DOI:10.12158/j.2096-3203.2023.04.016

基于深度强化学习的多阶段信息物理协同拓扑攻击方法

伊娜[1]，徐建军[1]，陈月[2]，孙迪康[1]
（1. 东北石油大学电气信息工程学院，黑龙江 大庆 163318；
2. 黑龙江八一农垦大学工程学院，黑龙江 大庆 163319）

摘　要：随着智能电网的发展及通信设备不断引入到信息物理系统（cyber physical system，CPS）中，CPS 正面临一种破坏性更强的新型攻击方式——信息物理协同攻击（coordinated cyber physical attack，CCPA），其隐蔽性与威胁性易导致系统出现级联故障。首先，基于攻击者的视角，提出一种多阶段信息物理协同拓扑攻击模型，单阶段的物理攻击使线路中断，双阶段的网络攻击分别用来掩盖物理攻击的断开线路和制造一条新的虚假断开线路。其次，结合深度强化学习（deep reinforcement learning，DRL）理论，提出一种基于深度 Q 网络（deep Q-network，DQN）的最小攻击资源确定方法。然后，给出攻击者考虑上层最大化物理攻击效果和下层最小化攻击代价的具体模型及求解方法。最后，以 IEEE 30 节点系统为例，验证了所提多阶段攻击模型的有效性。仿真结果表明，多阶段信息物理协同拓扑攻击较单一攻击更加隐蔽且有效，对电网的破坏程度更大，为防御此类攻击提供了参考。

关键词：信息物理系统（CPS）；信息物理协同攻击（CCPA）；拓扑攻击；负荷重分配攻击；深度强化学习（DRL）；深度 Q 网络（DQN）算法

中图分类号：TM732　　**文献标志码**：A　　**文章编号**：2096-3203（2023）04-0149-10

0　引言

随着电网信息物理系统（cyber physical system，CPS）的发展，现代电力系统的运行和控制越来越依赖于信息与通信技术（information and communications technology，ICT）[1]。一方面，ICT 是智慧城市发展的驱动力；另一方面，这些技术可能是系统运行和控制的致命弱点。网络的开放性使其极易受到恶意攻击，因此电力 CPS 网络安全问题不容忽视[2-4]。通常情况下，恶意攻击可以分成 3 类：物理攻击、网络攻击和信息物理协同攻击（coordinated cyber physical attack，CCPA）[5]。

物理攻击的目标是电力系统基础设施[6]。攻击者针对发电机、输电线路、变压器等组件发起物理攻击[7-8]，改变系统的拓扑，致使发生大面积停电事故。

网络攻击的目标是 ICT，尤其是数据采集与监视控制（supervisory control and data acquisition，SCADA）系统[9]。根据信息安全的属性，网络攻击可以分为 3 类：保密性攻击、可用性攻击和完整性攻击[10]。保密性攻击是指攻击者通过未经授权的访问或窃听获取电力系统中的保密信息，其建模方法包括 Petri 网[11]、贝叶斯网络[12]、攻击图[13]等。可用性攻击是指通过阻碍正常通信或增加系统延时的手段使系统无法获得信号[14-16]。完整性攻击主要破坏控制或量测信号，其中，虚假数据注入攻击（false data injection attack，FDIA）是完整性攻击的常见方式[17-20]。上述研究尽管达到了预期的攻击效果，但忽略了物理攻击和网络攻击的组合效果。

CCPA 的目标包括电力系统基础设施和 ICT，一旦攻击成功，会造成比单一攻击更严重的后果[21]。根据物理攻击是否具有隐蔽性，CCPA 可以分成 2 类，一类是非隐蔽性的 CCPA，即不管是否检测到物理攻击，只追求比单一攻击更大的复合效果[22-23]；另一类是隐蔽性的 CCPA，即攻击者首先破坏电力系统的某些组件，再进行网络攻击以掩盖物理攻击[2,24]。无论是否为隐蔽攻击，CCPA 的研究多数涉及负荷重分配（load redistribution，LR）攻击。文献[25]首次研究了一种特殊的 FDIA-LR 攻击，通过 LR 攻击使电网调度人员错误地掌握电网的负荷分配情况，进而做出错误的调度决策[26-27]。在非隐蔽性的 CCPA 中，文献[22]基于双层模型的方法，分析了 LR 攻击分别与线路和发电机之间的协同攻击场景。隐蔽性的 CCPA 可以分为线路维持攻击和线路移除攻击[28]。在线路维持攻击中，文献[2]提出与物理攻击相互协调的两步 LR 攻击以掩盖物理攻击中的断开线路。在线路移除攻击中，文献[28]针对线路移除攻击提出一种局部拓扑攻击模型，通过获取较少的网络信息确定线路可行攻击区域。

以上方法虽然证明了 CCPA 的潜力，但并未考虑线路维持攻击和线路移除攻击相结合的恶意攻击效果。另外，传统方法均假设攻击者可以根据历

收稿日期：2023-01-18；修回日期：2023-03-09

史负荷数据准确预测电力系统的负荷，而事实上，攻击瞬间的负荷存在不确定性，因此，引入深度强化学习（deep reinforcement learning，DRL）算法[29]，假设攻击瞬间的负荷是随机变量并且遵循一定的概率分布。与此同时，当实施 CCPA 时，一些关键线路会被严格保护，由于级联效应，断开这些关键线路可能导致电力系统连锁故障。鉴于高度保护性，只能通过间接的方法使关键线路跳闸。

为此，文中提出一种基于 DRL 的多阶段信息物理协同拓扑攻击方法，攻击的目的为使关键线路潮流越限。首先，在隐蔽性的 CCPA 中提出一种新的攻击场景，在物理攻击后，考虑线路维持攻击和线路移除攻击相结合的恶意攻击效果，另外，通过间接攻击使关键线路过载；其次，结合 DRL 理论，提出一种基于深度 Q 网络（deep Q-network，DQN）的最小攻击资源确定方法；最后，在攻击者的视角，提出一种双层 CCPA 模型，上层模型以最大化物理攻击效果为目标，下层模型以最小化网络攻击代价来掩盖物理攻击并制造一条新的虚假断开线路。仿真结果表明，相比于单一攻击方式，文中提出的攻击方法破坏性更强，有助于挖掘电力系统中的薄弱环节，为制定相应的防御策略提供参考。

1 信息物理协同拓扑攻击

1.1 CCPA 形式

随着信息通信系统与电力系统深度融合，CPS 不再局限于受到单一攻击。CCPA 如图 1 所示。

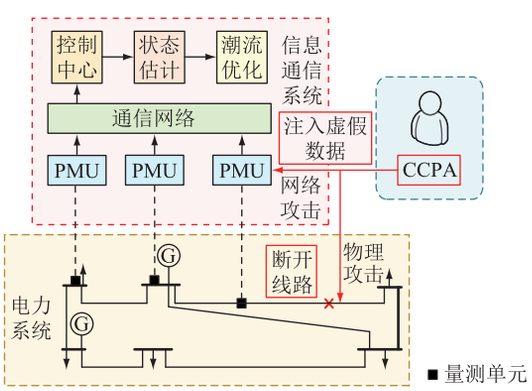


图 1 CCPA 示意

Fig.1 Schematic diagram of CCPA

在信息通信系统中，同步相量测量单元（phasor measurement unit，PMU）将电力系统中的线路潮流量测数据传输到控制中心，对其进行状态估计与不良数据辨识，然后再进行潮流优化等操作。信息物理协同攻击主要包括两部分：一是在电力系统中（物理侧）进行物理攻击，断开目标线路；二是在信息通信系统（信息侧）进行网络攻击，使调度中心观测到的线路状态为攻击者篡改后的状态，对电网的整体安全造成威胁[30-32]。CCPA 的隐蔽性与破坏性更强，能有效地发现电网中的脆弱点。

1.2 多阶段信息物理协同拓扑攻击模型

电力系统的拓扑可以用母线-支路模型 $T=(B,L)$ 表示，其中 B 为母线，L 为输电线路。如果攻击者在物理侧进行物理攻击，断开输电线路 M_l，电力系统的拓扑结构变为 $T_1=(B,L\backslash\{M_l\})$，通过检查输电线路的开关状态信息，控制中心可以检测出这种物理拓扑攻击，如图 2 所示。

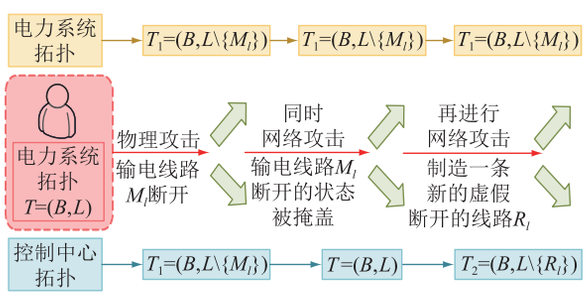


图 2 多阶段协同拓扑攻击过程

Fig.2 Multi-stage coordinated topology attacks process

为了使控制中心观测到的拓扑为 $T_2=(B,L\backslash\{R_l\})$，攻击者须掩盖物理侧的跳闸线路 M_l 并在信息侧制造一条新的虚假断开线路 R_l，即对负荷量测值和线路潮流量测值进行篡改，发起一种特殊的 FDIA——LR 攻击。精心设计的 LR 攻击服从基尔霍夫电压定律和电流定律，并成功绕过了不良数据检测机制，干扰电力系统的正常运行。此外，多阶段的协同拓扑攻击模型会使具有充分裕度特征的电网变得脆弱，甚至造成电力系统崩溃。

由于线性化的直流模型鲁棒性更强，文中以直流潮流模型为例，实施多阶段信息物理协同拓扑攻击。图 3（a）和图 3（b）分别为物理攻击前和物理攻击后的电力系统状态，D_i、D_j 分别为节点 i、j 攻击前的负荷量测值；P_{ij} 为线路 ij 的功率。

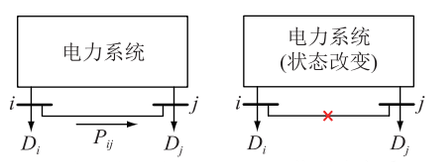
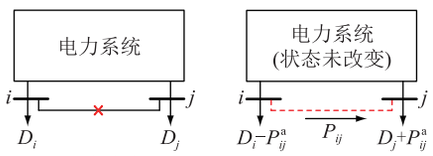


图 3 物理攻击前后的电力系统状态

Fig.3 Power system state before and after the physical attack

线路维持拓扑攻击主要用来掩盖物理侧的跳闸线路。图 4（a）为网络攻击前的电力系统状态，发起线路维持拓扑攻击后首先将线路的开关状态篡改为闭合，然后向节点负荷和线路潮流量测量中注

入虚假数据,使物理侧的跳闸线路仍然是连接状态,如图 4(b)所示。



(a) 网络攻击前 (b) 网络攻击后

图 4 线路维持拓扑攻击前后的电力系统状态

Fig.4 Power system state before and after the line-maintaining topology attack

攻击后的线路功率满足:

$$P_{ij}^{a}=\frac{\delta_i-\delta_j}{x_{ij}} \tag{1}$$

式中:P_{ij}^{a}为攻击后线路 ij 上的功率;δ_i、δ_j分别为节点 i、j 的电压相角;x_{ij}为线路 ij 的电抗。

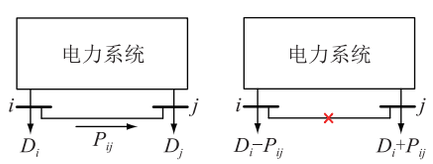
线路维持拓扑攻击后的节点负荷量测值为:

$$D_i^{a}=D_i-P_{ij}^{a} \tag{2}$$

$$D_j^{a}=D_j+P_{ij}^{a} \tag{3}$$

式中:D_i^{a}、D_j^{a} 分别为节点 i、j 攻击后的负荷量测值。

线路移除拓扑攻击主要用来生成信息侧的跳闸线路。图 5(a)为网络攻击前的电力系统状态,随后发起线路移除拓扑攻击,即向目标线路的量测数据中注入虚假数据,使其在信息侧为断开状态,如图 5(b)所示。



(a) 网络攻击前 (b) 网络攻击后

图 5 线路移除拓扑攻击前后的电力系统状态

Fig.5 Power system state before and after the line-removing topology attack

攻击后的节点负荷量测值应满足:

$$D_i^{a}=D_i+P_{ij} \tag{4}$$

$$D_j^{a}=D_j-P_{ij} \tag{5}$$

1.3 LR 攻击

攻击者在发起一次 LR 攻击时,控制中心通常会基于历史负荷数据设定一个负荷变化百分比阈值 τ,尽管在实施攻击时攻击者倾向于尽可能多地向负荷节点注入虚假数据,但是调度人员会检测到这些线路上的潮流突变并实时重新调度。因此,攻击者应使节点负荷量测值的变化在控制中心“不可检测”的范围内[28]。除此之外,由于 τ 的限制,很多情况下线路维持拓扑攻击和线路移除拓扑攻击不足以掩盖物理侧或生成信息侧的跳闸线路,因此攻击者还需要同时向跳闸线路的周围节点注入虚假数据,更改其量测值,即发起 LR 攻击。

站在攻击者的角度,力求通过消耗最少的攻击资源达到最优的攻击效果,LR 攻击模型表示为:

$$\min\sum_{i\in B}c_{B,i}\zeta_{B,i}+\sum_{l\in L}c_{L,l}\zeta_{L,l} \tag{6}$$

$$-\tau\zeta_{B,i}P_i\leqslant\Delta P_i\leqslant\tau\zeta_{B,i}P_i\quad i\in B\backslash B_{\text{pmu}} \tag{7}$$

$$-\zeta_{L,l}P_{ij}^{\max}\leqslant P_{ij}+\Delta P_{ij}\leqslant\zeta_{L,l}P_{ij}^{\max}\quad l\in L\backslash L_{\text{pmu}} \tag{8}$$

$$P_{ij}^{a}=\frac{\delta_i-\delta_j}{x_{ij}}\quad ij\in L;i,j\in B \tag{9}$$

$$-\pi\leqslant\delta_n\leqslant\pi \tag{10}$$

$$P_i^{a}-\sum_{i\in I}G_i+\sum_{j\in J}P_{ij}^{a}=0 \tag{11}$$

$$\sum_{i=1}^{K}\Delta P_i=0 \tag{12}$$

$$\Delta P_i=0\quad i\in B_{\text{pmu}} \tag{13}$$

$$\Delta P_{ij}=0\quad ij\in L_{\text{pmu}} \tag{14}$$

式中:$c_{B,i}$、$c_{L,l}$ 分别为篡改节点 i 的负荷量测值和篡改线路 l 的潮流量测值所需要的攻击资源;$\zeta_{B,i}$、$\zeta_{L,l}$ 为二进制变量,如果节点 i 的负荷量测值和线路 l 的潮流量测值被篡改,其值为 1,否则为 0;P_i 为节点 i 的负荷;ΔP_i 为注入到节点 i 的虚假负荷量测值;B_{pmu} 为配置 PMU 的节点集合;ΔP_{ij} 为注入到线路 ij 的功率虚假数据;$P_{ij}^{\max}$ 为线路 ij 的热稳定极限;L_{pmu} 为 PMU 能观测到的线路集合;δ_n 为攻击后节点 n 的电压相角;P_i^{a} 为攻击后节点 i 的负荷;G_i 为节点 i 的发电量;I 为节点 i 的发电机数量;J 为与节点 i 相连的节点集合;K 为负荷节点数量。式(7)表示注入到负荷节点的虚假数据应在一定的范围内,式(8)表示攻击后的潮流不超过其极限功率,式(9)为 LR 攻击须满足的潮流计算规律,式(10)为节点电压相角约束条件,式(11)为节点功率平衡约束,式(12)表示注入到负荷节点的攻击量和为 0,式(13)、式(14)表示部署 PMU 的节点和可被观测到线路的量测值均不可篡改。

2 基于 DRL 的最小攻击资源确定方法

2.1 DRL 原理

文中提出一种基于 DRL 的算法来确定发起 LR 攻击的最小攻击资源,如图 6 所示。将攻击者视为 DRL 算法中的智能体,交互环境为电力系统。智能体的行为 a_t 主要执行攻击动作,状态 s_t 为环境中的量测状态信息,DRL 的目标为确定最小的攻击资源以掩盖物理侧的跳闸线路和生成信息侧的虚假跳闸线路。

在 DRL 模型中,状态、动作和奖励函数的定义

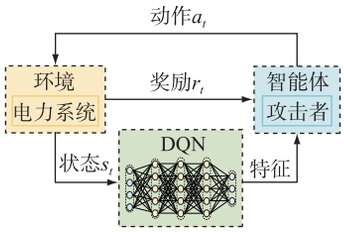


图 6 基于 DRL 最小攻击资源模型示意

Fig.6 Schematic diagram of minimum attack resource model based on DRL

如下：

（1）状态。定义状态变量为节点负荷和线路潮流的量测状态向量，初始状态是所有元素均为 0 的向量。

（2）动作。将文中智能体的控制动作为执行攻击操作，即改变节点负荷量测值和线路潮流量测值，被攻击的节点和线路的量测状态从 0 变为 1。

（3）奖励函数。智能体在执行一个动作后，会获得正反馈或负反馈信号，奖励函数会对智能体的决策行为产生影响。定义奖励函数为：

$$r_t = -N_1 c_{B,i} - N_2 c_{L,l} \tag{15}$$

式中：r_t 为 t 时刻的奖励函数；N_1、N_2 分别为需要更改的节点负荷量测值和线路潮流量测值数量。

通过与环境进行不断地交互，智能体学习最佳动作集合，获得总体最大奖励，t 时刻累计回报为：

$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \tag{16}$$

式中：$\gamma \in [0,1]$，为折扣系数。

强化学习目标是寻找一系列决策，实现总体奖励最大化。根据贝尔曼方程更新动作-价值函数 Q：

$$Q(s_t,a_t) = E(r_t + \gamma \max_{a_{t+1}} Q(s_{t+1},a_{t+1}) \mid s_t,a_t) \tag{17}$$

式中：s_t 为 t 时刻的状态；a_t 为 t 时刻的动作；s_{t+1} 为 $t+1$ 时刻的状态；a_{t+1} 为 $t+1$ 时刻的动作；$E(\cdot)$ 为期望函数。

通过贪婪策略，选取 Q 值最大的动作：

$$a^* = \operatorname{argmax}_{a_t} Q^*(s_t,a_t) \tag{18}$$

式中：a^* 为 Q 值最大的动作；Q^* 为最佳 Q 值。

2.2 DQN 训练算法

随着问题规模的增加，状态数与动作数逐渐增加，为了解决 Q-学习算法中无法存储高维数据的缺陷，引入 DQN[29]。DQN 存在 2 个结构相同但参数不同的神经网络，分别为 Q 神经网络和 Q-target 神经网络。Q 网络随着迭代更新网络参数，当 Q 网络的训练达到一定迭代次数后，复制其网络参数，实现 Q-target 网络参数的更新。DQN 算法更新损失函数表示为：

$$L(\theta_h) = E(r_t + \gamma \max_{a_{t+1}} Q(s_{t+1},a_{t+1};\theta_{t+1,h}) - Q(s_t,a_t;\theta_{t,h})) \tag{19}$$

式中：$\theta_{t,h}$、$\theta_{t+1,h}$ 分别为第 h 步时 Q 网络和 Q-target 网络参数。

DQN 算法的训练过程如图 7 所示。训练开始时，2 个模型使用相同的参数。Q 网络从环境中获取当前状态 s_t 后，执行动作 a_t 反馈给环境，此时状态变为 s_{t+1}，输入到 Q-target 网络中。将目标 Q 值与 Q 网络中的 $Q(s_t,a_t;\theta_{t,h})$ 值代入损失函数，更新 Q 网络中的参数，然后 Q 网络从环境中获取新状态 s_{t+1}，重复训练过程。通过在一段时间内固定 Q-target 网络，可以保证梯度的稳定下降，减轻模型的波动性。经验回放池不仅可以储存历史样本（s_t，a_t，r_t，s_{t+1}），还能在缓存中均匀随机地采样并对样本进行学习，在保证训练效果稳定的同时提高了样本利用率。

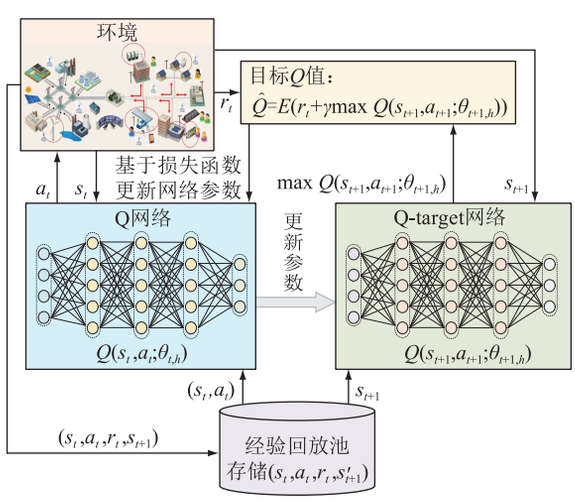


图 7 DQN 算法训练流程

Fig.7 Training flow chart for DQN algorithm

2.3 基于 DQN 算法的 LR 攻击策略

为了利用最小的攻击资源掩盖物理跳闸线路，提出算法 1 和算法 2 求解式(6)。算法 1 为基于 LR 攻击的 DQN 训练过程，算法 2 为利用 LR 攻击策略去掩盖物理侧跳闸线路的训练过程。

2.3.1 基于 LR 攻击的 DQN 算法流程

步骤 1：经验回放池初始化，以随机权重 θ 初始化 Q 网络和 Q-target 网络。

步骤 2：初始化状态 s_1，根据式(1)计算物理侧跳闸线路 M_l 的潮流量测量的改变量。

步骤 3：基于 ε-贪婪策略执行动作 a_t，选择一个负荷节点；否则选择动作 $a_t = \operatorname{argmax}_a Q(s_t,a_t)$。

步骤 4:执行动作 a_t,在阈值范围内改变节点负荷量测值和线路潮流量测值,得到奖励 r_t 和新状态 s_{t+1}。

步骤 5:将经验(s_t,a_t,r_t,s_{t+1})储存在经验回放池中;在经验回放池中随机抽取小批量经验样本,利用抽取的样本与最近的样本进行训练,打乱经验数据之间的相关性。

步骤 6:如果 episode 在 $d+1$ 步终止,则目标动作价值 $y_d=r_d$;否则 $y_d=r_d+\gamma\max\limits_{a_t}Q(s_{d+1},a_t;\theta)$。

步骤 7:通过梯度下降法,根据式(19)的损失函数更新 Q 网络参数,每经过 C 步,将 Q-target 网络的参数替换为 Q 网络的参数。

步骤 8:如果 $t=T$(T 为时间步),进行步骤 9;否则,$t=t+1$,跳转至步骤 3。

步骤 9:如果 episode 值达到终止值,进行步骤 10;否则,episode 值加 1,跳转至步骤 2。

步骤 10:保存 Q 网络,训练结束。

在算法 1 中,输入为电力系统基准、奖励函数 r 和物理侧跳闸线路 M_l,输出为 DQN 训练参数 θ。首先对经验回放池、批次、回合数、折扣系数、权重等参数进行初始化,episode 取 1。进入循环后,对环境状态 s_1进行初始化,计算物理侧跳闸线路 M_l的潮流测量的改变量。每个 episode 以时间 $t=1$ 开始,以 $t=T$ 结束。在这个循环中,选择动作 a_t后执行动作 a_t改变量测值,得到奖励 r_t和新状态 s_{t+1},将经验(s_t,a_t,r_t,s_{t+1})储存在经验回放池中。在经验回放池中随机抽取小批量经验样本,以打乱经验数据之间的相关性。如果 episode 在 $d+1$ 步终止,则目标动作价值 $y_d=r_d$;否则 $y_d=r_d+\gamma\max\limits_{a_t}Q(s_{d+1},a_t;\theta)$。最后,通过梯度下降法更新 Q 网络参数,每经过 C 步,将Q-target网络的参数替换为 Q 网络的参数,结束循环。

2.3.2 LR 攻击策略掩盖物理侧跳闸线路的训练流程

步骤 1:加载算法 1 中训练得到的网络参数 θ。

步骤 2:初始化状态 s_1,获取允许负荷测量变化的阈值和传输线路的潮流限值。

步骤 3:根据 Q 网络计算动作-价值函数 Q。

步骤 4:执行动作 a_t,$a_t=\operatorname{argmax}\limits_{a_t}Q(s_t,a_t;\theta)$。

步骤 5:如果 $t=T$,进行步骤 6;否则,$t=t+1$,跳转至步骤 2。

步骤 6:输出要改变的量测量,训练结束。

在算法 2 中,输入为电力系统基准状态、负荷节点、线路潮流和物理攻击后系统的拓扑,输出为改变的节点负荷量测量和线路潮流量测量。首先,加载算法 1 中训练得到的参数 θ;然后,进入循环,获取 LR 攻击掩盖物理侧跳闸线路的策略。初始时,根据 Q 网络计算动作-价值函数 Q,最终以 $a_t=\operatorname{argmax}\limits_{a_t}Q(s_t,a_t;\theta)$ 执行动作,确定改变的量测量。

2.3.3 LR 攻击策略生成信息侧虚假跳闸线路的训练流程

在训练完算法 1 和算法 2 后,使用与算法 1 相同的状态、动作、奖励函数和结构建立 DQN 模型,设计算法 3 和算法 4,用以确定 LR 攻击策略生成信息侧虚假跳闸线路。与算法 1、算法 2 不同的是,算法 3 和算法 4 先发起线路移除拓扑攻击生成信息侧跳闸线路,然后再发起 LR 攻击生成侧跳闸信号。算法 3 的训练流程与算法 1 相同,输入为电力系统基准、奖励函数 r 和信息侧跳闸线路 R_l。在训练完成后,基于 LR 攻击策略生成信息侧跳闸线路网络结构的算法 4 的训练流程与算法 2 相同,输入为电力系统基准状态、负荷节点、线路潮流和网络攻击后系统的拓扑。由于算法 3 和 4 具体的训练流程与算法 1 和 2 大致相同,在此不再赘述。

3 信息物理协同拓扑攻击策略的双层优化模型

文中的信息物理协同拓扑攻击模型综合考虑了三阶段攻击的交互过程,如图 8 所示,从攻击者的视角构建了双层优化模型。在上层模型中,攻击者寻找最优的物理攻击策略使关键线路过载量最大,给电网造成严重的损失;在下层模型中,攻击者通过篡改最少的量测量实施 LR 攻击,使控制中心接收到的拓扑与电力系统真实的拓扑差异性变大,给电网带来灾难性的后果。

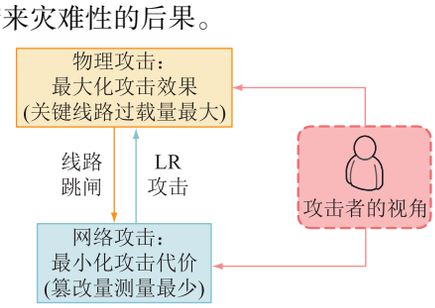


图 8 信息物理协同拓扑攻击策略的双层模型

Fig.8 Bilevel model for coordinated cyber-physical topology attack strategies

在上层模型中,当物理攻击和网络攻击资源一定时,攻击者寻求最优的物理攻击策略使关键线路过载量最大,目标函数可以表示为:

$$\max\sum_{M_l\in L}|\Delta O_l|\,p_{M_l} \tag{20}$$

$$\sum_{M_l \in L} R_{M_l} p_{M_l} + \sum_{i \in B} c_{B,i} \zeta_{B,i} + \sum_{l \in L} c_{L,l} \zeta_{L,l} \leqslant R \quad (21)$$

$$P_l^{(1)} = \frac{\delta_i - \delta_j}{x_{ij}} \quad ij \in L; i,j \in B \quad (22)$$

$$P_l^{(2)} = P_l^{(1)} + \Delta P_l \quad l \in L \quad (23)$$

$$\Delta O_l = P_l^{(2)} - P_l^{\max} \quad l \in L \quad (24)$$

式中：ΔO_l为关键线路 l 的过载量；p_{M_l} 为二进制变量，如果线路 M_l被物理攻击，其值为 1，否则为 0；R_{M_l} 为物理攻击线路 M_l需要的攻击资源；R 为攻击资源总量；$P_l^{(1)}$ 为系统未遭受到物理攻击时关键线路 l 的线路潮流值；$P_l^{(2)}$ 为线路 M_l被物理攻击后关键线路 l 的潮流值；ΔP_l为潮流改变量；$P_l^{\max}$ 为关键线路 l 的热稳定极限值。式(21)表示实施 CCPA 的总资源约束，式(22)表示系统须满足其潮流计算规律。

在下层模型中，当物理攻击目标确定后，攻击者基于 DQN 算法寻求最少的攻击资源发起 LR 攻击，使篡改的量测量最少，即先掩盖物理攻击线路 M_l，再制造一条新的虚假断开线路 R_l，目标函数可以表示为式(6)，约束条件为式(7)—式(14)。

文中所建立的双层优化模型的上、下层之间存在耦合关系，上层优化的目标为使关键线路的过载量最大，优化函数为混合整数线性规划模型，利用 MATLAB 调用 Cplex 求解器可以快速求解。求解完毕后将最优结果传递给下层模型，采用 DQN 算法寻求最少的攻击资源。

4 数值仿真与结果分析

4.1 IEEE 30 节点系统参数设置

为了更好地验证所提多阶段信息物理协同拓扑攻击模型的有效性，以 IEEE 6 机 30 节点系统为例进行仿真分析，该系统有 30 个母线和 41 条线路，如图 9 所示。对部分参数进行修改，节点 5 和 16 的负荷分别设置为 73.6 MW 和 23.5 MW，其余参数从 MATPOWER 潮流计算软件包 case30.m 文件中获取。$c_{B,i}$、$c_{L,l}$攻击资源的定义与文献[33]相同，负荷变化 τ 的上限为真实值的 50%。

4.2 最优拓扑攻击策略

在 IEEE 30 节点系统中，通过启发式的拓扑搜索方法得到在节点 1、9、14、19、25、30 部署 PMU 是最优的配置方案。系统中的关键线路 O_{l1}和 O_{l2}分别对应线路 4—6 和 12—16。攻击者首先通过单阶段的物理攻击使线路 O_{l1}过载，再通过基于 DQN 算法的两阶段 LR 攻击分别掩盖物理攻击的断开线路 M_l并制造一条虚假断开线路 R_l使线路 O_{l2}过载。通过

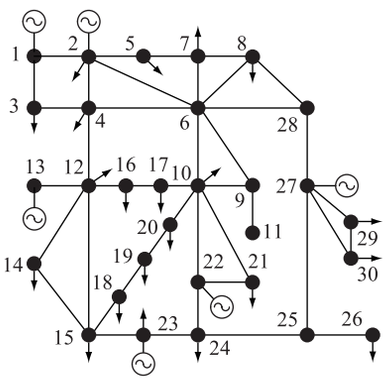


图 9 IEEE 6 机 30 节点系统

Fig.9 IEEE 6-machine 30-bus system

对所提模型进行求解，确定最优的拓扑攻击策略，即物理攻击的 M_l为线路 6(即线路 2—6)，网络攻击的 R_l为线路 26(即线路 10—17)，分别如图 10 和图 11 所示。在攻击者的视角，首先对系统发起物理攻击，造成线路 2—6 断线，进而导致线路 4—6 过载，同时可能会引发相关断路器在短时间跳闸。如果系统运行人员没有快速响应，线路 4—6 潜在的中断可能会导致线路 6—7 的连续过载和中断。随着过载线路的增多，系统的故障情况愈发严重，使电网更加脆弱，最终导致系统出现连锁故障。因此，即使没有发起后续的网络攻击，线路 4—6 的过载也会恶化物理攻击的影响。同理，线路 O_{l2}的过载也会引发系统的连锁故障。

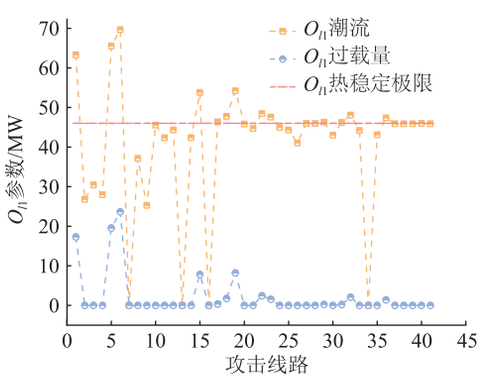


图 10 物理攻击的最优策略

Fig.10 Optimal strategy for physical attack

4.3 物理侧的最优 LR 攻击策略

在物理攻击后，线路 2—6 的真实潮流为 0，为了掩盖线路中断，其线路潮流应为 34.38 MW。如果只发起线路维持拓扑攻击，节点 2 的负荷量测值应由 21.7 MW 改为−12.68 MW，节点 6 的负荷量测值应由 0 改为 34.38 MW。然而，由于篡改阈值的限制，应协同改变与节点 2 和节点 6 相连节点的负荷量测值。

根据算法 1 训练 DQN 以学习掩盖物理侧跳闸

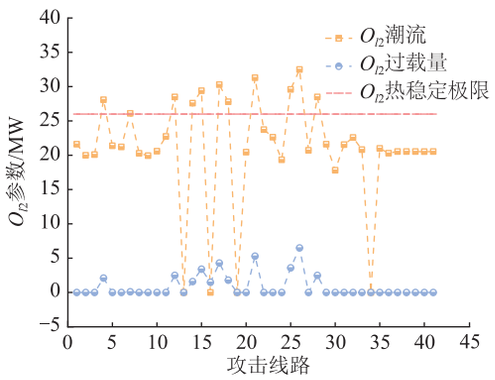


图 11 网络攻击的最优策略

Fig.11 Optimal strategy for cyber attack

线路的 LR 攻击策略,训练的 Q 网络是三层全连接神经网络结构,对模型进行了 2 000 次迭代,训练过程如图 12 所示。ε 的初始值设置为 0.6,每轮迭代减小 0.004,最终保持在 0.001。训练开始时,由于智能体处于探索状态,因此曲线波动较大。在前 800 次迭代中,智能体从搜索空间中随机选择攻击策略,随着智能体与环境不断地进行交互,回合奖励逐渐趋于稳定,在 800 次迭代时基本收敛,确定最优的攻击策略。当迭代次数大于 800 后,由于 DQN 智能体在计算回合奖励过程中的自身误差,曲线仍存在一定波动,但回合奖励逐渐趋于稳定且浮动范围逐渐减小,因此可以认为算法达到收敛。

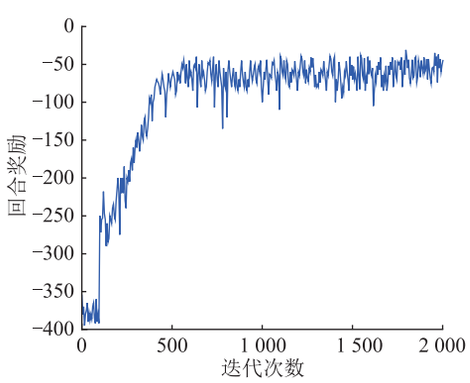


图 12 DQN 智能体掩盖物理侧跳闸线路训练奖励曲线

Fig.12 Training reward curve of DQN agent for masking the physical line outage

基于训练后的 DQN 模型生成需要更改的测量量。通过发起 LR 攻击,得到节点 2、5、7 和 8 的负荷量测值需要同时被篡改,网络攻击前后节点负荷量测值具体篡改结果如表 1 所示。与此同时,线路 2—5、2—6、5—7、6—7、8—28、6—28 潮流量测值也需要被篡改,具体篡改结果如表 2 所示。图 13 显示了为掩盖物理侧跳闸线路需要更改的量测数量,在 2 000 次迭代后,耗用的最少攻击资源减少到 16。当有效地实施后续协同的网络攻击时,由于协同攻击具有隐蔽性,物理攻击造成的线路中断会被掩盖,反馈给状态估计系统的数据与物理攻击前一致,诱导控制中心做出错误决策,给电网的稳定运行带来更严重的影响。除此之外,物理攻击仅会对局部电网造成负荷损失,而协同攻击会对全网多点造成负荷损失,使系统更加脆弱。

表 1 为掩盖物理跳闸线路 LR 攻击前和攻击后的节点负荷量测值

Table 1 Load measurements before and after LR attack for masking the physical line outage MW

节点	网络攻击前节点负荷量测值	需要改变的量测值	网络攻击后节点负荷量测值
2	21.700 0	7.984 3	29.684 3
5	73.600 0	−34.384 3	39.215 7
7	22.800 0	11.400 0	34.200 0
8	30.000 0	15.000 0	45.000 0

表 2 为掩盖物理跳闸线路 LR 攻击前和攻击后的线路潮流量测值

Table 2 Power flow measurements before and after LR attack for masking the physical line outage

MW

线路	网络攻击前线路潮流量测值	需要改变的量测值	网络攻击后线路潮流量测值
5(2—5)	65.970 0	−42.368 5	23.601 5
6(2—6)	0	34.384 3	34.384 3
8(5—7)	−7.630 0	−7.984 3	−15.614 3
9(6—7)	30.430 0	19.384 3	49.814 3
40(8—28)	−5.020 0	−15.000 0	−20.020 0
41(6—28)	−0.090 0	15.000 0	14.910 0

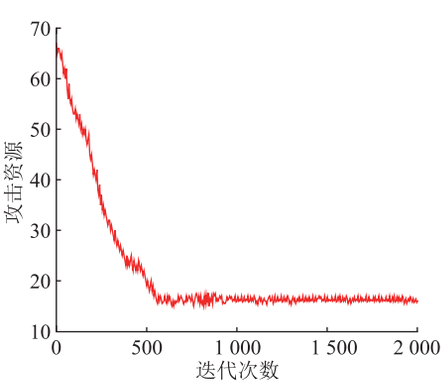


图 13 掩盖物理侧跳闸线路耗用攻击资源

Fig.13 Attack resources for masking the physical line outage

4.4 信息侧的最优 LR 攻击策略

在第一阶段 LR 攻击成功掩盖物理攻击后,发起第二阶段的 LR 攻击,根据算法 3 训练 DQN 以学习生成信息侧跳闸线路的攻击策略,对模型进行 2 500次迭代,训练过程如图 14 所示。

由于负荷变化上限的约束,需要同时篡改节点

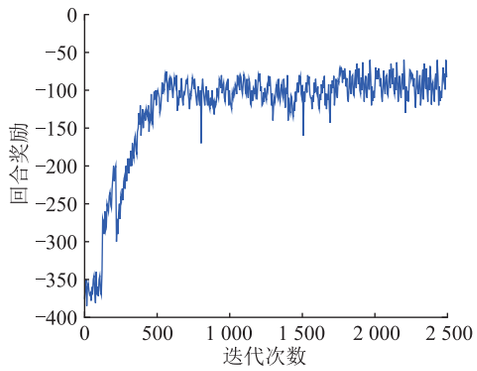


图 14　DQN 智能体生成信息侧跳闸线路训练奖励曲线

Fig.14　Training reward curve of DQN agent for creating the cyber line outage

10、16、17 和 21 的负荷量测值，线路 2—4、2—6、6—10、4—12、12—16、16—17、10—17 和 10—21 的潮流量测值，攻击前后的详细篡改信息如表 3 和表 4 所示。

表 3　为生成网络跳闸线路 LR 攻击前和攻击后的节点负荷量测值

Table 3　Load measurements before and after LR attack for creating the cyber line outage　MW

节点	网络攻击前节点负荷量测值	需要改变的量测值	网络攻击后节点负荷量测值
10	7.800 0	−2.000 0	5.800 0
16	33.461 0	−9.961 0	23.500 0
17	4.500 0	4.500 0	9.000 0
21	10.039 0	7.461 0	17.500 0

表 4　为生成网络跳闸线路 LR 攻击前和攻击后的线路潮流量测值

Table 4　Power flow measurements before and after LR attack for creating the cyber line outage　MW

线路	网络攻击前线路潮流量测值	需要改变的量测值	网络攻击后线路潮流量测值
3(2—4)	19.640 0	6.500 0	26.140 0
6(2—6)	40.310 0	−6.500 0	33.810 0
12(6—10)	10.210 0	−6.500 0	3.710 0
15(4—12)	8.470 0	6.500 0	14.970 0
19(12—16)	26.000 0	6.500 0	32.500 0
21(16—17)	−7.461 0	16.461 0	9.000 0
26(10—17)	11.961 0	−11.961 0	0
27(10—21)	−8.901 0	7.461 0	−1.440 0

图 15 显示了为生成信息侧跳闸线路需要更改的量测数量，在 2 500 次迭代后，最少攻击资源减少到 20。此时控制中心观测到的拓扑结构与系统真实的结构差异性逐渐变大，致使调度人员频繁做出错误的决策。当有效地发起 CCPA 时，会给系统造成更大面积的停电，面对 CCPA 的电网更加脆弱。除此之外，电力系统的脆弱点不再受重负载输电线路的限制，一些轻载输电线路极有可能被恶意利用，干扰电力系统的稳定运行。

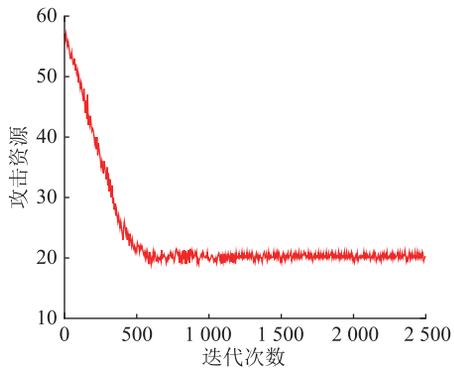


图 15　生成信息侧跳闸线路耗用攻击资源

Fig.15　Attack resources for creating the cyber line outage

5　结论

在总结和分析已有研究结果的基础上，文中提出了一种基于 DRL 的多阶段信息物理协同拓扑攻击方法，该模型考虑了物理攻击后，将线路维持拓扑攻击和线路移除拓扑攻击相结合形成两阶段网络攻击方式，其恶意攻击的结果是过载系统中的两条关键线路，以挖掘网络攻击下系统的脆弱性。多阶段信息物理协同拓扑攻击模型为电网防御者制定防御策略提供了参考与指导依据。通过在 IEEE 30 节点系统进行仿真得到如下结论：

（1）物理攻击或网络攻击会使电力系统中的关键线路过载，如果运行人员没有快速采取缓解措施，故障的蔓延可能会导致系统出现连锁故障。

（2）线路维持拓扑攻击通过注入虚假量测数据可以掩盖物理层的线路断开状态；线路移除拓扑攻击通过注入虚假量测数据可以生成信息层的虚假断开线路。

（3）基于多阶段的信息物理协同拓扑攻击模型会使控制中心观测到的拓扑与系统真实拓扑差异性变大，致使调度人员频繁做出错误决策，增大系统发生连锁故障的可能性。文中目前的工作是基于攻击方的视角研究 CCPA 模型对电力系统的影响，下一步将对防御模型进行研究。

参考文献：

[1] 徐飞阳，薛安成，常乃超，等. 电力系统自动发电控制网络攻击与防御研究现状与展望[J]. 电力系统自动化，2021，45(3)：3-14.

XU Feiyang，XUE Ancheng，CHANG Naichao，et al. Research status and prospect of cyber attack and defense on automatic generation control in power system[J]. Automation of Electric Power Systems，2021，45(3)：3-14.

[2] LI Z Y,SHAHIDEHPOUR M,ALABDULWAHAB A,et al. Bi-level model for analyzing coordinated cyber-physical attacks on power systems[J]. IEEE Transactions on Smart Grid,2016,7(5):2260-2272.

[3] YI N,WANG Q,YAN L M,et al. A multi-stage game model for the false data injection attack from attacker's perspective[J]. Sustainable Energy,Grids and Networks,2021,28:100541.

[4] 钱胜,王琦,颜云松,等. 计及网络攻击影响的安全稳定控制系统风险评估方法[J]. 电力工程技术,2022,41(3):14-21.
QIAN Sheng,WANG Qi,YAN Yunsong,et al. Risk assessment method of security and stability control system considering the impact of cyber attacks[J]. Electric Power Engineering Technology,2022,41(3):14-21.

[5] JENA P K,GHOSH S,KOLEY E. A binary-optimization-based coordinated cyber-physical attack for disrupting electricity market operation[J]. IEEE Systems Journal,2021,15(2):2619-2629.

[6] 田继伟,王布宏,李腾耀,等. 智能电网虚假数据注入攻击研究进展与展望[J]. 网络空间安全,2019,10(9):73-84.
TIAN Jiwei,WANG Buhong,LI Tengyao,et al. Research progress and prospects of false data injection attacks in smart grid[J]. Cyberspace Security,2019,10(9):73-84.

[7] HUANG Y D,HE T,CHAUDHURI N R,et al. Preventing outages under coordinated cyber-physical attack with secured PMUs[J]. IEEE Transactions on Smart Grid,2022,13(4):3160-3173.

[8] 韦晓广,高仕斌,李多,等. 基于连锁故障网络图和不同攻击方式的输电线路脆弱性分析[J]. 中国电机工程学报,2018,38(2):465-474,677.
WEI Xiaoguang,GAO Shibin,LI Duo,et al. Cascading fault graph for the analysis of transmission network vulnerability under different attacks[J]. Proceedings of the CSEE,2018,38(2):465-474,677.

[9] LIANG G Q,ZHAO J H,LUO F J,et al. A review of false data injection attacks against modern power systems[J]. IEEE Transactions on Smart Grid,2017,8(4):1630-1638.

[10] 汤奕,王琦,倪明,等. 电力信息物理融合系统中的网络攻击分析[J]. 电力系统自动化,2016,40(6):148-151.
TANG Yi,WANG Qi,NI Ming,et al. Analysis of cyber attacks in cyber physical power system[J]. Automation of Electric Power Systems,2016,40(6):148-151.

[11] CHEN T M,SANCHEZ-AARNOUTSE J C,BUFORD J. Petri net modeling of cyber-physical attacks on smart grid[J]. IEEE Transactions on Smart Grid,2011,2(4):741-749.

[12] 罗新宇,段斌,吴俊锋,等. 基于证据推理的风电场 SCADA 系统安全脆弱性定量评估方法[J]. 电力系统自动化,2020,44(11):25-31.
LUO Xinyu,DUAN Bin,WU Junfeng,et al. Evidence reasoning based quantitative evaluation method for vulnerability of SCADA system security in wind farms[J]. Automation of Electric Power Systems,2020,44(11):25-31.

[13] WANG H,CHEN Z F,ZHAO J P,et al. A vulnerability assessment method in industrial Internet of Things based on attack graph and maximum flow[J]. IEEE Access,2018,6:8599-8609.

[14] 黄博南,詹凤楠,张天闻,等. 一种针对电-热综合能源系统经济调度的 DoS 最优攻击策略[J]. 中国电机工程学报,2020,40(21):6839-6854.
HUANG Bonan,ZHAN Fengnan,ZHANG Tianwen,et al. An optimal DoS attack strategy against the economic dispatch for electric-thermal integrated energy system[J]. Proceedings of the CSEE,2020,40(21):6839-6854.

[15] WANG D,LI J,JIA H,et al. Prospects of key technologies of integrated energy systems for rural electrification in China[J]. Global Energy Interconnection,2021,4(1):3-17.

[16] LU X,WANG J,LIU G,et al. Station-and-network-coordinated planning of integrated energy system considering integrated demand response[J]. Global Energy Interconnection,2021,4(1):39-47.

[17] DENG R L,ZHUANG P,LIANG H. False data injection attacks against state estimation in power distribution systems[J]. IEEE Transactions on Smart Grid,2019,10(3):2871-2881.

[18] 王琦,邰伟,汤奕,等. 面向电力信息物理系统的虚假数据注入攻击研究综述[J]. 自动化学报,2019,45(1):72-83.
WANG Qi,TAI Wei,TANG Yi,et al. A review on false data injection attack toward cyber-physical power system[J]. Acta Automatica Sinica,2019,45(1):72-83.

[19] 肖鹏,王柯强,黄振林. 基于 IABC 和聚类优化 RBF 神经网络的电力信息网络安全态势评估[J]. 智慧电力,2022,50(6):100-106.
XIAO Peng,WANG Keqiang,HUANG Zhenlin. Security situation assessment of power information network based on IABC & clustering optimized RBF neural network[J]. Smart Power,2022,50(6):100-106.

[20] 计丽妍,李存斌,贾雪枫,等. 多证据融合下电力信息物理系统风险评估研究[J]. 智慧电力,2021,49(10):23-29.
JI Liyan,LI Cunbin,JIA Xuefeng,et al. Risk assessment of cyber-physical power system based on multi-evidence fusion[J]. Smart Power,2021,49(10):23-29.

[21] LIANG G Q,WELLER S R,ZHAO J H,et al. The 2015 Ukraine blackout:implications for false data injection attacks[J]. IEEE Transactions on Power Systems,2017,32(4):3317-3318.

[22] XIANG Y M,WANG L F,LIU N. Coordinated attacks on electric power systems in a cyber-physical environment[J]. Electric Power Systems Research,2017,149:156-168.

[23] TIAN J W,WANG B H,LI T Y,et al. Coordinated cyber-physical attacks considering DoS attacks in power systems[J]. International Journal of Robust and Nonlinear Control,2020,30(11):4345-4358.

[24] CHUNG H M,LI W T,YUEN C,et al. Local cyber-physical attack for masking line outage and topology attack in smart grid[J]. IEEE Transactions on Smart Grid,2019,10(4):4577-4588.

[25] YUAN Y L,LI Z Y,REN K. Modeling load redistribution at-

tacks in power systems[J]. IEEE Transactions on Smart Grid, 2011,2(2):382-390.

[26] PINCETI A,SANKAR L,KOSUT O. Detection and localization of load redistribution attacks on large-scale systems[J]. Journal of Modern Power Systems and Clean Energy,2022,10(2):361-370.

[27] LIU Y G,GAO S B,SHI J,et al. Sequential-mining-based vulnerable branches identification for the transmission network under continuous load redistribution attacks[J]. IEEE Transactions on Smart Grid,2020,11(6):5151-5160.

[28] LIU X,LI Z Y. Local topology attacks in smart grids[J]. IEEE Transactions on Smart Grid,2017,8(6):2617-2626.

[29] WANG Z H,HE H B,WAN Z Q,et al. Coordinated topology attacks in smart grid using deep reinforcement learning[J]. IEEE Transactions on Industrial Informatics, 2021, 17(2):1407-1415.

[30] LI J,SUN C,SU Q. Analysis of cascading failures of power cyber-physical systems considering false data injection attacks [J]. Global Energy Interconnection,2021,4(2):204-213.

[31] WANG Z,QI D,MEI J,et al. Real-time controller hardware-in-the-loop co-simulation testbed for cooperative control strategy for cyber-physical power system[J]. Global Energy Interconnection,2021,4(2):214-224.

[32] 田猛,董政呈,王先培,等. 目标冲突下电力信息物理协同攻击分析[J]. 电网技术,2019,43(7):2336-2344.
TIAN Meng,DONG Zhengcheng,WANG Xianpei,et al. Analysis of electrical coordinated cyber physical attacks under goal conflict[J]. Power System Technology,2019,43(7):2336-2344.

[33] LI Z Y,SHAHIDEHPOUR M,ALABDULWAHAB A,et al. Analyzing locally coordinated cyber-physical attacks for undetectable line outages[J]. IEEE Transactions on Smart Grid, 2018,9(1):35-47.

作者简介:

伊娜

伊娜(1997),女,博士在读,研究方向为信息物理系统网络安全(E-mail:nepuyina@163.com);

徐建军(1971),男,博士,教授,研究方向为电力系统稳定控制;

陈月(1996),女,硕士,助教,研究方向为多源信息融合。

A multi-stage coordinated cyber-physical topology attack method based on deep reinforcement learning

YI Na[1], XU Jianjun[1], CHEN Yue[2], SUN Dikang[1]

(1. School of Electrical Engineering and Information,Northeast Petroleum University,Daqing 163318,China;
2. College of Engineering,Heilongjiang Bayi Agricultural University,Daqing 163319,China)

Abstract:With the development of smart grid and the continuous introduction of communication equipments into cyber physical system (CPS),CPS is confronted with a new attack mode with more destructive—coordinated cyber physical attack (CCPA). CCPA is not only hidden but also threatening,which is easy to cause cascading failures. Firstly,from the perspective of the attacker,a multi-stage coordinated cyber-physical topology attack model is proposed. The single-stage physical attack first trips a transmission line,and the two-stage cyber attack is used to mask the outage signal of the disconnected line in the physical layer and then create a new fake tripped line in the cyber layer. Secondly,combined with deep reinforcement learning (DRL) theory, the method for determining the minimum attack resources based on deep Q-network (DQN) is proposed. Then,the specific model and solution method for the attacker are given,taking the maximization of the physical attack effect in the upper layer and minimization of the attack cost in the lower layer into consideration. Finally,the IEEE 30-bus system is taken as an example to verify the effectiveness of the proposed multi-stage attack model. The simulation results demonstrate that the multi-stage coordinated cyber-physical topology attack is more hidden and effective than the single attack,and the damage to the power grid is greater,which provides a reference for the defender against such attacks.

Keywords:cyber physical system (CPS);coordinated cyber physical attack (CCPA);topology attack;load redistribution attack;deep reinforcement learning (DRL);deep Q-network (DQN) algorithm

(编辑　钱悦)