攻击方收益函数，其中，a_d、a_a、θ_d 分别为防御方动作、攻击方动作、防御方类型
$R_d(a_d,a_a,\theta_d)$	防御方收益函数
r_a	攻击命令成功执行时攻击方的收益
l_a	攻击命令造成的攻击方信息损失
c_a	攻击方的攻击成本
c_d	防御方维护系统的成本
c_{fa}	防御方伪造允许输出的成本
c_{fb}	防御方伪造阻塞输出的成本，且 $c_{fa}>c_{fb}$
P	防御类型的先验信念集合，$P(T_d)=\{P_0,1-P_0\}$
P'	防御类型的后验信念集合，由贝叶斯法则计算
p_{va}	攻击方验证后识别出虚假允许输出的概率
p_{vb}	攻击方验证后识别出虚假阻塞输出的概率，且 $p_{va}>p_{vb}$
$p_t=\{p_{t1},p_{t2},p_{t3}\}$	真实生产系统产生允许输出、虚假阻塞输出和阻塞输出的概率，且 $p_{t1}>p_{t2}>p_{t3}$

2.2　博弈过程

HoneyED 博弈过程如图 1 所示。带有初始信念分布的攻击者首先评估两类动作的期望收益，选择是否进行攻击以及进行何种攻击。若攻击者选择退出则博弈结束，否则蜜罐根据收到的攻击命令选择动作响应；若蜜罐选择伪造信息，攻击方将以一定的概率识别出欺骗行为，并直接退出博弈，否则攻击方认为其是真实的攻击执行信息或阻塞信息，并根据反馈信息修改其对防御方类型的信念分布，继续评估两类动作的价值并选择下一步动作，直到其选择中途退出或者完成攻击任务退出博弈。博弈过程由多个博弈阶段组成，每个阶段称为一轮博弈。攻防双方各执行一次动作即进行了一轮博弈，随后攻击方更新信念进入下一轮博弈。

3　纯策略均衡求解

本节假设攻击方完成任务需要防御方成功执行 $n(n\geqslant 1)$次允许动作，理论推导求解攻防纯策略均衡。

3.1　带有欺骗证据的一步蜜罐博弈

考虑攻击方只需要一个攻击指令被成功执行即可完成任务的情况，本文称之为带有欺骗证据的一步蜜罐博弈（1SA-HoneyED, honeypot game with evidence for deception that requires one successful action）。图 2 给出了 1SA-HoneyED 博弈过程，其中，攻击方前期通过侦察确定网络中的任一台主机部署蜜罐的概率为 P_0。即使在一步蜜罐博弈中，博弈过程仍有可能包含多个交互轮次（一轮博弈），例如，攻击方发送的前几个命令全部被蜜罐伪造阻塞输出，而攻击方未能识别出来，直到最后一个命令被蜜罐允许执行才得以攻击成功而退出。

图 2 中，空心圆圈表示决策点，黑色实心圆圈表示博弈结束，虚线方框表示以初始信念 P_0 开始的一轮博弈（攻击方和防御方各执行一次动作）。若蜜罐选择 block 或者 f-block 而未被识别，博弈将以攻击方后验信念 P' 重新开始新的一轮；若蜜罐选择 allow 或者 f-block 而未被识别，或者 f-allow（无论是否被识别），将导致攻击方认为自己完成了攻击任务或者确认对方为蜜罐而选择退出博弈，导致博弈结束。从博弈开始（初始信念 P_0）到博弈结束的完整过程称为一个 1SA-HoneyED。

关于 1SA-HoneyED，有以下结论。

定理 1　1SA-HoneyED 是有限博弈。

1SA-HoneyED 结束有以下可能：一是攻击方认为自己完成任务，由防御方选择 allow 和 f-allow 未被识别导致；二是攻击方发现了欺骗行为，直接认定防御方为蜜罐选择退出，由 f-allow 被识别和 f-block 被识别导致。博弈进入下一轮的条件是蜜罐选择 block 或者选择 f-block 而未被识别，判断博弈是否为有限过程需要分析这 2 种情况下博弈过程是否能一直持续。下面，给出这 2 种情况一定导致 1SA-HoneyED 结束的证明，为此先证明 4 个引理。

引理 1　1SA-HoneyED 中，allow 和 f-allow 不是每轮博弈中蜜罐的最优动作。

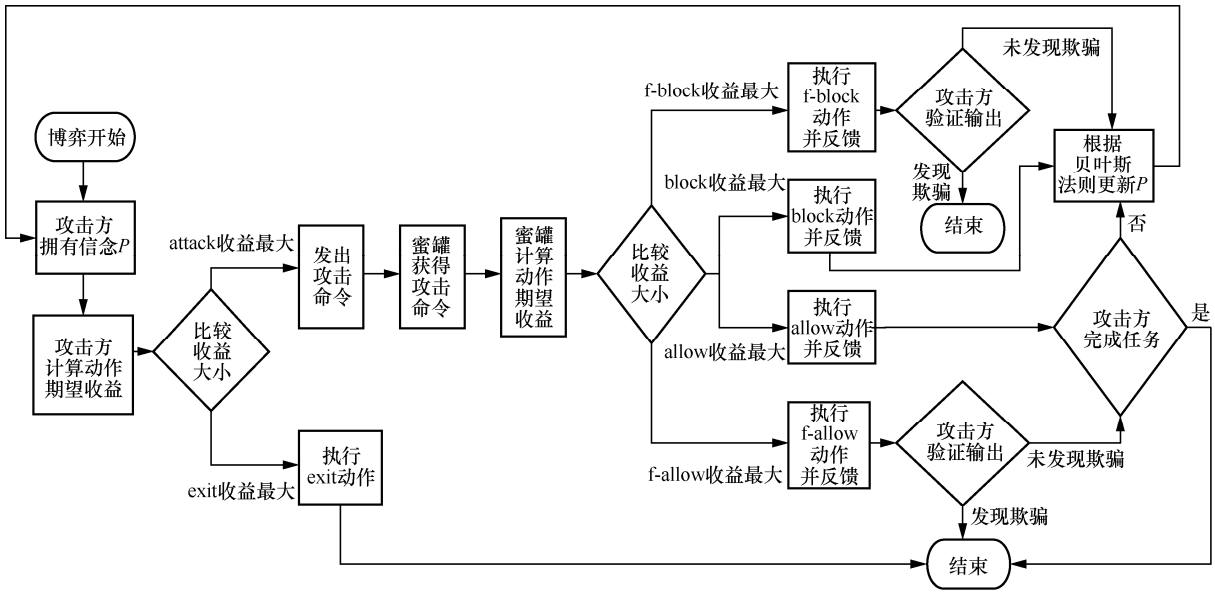


图 1　HoneyED 博弈过程

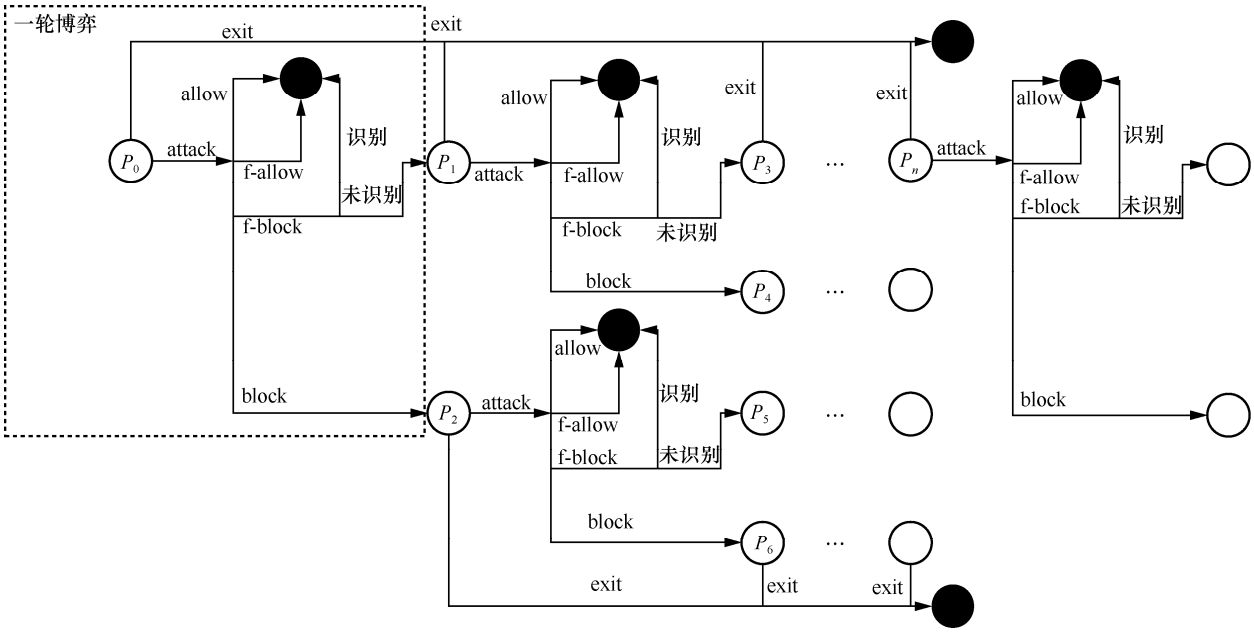


图 2　1SA-HoneyED 博弈过程

证明　若蜜罐在第 n 轮博弈中选择 allow 或者 f-allow，其获得的总期望收益为 $l_a - r_a - c_d$ 和 $l_a - c_d - c_{fa}$，选择 block 将至少获得 $l_a - c_d$ 的总收益，因为

$$l_a - c_d > l_a - c_d - c_{fa} > l_a - r_a - c_d \tag{1}$$

所以第 n 轮蜜罐选择 allow 或 f-allow 获得的并不是最大总收益，从而 allow 或 f-allow 也不是最优动作。证毕。

引理 2　1SA-HoneyED 中，攻击方对防御方为蜜罐的怀疑随博弈轮次单调上升。

证明　假设第 n 轮攻击方怀有初始信念 P_n，后验信念为 $P_n^{'}$。若蜜罐选择 f-block，该输出有 p_{vb} 的概率被攻击方识别并直接退出，因此当面临攻击方无法识别的 f-block 输出时，其认为蜜罐产生此输出的可能性是 $P_n(1-p_{vb})$，而真实生产系统产生此输出的可能性是 $p_{t2}(1-P_n)$，因此后验信念更新为

$$P_n^{'} = \frac{P_n(1-p_{vb})}{p_{t2}(1-P_n)+P_n(1-p_{vb})} \tag{2}$$

若蜜罐选择 block，则

$$P_n^{'} = \frac{P_n}{p_{t3}(1-P_n)+P_n} \tag{3}$$

假设 $1-p_{vb} > p_{t2}$，又因为 $1 > p_{t3}$，所以无论蜜罐选择何种动作，均满足 $P_n^{'} > P_n$。证毕。

引理 3　1SA-HoneyED 中，当信念增长到 $P_{\text{exit}}^{\text{1SA}} = \dfrac{p_{t1}r_a - c_a}{l_a + p_{t1}r_a}$ 时，攻击方的最优动作是选择 exit 退出博弈。

证明　基于引理 1 和引理 2 分析攻击方收益，假设某轮攻击方怀有初始信念 P_n。用 $f_{\text{attack}}(P_i)$ 表示以 P_i 为初始信念的一步蜜罐博弈攻击方能获得的总期望收益，已知每轮蜜罐不会选择 allow 或 f-allow，若蜜罐选择 block，攻击方获得即时期望收益为

$$r_n = P_n(-c_a - l_a) + p_{t1}(1-P_n)(r_a - c_a) - c_a(1-P_n)(1-p_{t1}) = P_n(-l_a - p_{t1}r_a) + p_{t1}r_a - c_a \tag{4}$$

若蜜罐选择 f-block，则

$$r_n = P_n(-c_a - l_a) + p_{t1}(1-P_n)(r_a - c_a) - c_a(1-P_n)(1-p_{t1}) = P_n(-l_a - p_{t1}r_a) + p_{t1}r_a - c_a \tag{5}$$

令 $r_n \leqslant 0$，可得 $P_n \geqslant \dfrac{p_{t1}r_a - c_a}{l_a + p_{t1}r_a}$。由于 $\dfrac{p_{t1}r_a - c_a}{l_a + p_{t1}r_a} < 1$，因此存在信念 $P_{\text{exit}}^{\text{1SA}} = \dfrac{p_{t1}r_a - c_a}{l_a + p_{t1}r_a}$。在此信念下，攻击方选择 attack 的即时期望收益等于 0。由于随着博弈轮次的增加，攻击方对防御方为蜜罐的怀疑会逐渐增加，而即时期望收益 r_n 是关于信念的减函数，因此后续每轮博弈的即时期望收益均小于 0，而总期望收益 $f_{\text{attack}}(P_i)$ 是后续各轮博弈即时期望收益的加权值，因此在后验信念下攻击方选择 attack 的总期望收益小于 0，攻击方应该在信念达到 $P_{\text{exit}}^{\text{1SA}}$ 时就选择 exit 中途退出博弈。证毕。

引理 4　随着蜜罐不断选择 block 和 f-block 未被识别，攻击方信念一定达到 $P_{\text{exit}}^{\text{1SA}}$。

证明　由引理 2 可得，攻击方信念随博弈轮次单调上升，按照式(2)和式(3)更新，且通过式(2)和式(3)的比较可得，对于同一先验信念，block 导致的后验信念更大，因此需证明蜜罐不断选择 f-block 未被识别将导致攻击方信念达到 $P_{\text{exit}}^{\text{1SA}}$。由于后验信念 P_n' 也是下一轮的先验信念，因此用 P_{n+1} 重新表示。由式(3)可得

$$\frac{1}{P_{n+1}}-1=\frac{p_{t2}}{1-p_{vb}}\left(\frac{1}{P_n}-1\right) \tag{6}$$

因此可推得 P_n 的通用表达式为 $\dfrac{1}{\left(\dfrac{p_{t2}}{1-p_{vb}}\right)^n\left(\dfrac{1}{P_0}-1\right)+1}$。

由于 $\dfrac{p_{t2}}{1-p_{vb}}<1$，可得 $\lim\limits_{n\to\infty}P_n=1$，因此一定存在 m，使 $P_m>P_{\text{exit}}^{\text{1SA}}$，蜜罐不断选择 f-block 未被识别将导致攻击方信念达到 $P_{\text{exit}}^{\text{1SA}}$。证毕。

定理 1 证明过程如下。由于 f-block 未被识别和 block 会导致攻击方信念不断增长，当信念至多增长到 $P_{\text{exit}}^{\text{1SA}}$ 时，攻击方应该选择 exit 中途退出博弈，可以得到 f-block 未被识别和 block 一定导致 1SA-HoneyED 结束，因此 1SA-HoneyED 是有限博弈。

假设 $P_0^{\min}=\dfrac{p_{t3}(p_{t1}r_a-c_a)}{(l_a+c_a)+p_{t3}(p_{t1}r_a-c_a)}\leqslant P_0$，即蜜罐选择一次 block 后会导致攻击方直接退出博弈，在此假设下，1SA-HoneyED 的均衡结果如表 2 所示。

由于当攻击方信念达到 $P_{\text{exit}}^{\text{1SA}}$ 时，会直接选择退出博弈，因此表 2 第一行初始信念区间 $[P_{\text{exit}}^{\text{1SA}},1]$ 下的均衡为(exit,block)。

考虑某轮博弈攻击方信念为 P，蜜罐选择 f-block 未被识别导致的后验信念大于或等于 $P_{\text{exit}}^{\text{1SA}}$ 且 $P<P_{\text{exit}}^{\text{1SA}}$，得到区间 $\left[\dfrac{p_{t2}(p_{t1}r_a-c_a)}{(1-p_{\text{vb}})(l_a+c_a)+p_{t3}(p_{t1}r_a-c_a)},P_{\text{exit}}^{\text{1SA}}\right)$。在此区间内，蜜罐选择 f-block 未被识别或者 block 均会导致后验信念大于 $P_{\text{exit}}^{\text{1SA}}$，因此蜜罐选择 f-block 获得总收益 $l_a-c_d-c_{fb}$，选择 block 获得总收益 l_a-c_d，蜜罐应该选择 block。因此若初始信念位于 $\left[\dfrac{p_{t2}(p_{t1}r_a-c_a)}{(1-p_{vb})(l_a+c_a)+p_{t2}(p_{t1}r_a-c_a)},P_{\text{exit}}^{\text{1SA}}\right)$，均衡结果为攻击方首先选择 attack，在收到蜜罐的 block 响应后选择 exit 退出博弈。以此类推，基于式(3)将初始信念从 $P_{\text{exit}}^{\text{1SA}}$ 不断逼近最小值 $P_0^{\min}$，比较蜜罐所能获得的总收益大小，最终得到如表 2 所示的不同条件

表 2　　**1SA-HoneyED 的均衡结果**

假设条件	初始信念区间	均衡结果
$1-\dfrac{p_{t2}}{\sqrt[n]{p_{t3}}}\geqslant p_{vb}>1-\dfrac{p_{t2}}{\sqrt[n-1]{p_{t3}}}$	$\left[P_{\text{exit}}^{\text{1SA}},1\right]$	（exit, block）
$p_{vb}<\dfrac{l_a-c_d-c_{fb}}{l_a-c_d}$	$\left[\dfrac{p_{t2}\alpha}{(1-p_{vb})\beta+p_{t2}\alpha},P_{\text{exit}}^{\text{1SA}}\right)$	（attack-exit, block）
	$\left[\dfrac{p_{t2}^2\alpha}{(1-p_{vb})^2\beta+p_{t2}^2\alpha},\dfrac{p_{t2}\alpha}{(1-p_{\text{vb}})\beta+p_{t2}\alpha}\right)$	（attack^2 -exit, f-block-block）
	$\left[\dfrac{p_{t2}^3\alpha}{(1-p_{vb})^3\beta+p_{t2}^3\alpha},\dfrac{p_{t2}^2\alpha}{(1-p_{vb})^2\beta+p_{t2}^2\alpha}\right)$	（attack^3 -exit, f-block^2 -block）
	⋮	
	$\left[\dfrac{p_{t2}^{n-1}\alpha}{(1-p_{vb})^{n-1}\beta+p_{t2}^{n-1}\alpha},\dfrac{p_{t2}^{n-2}\alpha}{(1-p_{vb})^{n-2}\beta+p_{t2}^{n-2}\alpha}\right)$	（attack^{n-1} -exit, f-block^{n-2} -block）
	$\left[P_0^{\min},\dfrac{p_{t2}^{n-1}\alpha}{(1-p_{vb})^{n-1}\beta+p_{t2}^{n-1}\alpha}\right)$	attack^n -exit, f-block^{n-1} -block）
$1-\dfrac{p_{t2}}{\sqrt[n]{p_{t3}}}\geqslant p_{vb}>1-\dfrac{p_{t2}}{\sqrt[n-1]{p_{t3}}}$	$\left[P_{\text{exit}}^{\text{1SA}},1\right]$	（exit, block）
$p_{vb}\geqslant\dfrac{l_a-c_d-c_{fb}}{l_a-c_d}$	$\left[P_0^{\min},P_{\text{exit}}^{\text{1SA}}\right)$	（attack-exit,block）

下 1SA-HoneyED 的均衡结果。为了表示方便，用 α 表示 $p_{t1}r_a - c_a$，用 β 表示 $l_a + c_a$。表 2 中，attack^n 表示攻击方选择 n 次 attack。

3.2 带有欺骗证据的两步及 n 步蜜罐博弈

用类似的方法对带有欺骗证据的两步及 n 步蜜罐博弈进行分析，即攻击方需要成功执行两次和 $n(n \geqslant 3)$次行动才完成任务的博弈，分别简写为 2SA-HoneyED 和 nSA-HoneyED，得到的结论陈述如下，因篇幅限制，此处省略证明和分析过程。

定理 2 2SA-HoneyED 是有限博弈。

定理 3 nSA-HoneyED 是有限博弈。

由于 2SA-HoneyED 中当蜜罐选择 allow 或者 f-allow 后博弈将跳转到 1SA-HoneyED，因此基于 1SA-HoneyED 的均衡收益，可以得到假设条件和 2SA-HoneyED 均衡结果分别如表 3 和表 4 所示，表 4 仅展示其中 3 种情况及其对应的假设条件组合。为表示方便，用 $P_{\text{exit}}^{\text{2SA}}$ 表示攻击方退出博弈的信念阈值 $\dfrac{2p_{t1}r_a + 2p_{t2}r_a - c_a - p_{t1}c_a - 2p_{t2}c_a}{l_a + 2p_{t1}r_a + 2p_{t2}r_a - p_{t1}c_a - 2p_{t2}c_a}$，用 η 表示 $2p_{t1}r_a + 2p_{t2}r_a - c_a - p_{t1}c_a - 2p_{t2}c_a$。

3.3 均衡结果分析

分析 3.1 节和 3.2 节得到的均衡结果，本节可以得到如下趋势性结论。

1) 信念过大将导致攻击方中途退出博弈。

2) 由于 block 和 f-block 将大幅度增加攻击方信念，因此若攻击方完成任务需要多步，为引诱攻击方继续攻击，避免其中途退出，蜜罐一开始应执行或虚假执行攻击指令。

3) 当攻击方将要完成攻击任务时，为降低风险，蜜罐应阻塞或虚假阻塞攻击指令。

4) 当攻击方欺骗识别能力较低时，由于虚假动作能有效降低风险，因此蜜罐倾向于选择虚假动作，随着攻击方欺骗识别能力的提高，虚假动作被识破风险增大，导致后续博弈收益减小，蜜罐选择真实输出操作。

随着完成任务步数的增加，分类讨论情况增多，针对攻击方识别能力得出对应均衡解的难度加大，即使 2SA-HoneyED 也很难列举出所有可能情况的纯策略均衡。因此，需要设计算法以攻击方完成任务步数及识别能力为参数自动求解均衡策略。

4 nSA-HoneyED 近似混合策略均衡求解算法

纯策略均衡是混合策略均衡的特例，考虑到混合策略可以给其他博弈参与人造成不确定性，不易被对方准确猜测等特点，本节基于 Deep-CFR 设计实现近似混合策略均衡求解算法，并构建执行混合均衡策略的攻防智能体。

反事实遗憾最小化（CFR, counterfactual regret minimization）算法[13]是目前流行的大型不完美信息博弈的近似均衡求解算法，其基本思路是计算在信息集 s 下执行动作 a 所获得的收益与信息集 s 的价值之间的差异，即遗憾值，来调整在状态 s 下执行动作 a 的概率，通过最小化单个信息集的遗憾值来实现最小化全局遗憾值的目的。CFR 算法记录每一次迭代中智能体在信息集 s 的动作选择概率，利用其得到近似所有迭代过程动作选择概率的平均策略。Deep-CFR[14]利用神经网络的强大拟合能力，构建价值网络估计遗憾值，利用监督学习构建策略网络来近似所有迭代过程的平均策略，最终训练得到的策略网络就是执行混合均衡策略的攻防智能体。

4.1 算法设计

目前研究主要将 Deep-CFR 应用于德州扑克游戏的策略优化，这类游戏的特点是参与人收益固定（赢或输），且博弈过程中不涉及信念的更新。本文在 Deep-CFR 的基础上，面向 nSA-HoneyED 重新设计模拟博弈流程，得到 nSA-HoneyED 近似混合策略均衡求解算法（nSA-HoneyED-AMSEA, approximate mixed strategy equilibrium algorithm for nSA-HoneyED）。

原 Deep-CFR 算法包括 2 个部分：使用遍历数据训练网络的外层算法 Deep-CFR 和用于模拟博弈的遍历算法 TRAVERSE，Deep-CFR 调用 TRAVERSE 获得价值样本和策略样本。nSA-HoneyED-AMSEA 在原 TRAVERSE 的基础上增加了信念更新和收益计算环节，得到带有信念更新模拟博弈遍历算法（TRAVERSE-BU, TRAVERSE algorithm with belief updating）。TRAVERSE-BU 在每一轮攻防双方根据价值网络选择动作后，基于先验信念计算攻防双方的即时收益，然后根据贝叶斯法则计算后验信念，用于下一轮即时收益的计算和信念的进一步更新，最后综合相关即时收益计算对应信息集的遗憾值。详细过程如算法 1 和图 3 所示。

表 3　　**假设条件**

编号	假设条件	编号	假设条件
1	$p_{\mathrm{va}} \leqslant 1-\dfrac{p_{t1}(p_{t1}r_a - c_a)}{2p_{t1}r_a + 2p_{t2}r_a - c_a - p_{t1}c_a - 2p_{t2}c_a}$	14	$1-\dfrac{p_{t2}p_{t1}}{(1-p_{\mathrm{vb}})p_{t3}} < p_{\mathrm{va}} \leqslant 1-p_{t1}$
2	$c_a \geqslant \dfrac{{p_{t1}}^2 r_a + 2{p_{t3}}^2 r_a - 2p_{t3}r_a + 2p_{t3}p_{\mathrm{va}}r_a - 2{p_{t3}}^2 p_{\mathrm{va}}r_a}{p_{t1} - p_{t3}(1-p_{\mathrm{va}})(1+p_{t1}+2p_{t2})}$	15	$p_{\mathrm{va}} > 1-p_{t1}$
3	$p_{\mathrm{vb}} \leqslant 1-{p_{t2}}^2\sqrt{\dfrac{(2p_{t1}r_a + 2p_{t2}r_a - c_a - p_{t1}c_a - 2p_{t2}c_a)}{p_{t3}(p_{t1}r_a - c_a)}}$	16	$p_{\mathrm{va}} \leqslant 1-\dfrac{(1-p_{\mathrm{vb}})p_{t3}}{p_{t2}}$
4	$p_{\mathrm{vb}} \leqslant 1-\dfrac{p_{t2}(2p_{t1}r_a + 2p_{t2}r_a - c_a - p_{t1}c_a - 2p_{t2}c_a)}{2p_{t1}r_a + 2p_{t3}r_a - c_a - p_{t1}c_a - 2p_{t3}c_a}$	17	$p_{\mathrm{va}} > 1-\dfrac{(1-p_{\mathrm{vb}})p_{t3}}{p_{t2}}$
5	$p_{\mathrm{vb}} < \dfrac{l_a - c_d - c_{fb}}{l_a - c_d}$	18	$\dfrac{p_{t2}p_{t1}}{(1-p_{\mathrm{vb}})p_{t3}} < p_{va} \leqslant 1-\dfrac{{p_{t2}}^2 p_{t1}}{(1-p_{\mathrm{vb}})^2 p_{t3}}$
6	$p_{\mathrm{vb}} \geqslant \dfrac{l_a - c_d - c_{fb}}{l_a - c_d}$	19	$\dfrac{l_a - c_d - c_{fa} - c_{fd} - p_{\mathrm{vd}}l_a + p_{\mathrm{vd}}c_d}{2(l_a - c_d) - c_{fd} - p_{\mathrm{vd}}l_a + p_{\mathrm{vd}}c_d} \leqslant p_{\mathrm{va}} < \dfrac{l_a - c_d - c_{fa}}{2(l_a - c_d)}$
7	$p_{\mathrm{vb}} \leqslant 1-\dfrac{r_a + c_{fb}}{l_a - c_d}$	20	$p_{\mathrm{va}} < \dfrac{l_a - c_d - c_{fa} - c_{fd} - p_{\mathrm{vd}}l_a + p_{\mathrm{vd}}c_d}{2(l_a - c_d) - c_{fd} - p_{\mathrm{vd}}l_a + p_{\mathrm{vd}}c_d}$
8	$p_{\mathrm{vb}} > 1-\dfrac{r_a + c_{fb}}{l_a - c_d}$	21	$p_{\mathrm{va}} \geqslant \dfrac{l_a - c_d - c_{fa}}{2(l_a - c_d)}$
9	$p_{\mathrm{va}} \leqslant 1-\dfrac{p_{t2}p_{t1}}{(1-p_{\mathrm{vb}})p_{t3}}$	22	$\dfrac{r_a - c_{\mathrm{fa}}}{l_a - c_d - c_{fb} + (1-p_{\mathrm{vb}})(l_a - c_d)} \leqslant p_{va} < \dfrac{r_a - c_{\mathrm{fa}} + c_{\mathrm{fb}} + p_{\mathrm{vb}}(l_a - c_d)}{2(l_a - c_d)}$
10	$1-\dfrac{p_{t2}p_{t1}}{(1-p_{\mathrm{vb}})p_{t3}} < p_{va} \leqslant 1-\dfrac{p_{t2}}{1-p_{\mathrm{vb}}}$	23	$p_{\mathrm{va}} < \dfrac{r_a - c_{\mathrm{fa}}}{l_a - c_d - c_{fb} + (1-p_{\mathrm{vb}})(l_a - c_d)}$
11	$1-\dfrac{p_{t2}}{1-p_{\mathrm{vb}}} < p_{\mathrm{va}} \leqslant 1-p_{t3}$	24	$p_{va} \geqslant \dfrac{r_a - c_{fa} + c_{fb} + p_{\mathrm{vb}}(l_a - c_d)}{2(l_a - c_d)}$
12	$p_{\mathrm{va}} > 1-p_{t3}$	25	$p_{t3} \geqslant \dfrac{p_{t2}p_{t1}}{1-p_{\mathrm{vb}}}$
13	$1-\dfrac{p_{t2}}{1-p_{\mathrm{vb}}} < p_{\mathrm{va}} < 1-\dfrac{{p_{t2}}^2}{(1-p_{\mathrm{vb}})^2}$	26	$p_{t3} < \dfrac{p_{t2}p_{t1}}{1-p_{\mathrm{vb}}}$

表 4　　**2SA-HoneyED 均衡结果**

假设条件	信念区间	均衡结果
1,2,3,4,5,9,25 1,2,3,4,5,10,14,19,25	$[P_{\mathrm{exit}}^{\mathrm{2SA}}, 1]$	（exit, block）
1,2,3,4,5,10,15,18,21,25 1,2,3,4,5,11,14,19,25 1,2,3,4,5,11,15,18, 21,25	$\left[\dfrac{p_{t2}\eta}{(1-p_{vb})\beta + p_{t2}\eta}, P_{\mathrm{exit}}^{\mathrm{2SA}}\right)$	（attack-exit, block）
1,2,3,4,5,12,15,18,21,25 1,2,3,4,5,8,9, 26 1,2,3,4,5,8,10,14,19,26	$\left[\dfrac{p_{t3}\eta}{\beta + p_{t3}\eta}, \dfrac{p_{t2}\eta}{(1-p_{vb})\beta + p_{t2}\eta}\right)$	（ attack^2 -exit, f-block-block）
1,2,3,4,5,8,10,15,18,21,26 1,2,3,4,5,8,11,14,19,26 1,2,3,4,5,8,11,15,18, 21,26	$\left[P_0^{\min}, \dfrac{p_{t3}\eta}{\beta + p_{t3}\eta}\right)$	（ attack^2 -exit, block^2 ）
	$[P_{\mathrm{exit}}^{\mathrm{2SA}}, 1]$	（exit, block）
1,2,3,4,6,10,15,18,19 1,2,3,4,6,11,15,18,19 1,2,3,4,6,12,15,18,19	$\left[\dfrac{p_{t1}p_{t3}\alpha}{(1-p_{va})\beta + p_{t1}p_{t3}\alpha}, P_{\mathrm{exit}}^{\mathrm{2SA}}\right)$	（attack-exit, block）
	$\left[P_0^{\min}, \dfrac{p_{t1}p_{t3}\alpha}{(1-p_{va})\beta + p_{t1}p_{t3}\alpha}\right)$	（ attack^2 -exit, f-allow-block）
1,2,3,4,6,9,10 1,2,3,4,6,10,14 1,2,3,4,6,11,14	$\left[P_{\mathrm{exit}}^{\mathrm{2SA}}, 1\right]$	（exit, block）
1,2,3,4,6,10,15,18,21 1,2,3,4,6,11,15,18,21 1,2,3,4,6,12,15,18,21	$\left[P_0^{\min}, P_{\mathrm{exit}}^{\mathrm{2SA}}\right)$	（attack-exit, block）

算法 1 带有信念更新的 CFR 博弈遍历算法

输入 原始动作序列 h []，博弈参与人 p，表 1 定义的博弈模型各要素，各博弈参与人的价值网络参数 θ_1,θ_2，各博弈参与人的价值样本池 $\mathcal{M}_V$，策略样本池 $\mathcal{M}_\Pi$，迭代 t 和信念值 P

输出 参与人 p 的博弈遍历收益

function TRAVERSE-BU$\left(h,p,\theta_1,\theta_2,\mathcal{M}_V,\mathcal{M}_\Pi,t,P\right)$

1) if h 是终止状态 then
2) return 博弈参与人 p 的收益
3) else if $P(h)=p$ then
4) 基于价值网络的预测值 $V\left(I(h),a|\ \theta_p\right)$ 计算动作选择概率 $\sigma^t(I)$
5) for $a\in A(h)$ do
6) if $p=2$ then#是防御方决策
7) 基于信念 P 和动作 a 计算攻防收益
8) 根据动作 a 基于贝叶斯法则更新信念 P
9) $v(a)\leftarrow$ TRAVERSE-BU $(ha,p,\theta_1,\theta_2,\mathcal{M}_V,\mathcal{M}_\Pi,t,P)$
10) for $a\in A(h)$ do
11) $\tilde{r}(I,a)\leftarrow v(a)-\sum_{a'\in A(h)}\sigma(I,a')v(a')$

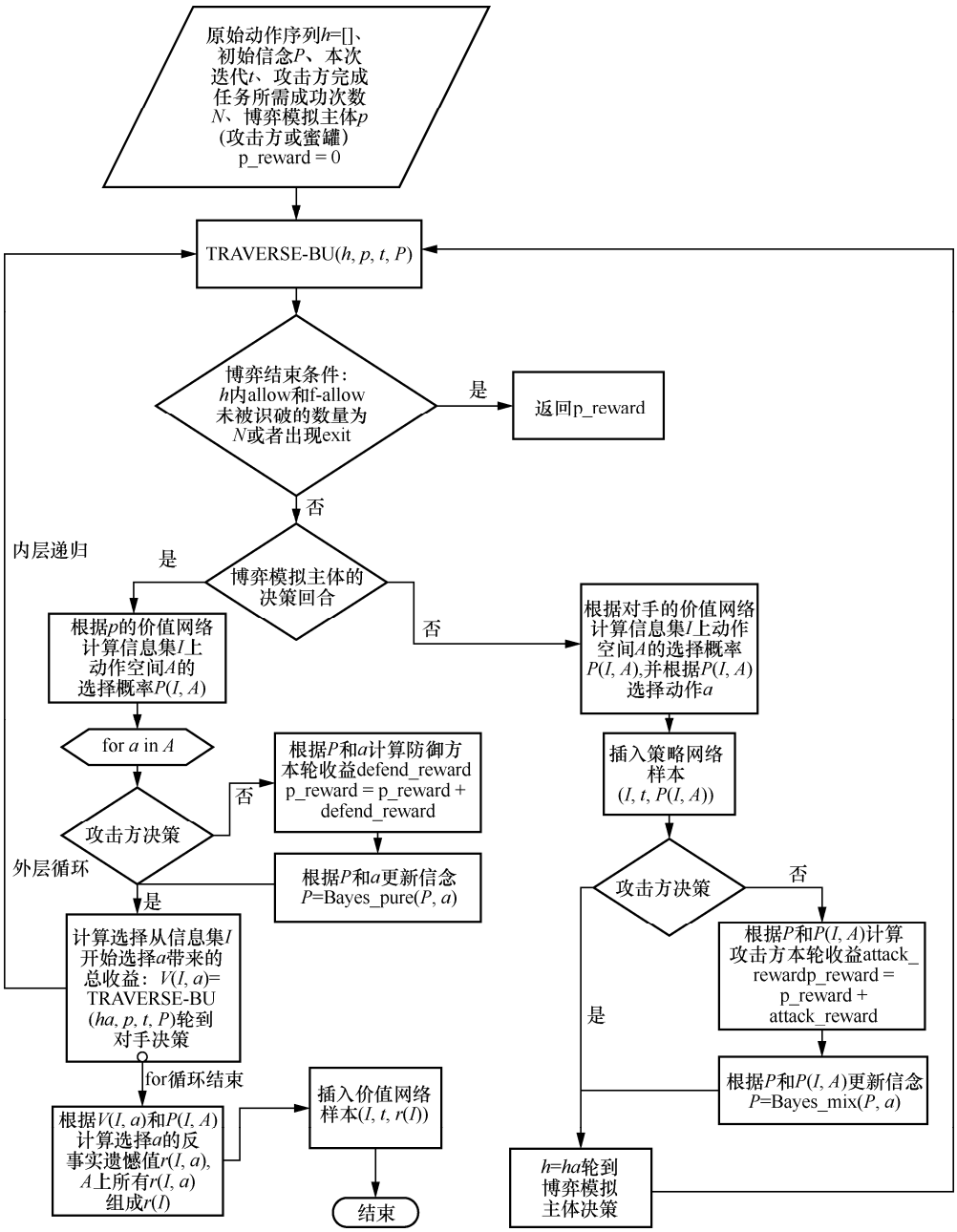


图 3 TRAVERSE-BU 算法流程

12) 价值样本$\left(I,t,\tilde{r}^t(I)\right)$插入样本池$\mathcal{M}_V$

13) else#是对手的决策回合

14) 基于价值网络的预测值$\left(I(h),a|\ \theta_{3-p}\right)$计算动作选择概率$\sigma^t(I)$

15) 策略样本$\left(I,t,\sigma^t(I)\right)$插入样本池$\mathcal{M}_\Pi$

16) 按照概率$\sigma^t(I)$选择a

17) if$3-p=2$ then#是防御方决策

18) 基于信念P和动作选择概率$\sigma^t(I)$计算攻防收益

19) 根据动作选择概率$\sigma^t(I)$基于贝叶斯法则更新信念P

20) return TRAVERSE-BU($ha, p, \theta_1, \theta_2, \mathcal{M}_V, \mathcal{M}_\Pi, t, P$)

4.2 实验结果与分析

本节通过仿真实验检验算法的有效性，验证所得策略的最优性。

4.2.1 实验设置

实验基于 SSH 攻击场景，参考文献[6,15]提出的量化方法确定博弈模型各要素取值，设定攻击方完成任务需要 2 步。为探讨攻击方欺骗识别能力大小对双方策略的影响，考虑 3 种欺骗识别概率组合：$\{p_{vb}=0.2,\ p_{va}=0.4\}$、$\{p_{vb}=0.4,\ p_{va}=0.8\}$和$\{p_{vb}=0.7,\ p_{va}=0.8\}$，分别代表低、中、高欺骗识别能力。

SSH 攻击中，攻击方可以执行系统信息查询、攻击工具下载、攻击工具运行等攻击指令，其中起决定性作用的是攻击工具的下载与运行，而蜜罐能从攻击工具下载和运行中获得关于攻击方工具库地址和攻击工具使用的相关有用信息。基于这一分析，对攻防双方动作空间进行简化，仅针对攻击方输出攻击工具下载与执行指令优化蜜罐策略。其中，attack 表示攻击方执行攻击工具下载与执行指令，allow、block、f-allow 和 f-block 表示蜜罐的响应动作，即正常执行、返回错误代码、伪造文件（如替换原始下载文件中的部分比特等）、伪造阻塞信息输出（如网页无法找到等）。

神经网络输入长度为 128 的历史动作序列，采用 Adam 优化器实现网络参数更新，每一次迭代模拟博弈 10 次。实验参数设置如表 5 所示。

表 5　　实验参数设置

实验参数	量化数值
攻击收益 r_a	7
信息收益 l_a	10
攻击成本 c_a	2
维护成本 c_d	1
f-allow 伪造成本 c_{fa}	3
f-block 伪造成本 c_{fb}	2
初始信念 P_0	0.1
p_{va}	0.4、0.8
p_{vb}	0.2、0.4、0.7
$p_t=\{p_{t1},p_{t2},p_{t3}\}$	{0.89,0.10,0.01}
学习率	0.001
批大小 batchsize	6 400

4.2.2 算法收敛性分析

图 4 展示了在$\{p_{vb}=0.7,\ p_{va}=0.8\}$组合下 2SA-HoneyED-AMSEA 运行过程中损失值随训练过程的变化情况。从图 4 可以看出，3 000 次训练即可达到良好的收敛效果。由于攻击方动作空间较小，因此攻击方价值网络比防御方收敛得快。

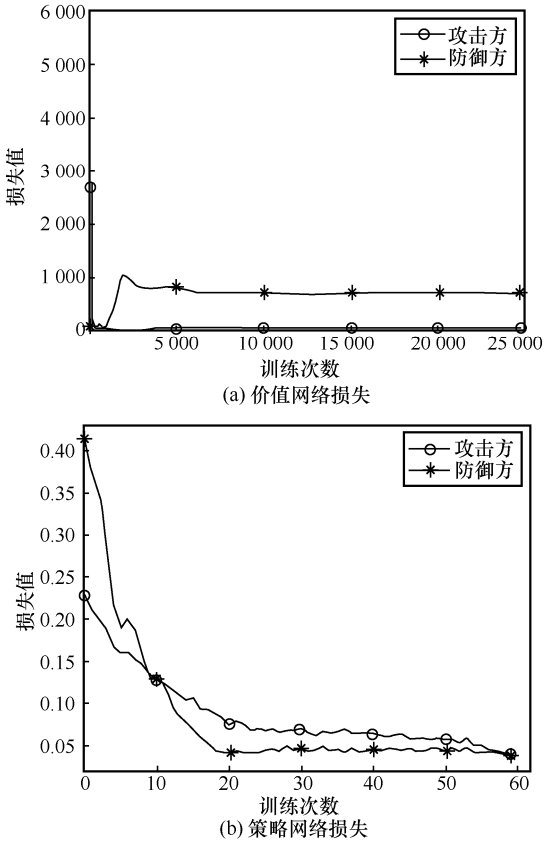


图 4　在 $\{p_{vb}=0.7,\ p_{va}=0.8\}$ 组合下 2SA-HoneyED-AMSEA 运行过程中损失值随训练过程的变化情况

4.2.3 两步博弈均衡结果分析

图 5 展示了 3 种欺骗识别概率组合下 2SA-HoneyED-AMSEA 运行得到的动作选择概率。图 5 中，S 表示博弈开始，A 表示攻击方决策，D 表示蜜罐决策，N 表示“自然”，黑色实心节点表示博弈结束。图 5 只展示了概率大于 0.1 的动作，且只画出了选择最大概率动作的博弈过程。

图 5 显示，欺骗输出动作虽然能减少即时风险，但是也减少了获得后续博弈收益的概率，因此随着攻击方识别能力的增强，后续博弈收益逐渐降低，蜜罐越来越倾向于选择真实输出。而在导致蜜罐更改最优策略的攻击方识别能力阈值上，实验与理论分析结果基本一致，详细分析如下。

首先，理论分析可以得到 1SA-HoneyED 中攻击方退出博弈的信念阈值为 $\frac{p_{t1}r_a - c_a}{l_a + p_{t1}r_a} \approx 0.3$。而图 5 中当在 1SA-HoneyED 中蜜罐选择 f-block 或 block 导致信念大于 0.3 时，攻击方均以较大概率选择退出博弈，说明以最大化期望收益为目的攻击方确实存在退出博弈的信念阈值，与理论分析结果契合。

另一方面，当 $p_{va}=0.8$ 时，蜜罐均选了 allow；当 $p_{va}=0.4$ 时，蜜罐选择了 f-allow，这一点也与理论分析吻合。后者表明当 $1-p_{va} < p_{t1}$ 且后验信念大于 $\frac{p_{t2}(p_{t1}r_a - c_a)}{(1-p_{vb})(l_a + c_a) + p_{t2}(p_{t1}r_a - c_a)}$ 时，若 $p_{va} < \frac{r_a - c_{fa}}{l_a - c_d} \approx 0.44$，则蜜罐将在 2SA-HoneyED 中选择 f-allow，不选择 allow。类似地，理论分析知 $p_{vb} < \frac{l_a - c_d - c_{fb}}{l_a - c_d} \approx 0.78$ 且后验信念小于 $\frac{p_{t2}(p_{t1}r_a - c_a)}{(1-p_{vb})(l_a + c_a) + p_{t2}(p_{t1}r_a - c_a)}$（低、中、高欺骗识别概率组合中该值分别为 0.04、0.06 和 0.11）时，蜜罐将在 1SA-HoneyED 中选择 f-block，否则选择 block。图 5 显示这 3 个阈值在实验中分别是 0.06、0.09 和 0.10，与理论结果相近。以低欺骗识别概率组合为例，实验阈值比理论阈值大是因为实验中双方采取的是混合策略，f-block 引起的怀疑较小，更容易导致攻击方继续攻击从而获得后续博弈收益，所以蜜罐更倾向于选择 f-block；而高欺骗识别概率组合中实验阈值比理论阈值小是因为 p_{vb} 较大，f-block 和 block 都将大概率导致攻击方退出，而 block 动作成本更小。

4.2.4 蜜罐策略的最优性

本节考察算法输出蜜罐策略的最优性，为此将其与三类常见策略进行对比：局部最优（只选择每一轮博弈的最优动作，即 block）、伪装（采取与生产系统相同的动作选择概率，即 allow、f-block、block 分别为 0.89、0.10、0.01）和虚假伪装策略（在伪装策略的基础上用 f-allow 代替 allow，即 f-allow、f-block、block 分别为 0.89、0.10、0.01）。实验中攻击方采用算法训练出的智能体。图 6 展示了对应的蜜罐收益。

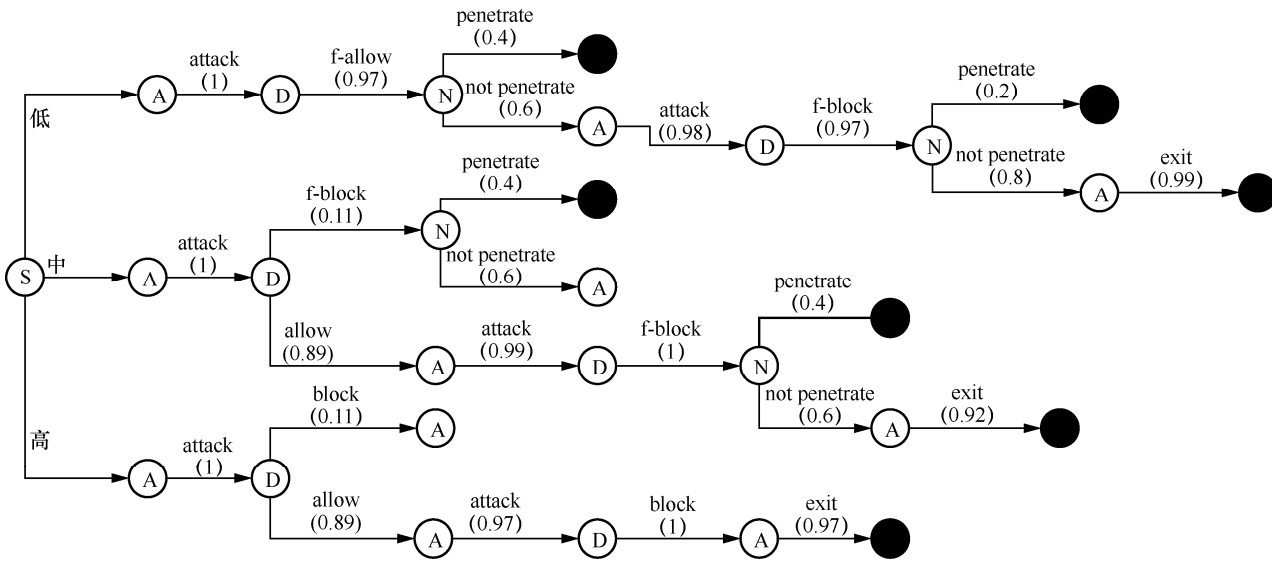


图 5 3 种欺骗识别概率组合下 2SA-HoneyED-AMSEA 运行得到的动作选择概率

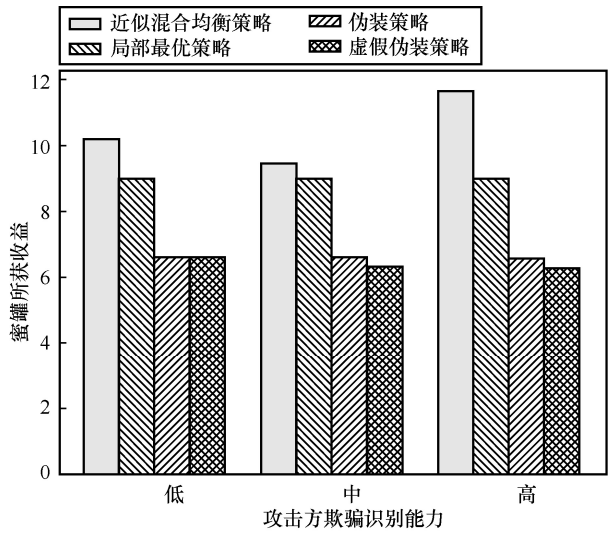


图 6　蜜罐收益对比

从图 6 可以看出，采用算法输出的策略时蜜罐收益最大。对此分析如下：攻击方在信念达到阈值后将退出博弈，而 f-block 和 block 均会导致信念的大幅度增长，因此局部最优策略没有综合考虑博弈全过程，实际上只能获得一轮收益；而伪装策略没有考虑在攻击方欺骗识别能力较低时，f-allow 相对于 allow 能获得更大的收益，同时在攻击方将要完成任务时，block 相对于 allow 是更好的选择；虚假伪装策略则没有考虑随着欺骗识别能力的提高，欺骗输出动作逐渐丧失优势；算法输出策略综合博弈全过程考虑了信念对攻击方策略的影响，能在尽量不提高攻击方怀疑的基础上针对攻击方的欺骗识别能力选择最佳动作，获得最大的收益。可以说，算法能根据攻击方策略和欺骗识别能力，适应性选择最佳动作，得到近似最优策略。

4.2.5　攻击任务复杂度对蜜罐策略的影响

用攻击方完成任务所需成功执行攻击命令数量代表攻击任务的复杂度，本节检查该变量对蜜罐最优策略的影响。图 7 展示了欺骗组合概率 $\{p_{vb}=0.2,\ p_{va}=0.4\}$ 下攻击方分别需要成功执行 2 次、4 次、8 次攻击时的实验结果。

从图 7 可以看出，当博弈为 2SA 时，蜜罐在初始博弈阶段以较大概率选择 f-allow，而后以较大概率选择 f-block；当博弈为 4SA 时，蜜罐在初始阶段以较大概率选择 allow，且概率逐渐减小，f-allow 的概率增大，而后以较大概率选择 f-block；当博弈为 8SA 时，蜜罐在初始阶段以更大概率选择 allow，而后变化趋势与 4SA 一致。

由于欺骗输出动作会以一定概率被攻击方识别，导致蜜罐是否获得后续博弈收益呈现一定的概率（连续选择 3 次 f-allow，后续攻防博弈继续进行的概率为 $(1-p_{va})^3$），因此当博弈所需成功执行攻击命令的次数增加时，蜜罐一开始将更加坚定地选择 allow，接着攻击方剩余攻击次数逐渐减小，蜜罐更倾向于选择 f-allow，攻击方信念逐渐上升，选择 exit 的概率越来越大，从而导致蜜罐更关注即时收益的获取，转而选择 f-block。

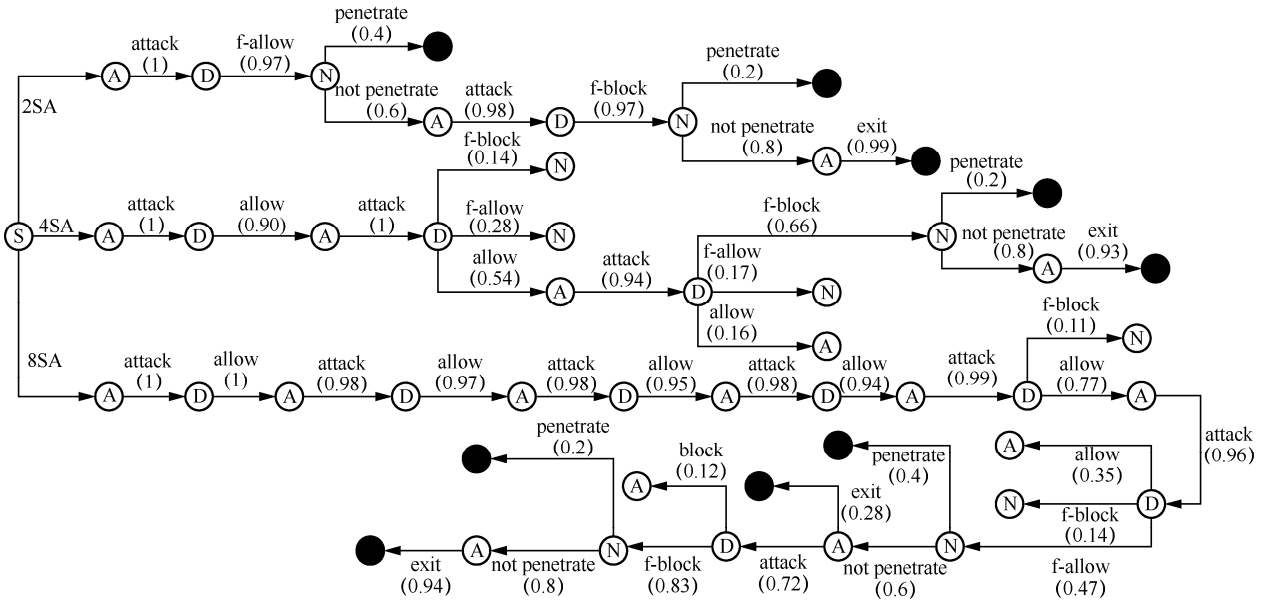


图 7　欺骗组合概率 $\{p_{vb}=0.2,\ p_{va}=0.4\}$ 下攻击方分别需要成功执行 2 次、4 次、8 次攻击时的实验结果

5 结束语

蜜罐行为策略的优化是提升蜜罐欺骗性能的重要因素，博弈论为其提供了很好的分析框架。然而，现有博弈模型动作简单、没有综合考虑博弈全过程、不符合实际攻防情况，同时所推导出的蜜罐策略只关注该轮收益的最大化，容易导致蜜罐丧失后续博弈带来的更多收益，为此本文建立了带有欺骗证据的蜜罐博弈机制，将蜜罐动作空间拓展，增加欺骗输出动作，并关注博弈全过程中攻击方对防御方类型信念的变化。针对具有不同欺骗识别能力的攻击方，本文求解了攻防纯策略均衡，并设计了基于 Deep-CFR 的近似混合策略均衡求解算法。实验表明，所提算法结果与理论分析相契合，面对欺骗识别能力弱的攻击方，蜜罐更倾向于采用欺骗输出动作。下一步工作包括进一步细化攻防双方的动作空间，优化智能求解算法。

参考文献：

[1] SPITZNER L. Honeypots: tracking hackers[M]. Reading: Addison-Wesley, 2003.

[2] WAGENER G, DULAUNOY A, ENGEL T. Self adaptive high interaction honeypots driven by game theory[C]//Symposium on Self-Stabilizing Systems. Berlin: Springer, 2009: 741-755.

[3] HAYATLE O, OTROK H, YOUSSEF A. A game theoretic investigation for high interaction honeypots[C]//Proceedings of 2012 IEEE International Conference on Communications. Piscataway: IEEE Press, 2012: 6662-6667.

[4] 王鹃, 杨泓远, 樊成阳. 一种基于多阶段攻击响应的 SDN 动态蜜罐[J]. 信息网络安全, 2021, 21(1): 27-40.
WANG J, YANG H Y, FAN C Y. A SDN dynamic honeypot with multi-phase attack response[J]. Netinfo Security, 2021, 21(1): 27-40.

[5] WAGENER G, STATE R, DULAUNOY A, et al. Heliza: talking dirty to the attackers[J]. Journal in Computer Virology, 2011, 7(3): 221-232.

[6] PAUNA A, IACOB A C, BICA I. QRASSH - a self-adaptive SSH honeypot driven by Q-learning[C]//Proceedings of 2018 International Conference on Communications (COMM). Piscataway: IEEE Press, 2018: 441-446.

[7] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Playing atari with deep reinforcement learning[J]. arXiv Preprint, arXiv: 1312.5602, 2013.

[8] PAUNA A, BICA I, POP F, et al. On the rewards of self-adaptive IoT honeypots[J]. Annals of Telecommunications, 2019, 74(7/8): 501-515.

[9] DOWLING S, SCHUKAT M, BARRETT E. New framework for adaptive and agile honeypots[J]. ETRI Journal, 2020, 42(6): 965-975.

[10] SURATKAR S, SHAH K, SOOD A, et al. An adaptive honeypot using Q-learning with severity analyzer[J]. Journal of Ambient Intelligence and Humanized Computing, 2022, 13(10): 4865-4876.

[11] PAWLICK J, ZHU Q. Deception by design: evidence-based signaling games for network defense[J]. arXiv Preprint, arXiv:1503.05458, 2015.

[12] PAWLICK J, COLBERT E, ZHU Q Y. Modeling and analysis of leaky deception using signaling games with evidence[J]. IEEE Transactions on Information Forensics and Security, 2018, 14(7): 1871-1886.

[13] ZINKEVICH M, JOHANSON M, BOWLING M, et al. Regret minimization in games with incomplete information[C]//Proceedings of the 20th International Conference on Neural Information Processing Systems. Piscataway: IEEE Press, 2007: 1729-1736.

[14] BROWN N, LERER A, GROSS S, et al. Deep counterfactual regret minimization[J]. arXiv Preprint, arXiv: 1811.00164, 2018.

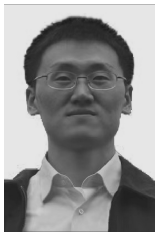
[15] RU Y Q, WANG Y F, LI J E, et al. Risk assessment of cyber attacks in ECPS based on attack tree and AHP[C]//Proceedings of 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD). Piscataway: IEEE Press, 2016: 465-470.

[作者简介]

宋丽华（1976− ），女，河北高碑店人，博士，陆军工程大学教授，主要研究方向为网络安全主动防御技术等。

姜洋洋（1998− ），男，湖北监利人，陆军工程大学硕士生，主要研究方向为网络空间安全、蜜罐、博弈论和强化学习等。

邢长友（1982− ），男，河南杞县人，博士，陆军工程大学教授，主要研究方向为软件定义网络、网络测量、网络主动防御、网络空间测绘与对抗等。



张国敏（1979− ），男，山东济南人，博士，陆军工程大学副教授，主要研究方向为网络安全、网络管理等。