DOI:10.20033/j.1003-7241.(2026)02-0054-05

基于深度 Q-learning 算法的智能电网管控模型研究

王 筠[1], 李志鹏[2], 项 旭[1], 张军堂[2], 石雷波[2]

(1.国网安徽电力信息通信分公司, 安徽 合肥 230041; 2.安徽明生恒卓科技有限公司, 安徽 合肥 230000)

摘 要:设计基于深度 Q-learning 算法的智能电网管控模型,将可验证声明(verifiable credential,VC)和分布式数字身份(decentralized identity,DID)作为应用程序身份凭证与软件定义网络(software-defined networking,SDN)控制器,结合动态信任评估算法与基于属性的访问控制策略,构建基于区块链的智能电网分布式 SDN 管控模型。在资源分配、网络拓扑动态变化以及安全威胁不断演变的情况下,实施基于区块链的分布式 SDN 网络的优化。实验测试结果表明,设计方法在通过深度 Q-learning 优化模型后累积奖励明显大幅增加,在多种安全性能方面表现出色,能够清除恶意域,确保网络环境的安全。

关键词:SDN 控制器;分布式 SDN 网络;深度 Q-learning 算法;区块链;智能电网管控模型

中图分类号: TP181;TN73 **文献标志码**: A **文章编号**: 1003-7241(2026)02-0054-05

Research on intelligent grid control model based on deep Q-learning algorithm

WANG Yun[1], LI Zhipeng[2], XIANG Xu[1], ZHANG Juntang[2], SHI Leibo[2]

(1. State Grid Anhui Electric Power Information and Communication Branch, Hefei 230041, Anhui, China;
2. Anhui Mingsheng Hengzhuo Technology Co., Ltd., Hefei 230000, Anhui, China)

Abstract: A smart grid management and control model based on deep Q-learning algorithm is designed. verifiable credential (VC) and decentralized identity(DID) are used as application credentials and software-defined networking (SDN) controllers,and a distributed SDN management and control model of smart grid based on blockchain is constructed by combining dynamic trust evaluation algorithm and attribute-based access control strategy. In the case of resource allocation, dynamic changes of network topology and evolving security threats,the optimization of distributed SDN network based on blockchain is implemented. The experimental test results show that the cumulative reward of the design method is significantly increased after optimizing the model through deep Q-learning, and it performs well in various security performances, which can clear malicious domains and ensure the security of the network environment.

Keywords: SDN controller; distributed SDN network; deep Q-learning algorithm; Blockchain; smart grid control model

智能电网已成为电力系统的重要发展方向,它融合了先进的通信技术、信息技术和控制技术,旨在实现电力系统的智能化、自动化和分布式管理。然而,智能电网在带来诸多便利的同时,也面临着严峻的网络安全挑战[1]。一方面,智能电网的广泛互联和高度集成使得系统更加复杂,传统的安全防护手段已难以适应新的安全威胁。网络攻击者可能利用系统漏洞、恶意软件等手段,对智能电网进行渗透和破坏,导致电力中断、数据泄露等严重后果。这些安全事件不仅会影响人们的正常生活和社会经济的稳定运行,还可能对国家安全构成威胁[2]。另一方面,智能电网的分布式特性和资源共享需求也对管控模型提出了更高的要求。传统的集中式管控模式已难以满足智能电网的分布式管理和资源优化需求,而分布式管控模型则能够更好地适应智能电网的特性和需求[3]。

智能电网管控模型作为智能电网技术的核心组成部分,其研究现状在国内外均呈现出蓬勃发展的态势。文献[4]从光伏电池中获取直流电,并利用电力电子技术将其转换为有用的直流和交流并与电网相连。使用微控制器检测发电电压和电流水平并实施监测和控制。利用物联网在云中存储和显示监测值,供客户查看。文献[5]探讨了区块链技术在智能电网安全控制中的应用,设计了一种区块链框架,并展示了区块结构及其所蕴含的高级区块链技术要素。尽管区块链本身不能直接满足隐私需求,但通过结合零知识证明(zero knowledge proof,ZKP)和椭圆曲线数字签名算法(elliptic curve digital signature algorithm,ECDSA)等强化加密手段,可以有效提升隐私防护水平。该文还深入探讨了区块链的数据不可篡改性,强调其在防范电网,特别是发电与配电系统遭受网络攻击方面的作用。此外,还介绍了一个基于智能合约的安全设备维护体系,旨在实现高效故障诊断。文献[6]介绍了一种将边缘

收稿日期:2024-10-20;**录用日期**:2024-12-11

基金项目:国家电网科技项目资助(SGGSPX00HLWJS2200097)

作者简介:王 筠(1975—),女,硕士,高级工程师,研究方向:项目管理、项目建设。

引用本文:王筠, 李志鹏, 项旭, 等. 基于深度 Q-learning 算法的智能电网管控模型研究[J]. 自动化技术与应用, 2026,45(2):54-57,142. (WANG Yun, LI Zhipeng, XIANG Xu, et al. Research on intelligent grid control model based on deep Q-learning algorithm[J]. Techniques of Automation and Applications, 2026,45(2):54-57,142.)

计算技术融入电力物联网系统的新方法,旨在对庞大的智能用电数据进行本地化处理,从而减轻通信网络的负担,避免数据传输拥堵,满足电力物联网对于快速响应与精确执行的高标准。文章详细说明了所设计的数据库结构、边缘计算服务器的部署以及数据处理流程,并提出了一种针对物联网智能终端设备异构性的解决方案,增强了系统的兼容性和边缘数据的即时处理能力。文献[7]提出了一种针对电力智能终端设备的“即时探测-灵活管理-主动监督”安全管控框架,旨在保障设备间数据与信息的安全流通,为电力行业安全、高效、快速发展提供重要支撑与参考。以上方法存在系统兼容性与标准化以及管理效率等方面的不足,因此设计一种基于深度 Q-learning 算法的智能电网管控模型。该模型具有以下优势及特点。

1) 基于区块链的智能电网分布式 SDN 管控模型通过将 VC 和 DID 作为身份凭证,结合 SDN 控制器与动态信任评估算法,以及基于属性的访问控制策略,并利用区块链技术,为智能电网提供了一个安全、可信、灵活且可扩展的分布式 SDN 管控方案。这个模型能够实时评估应用程序的信任值,动态调整访问权限,确保智能电网的安全性和可靠性。

2) 基于区块链的分布式 SDN 网络优化方案则侧重于在资源分配、网络拓扑动态变化以及安全威胁不断演变的情况下,通过 SDN 控制器的集中管理和自动化配置,实现网络资源的优化分配和网络的动态优化。同时,结合区块链技术的不可篡改性和分布式特性,该方案能够动态调整安全策略,快速响应网络中的异常行为和安全事件,提高网络的安全性和防御能力。

1 智能电网管控模型设计

1.1 分布式 SDN 管控模型设计

将 VC 和 DID 作为应用程序身份凭证与 SDN 控制器,结合动态信任评估算法与基于属性的访问控制(attribute based access control,ABAC)策略,提出基于区块链的智能电网分布式 SDN 管控模型。该模型的成员如下。

1) 访问实体。由两部分构成,首先在域 A 中设置一个 SDN 控制器(命名为 A_Ctrl),通过唯一标识符 DID_A 来识别。在 A_Ctrl 内嵌入与 DID_A 相关联的公私钥对及其信任验证证书 TVC_A,用于记录并维护 A_Ctrl 的信任评分。其次,在区域 A 中部署 SDN 应用(A_App),由 PID 唯一标识,并通过 A_Ctrl 管理。在 A_App 内存储由 A_Ctrl 颁发的属性证书 A_AttrCert,用于定义其在访问过程中可使用的属性集。需要注意的是,当 A_App 作为访问实体时,需提交其所在区域的 DID 及信任证书作为访问评估的参考,并作为 ABAC 策略信息源运作[8]。

2) 访问客体。指的是在其他域设置的一组持有资源的 SDN 控制器(B_Ctrl),通过 DID_B 识别,在 B_Ctrl 内嵌入与 DID_B 关联的公私钥对及信任验证证书 TVC_B,用于存储 B_Ctrl 的信任评分。同时 B_Ctrl 作为 ABAC 的策略执行点与策略信息点提供功能。

3) 验证者。为已加入信任域区块链网络的成员(K_Ctrl),通过多个 DID_K 区分,在每个成员内嵌入与 DID_K 相关联的公私钥对[9]。验证者校验访问实体与访问目标提供的凭证,将验证结果提交至访问控制核心请求处理,然后将处理结果通过共识机制记录于区块链上。同时,验证者承担 ABAC 策略管理点的角色。

4) 访问控制引擎。由两个模块组成,一是信誉评估模块,通过检索区块链上的连接信息,获取访问主体/访问客体的相关数据,应用动态信任评估算法,计算双方的信誉值。动态信任评估算法的信誉值计算公式为

$$R = W_{SC} \times S_H + W_B S_{SC} + W_T T_E \tag{1}$$

式中,S_H 是指合规行为得分,S_{SC} 是指违规行为得分,T_E 是指第三方评价得分,W_{SC}、W_B、W_T 分别是 S_H、S_{SC}、T_E 的权重[10]。

二是属性访问判定模块,基于访问主体提供的属性证书,通过函数 Φ(A_AttrCert,ReqAttrs)转换,利用 ABAC 判断访问主体有权访问的数据范围。最终,访问控制引擎将两个模块的输出返回给验证者,同时作为 ABAC 策略决策点运作。

5)区块链。多个区域的 SDN 控制器构成区块链,为访问控制模型提供信息检索、存储及验证服务,确保数据的完整性和不可篡改性。

1.2 分布式 SDN 网络优化

在资源分配、网络拓扑动态变化以及安全威胁不断演变的情况下,基于深度 Q-learning 实施基于区块链的分布式 SDN 网络的优化。具体的分布式 SDN 网络优化流程如下。

1) 初始化阶段

初始化经验回放池 D,用于存储训练过程中产生的经验(状态、动作、奖励、下一状态)。设定抽取经验的批量大小,用于每次从经验回放池中随机抽取经验实施训练[11]。并初始化动作值函数的权重,即深度神经网络的参数。

2) 学习阶段

对于每个训练回合,从 1 到 N 执行以下步骤。

步骤 1 初始化网络拓扑图 G 和初始状态 u_t。

步骤 2 对于每个时隙 t 从 1 到 T。

1) 根据 ε - 贪婪策略选择动作 b_t。

$$b_t = \mathrm{argmax}_b Q(u,b;\vartheta) \tag{2}$$

式中,b 是指动作空间中的一个动作;u 是指当前的状态;ϑ 是指 Q 值函数的参数。

以 $1-\varepsilon$ 的概率选择具有最高 Q 值的动作:使用深度神经网络,输入当前状态 u_t,输出所有可能动作的 Q 值,这些 Q 值表示在给定状态下执行每个动作所预期的长期回报。从深度神经网络的输出中,找到具有最高 Q 值的动作,这个动作代表了在当前状态下,根据深度神经网络的预测,最有可能带来最大长期回报的动作。生成一个介于 0~1 之间的随机数,并比较其与探索率 ε[12]。如果随机数小于 ε,则随机选择一个动作作为 u_t。如果随机数大于或等于 ε,则选择上一个步骤中找到的具有最高 Q 值

的动作作为 u_t。

2)执行动作 u_t, 包括选择出块周期、分配不同交换机的 slave 控制器以及边缘服务器。

3)对于每个控制器 c 从 1 到 C。

当满足下式

$$\begin{cases}\zeta^{\mathrm{P}} > \zeta^{\mathrm{STD}} \\ s_c > s^{\mathrm{STD}}\end{cases} \tag{3}$$

式中, ζ^{P} 是指控制器服务强度, ζ^{STD} 是指控制器服务强度的标准差, s^{STD} 是指控制器损失率的标准差, s_c 是指单个控制器 c 的损失率。

首先根据交换机传输的电力业务类型,删除不符合带宽要求的链路,生成新的子图 G'。对于每个链路 $(i, j) \in G$, 若其带宽 $B(i, j)$ 不满足传输的电力业务类型所需的最低带宽要求 B_{req}, 则将该链路从图中删除。

$$(i, j) \notin G',\ B(i, j) < B_{\mathrm{req}} \tag{4}$$

接着在子图 G' 上,使用 Dijkstra 算法计算从源节点 s 到目标节点 t 的最佳传输路径 $P(s,t)$, 并更新路径对应的丢包率、最小带宽以及迁移成本[13]。

最佳传输路径的计算公式为

$$P(s,t) = \arg\min C(P) \tag{5}$$

式中, $C(P)$ 是指路径 P 的成本。

最佳传输路径对应的丢包率的更新公式为

$$L(P(s,t)) = \prod_{(i,i)\in P(s,t)} w(i,i) \cdot L(i,i) \tag{6}$$

式中, $w(i,i)$ 是指链路 (i,i) 的权重; $L(i,i)$ 是指链路 (i,i) 的丢包率。

最佳传输路径对应的最小带宽的更新公式为

$$B_{\min}(P(s,t)) = \min_{(i,i)\in P(s,t)} B(i,i) \tag{7}$$

最佳传输路径对应的迁移成本的更新公式为

$$M(P(s,t)) = \sum_{(i,i)\in P(s,t)} M(i,i) + \sum_{n\in P(s,t)} M_n \tag{8}$$

式中, $M(i,i)$ 是指链路 (i,i) 的成本, M_n 是指节点 n 的处理成本[14]。

4) 更新奖励 z_t 和下一状态 u_{t+1}, 并将经验 (u_t, b_t, z_t, u_{t+1}) 存储到经验回放池 D 中。

5) 依据智能电网窃电检测混合模型[15-16],从经验回放池 D 中随机抽取一个小批量经验 (u_j, b_j, z_j, u_{j+1})。

6)根据目标网络更新目标值为

$$\beta_j = \begin{cases} z_j, \text{ stop from } u_{j+1} \\ z_j + \varphi Q(u_{j+1}, \arg\max_{b'}(Q(u_{j+1}, b', \vartheta'), \vartheta')), \text{ 其他} \end{cases} \tag{9}$$

式中, φ 是指折扣因子, b' 是指在下一个状态下采取的动作, ϑ' 是指目标网络的参数。

3) 结束循环

结束循环,完成算法执行,实现分布式 SDN 网络优化。

2 实验分析

2.1 实验设置

通过 python 内的 networkx 实现基于区块链的智能电网分布式 SDN 管控模型的网络拓扑结构。分别搭建两种网络拓扑结构-星型拓扑结构与网状拓扑结构,测试两种结构下的迁移成本,测试结果如图 1 所示。

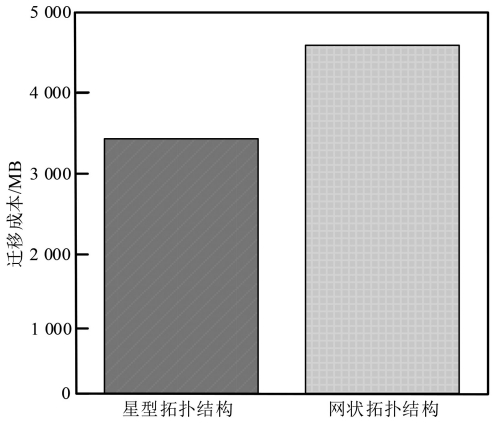

图 1 两种结构下的迁移成本测试结果

Fig. 1 Test results of migration cost under two structures

星型拓扑结构的迁移成本较低,因此选择星型拓扑结构作为基于区块链的智能电网分布式 SDN 管控模型的网络拓扑结构。

利用 tensorflow 实现深度 Q-learning 算法。实验过程中的参数设置情况见表 1。

表 1 参数设置情况

Tab. 1 Parameter settings

参数	项目	参数设置情况
SDN 域	数量	8 个
各域	初始信誉值	0.3
诚信域	数量	6 个
恶意域	数量	2 个
恶意访问次数	阈值	3
总访问	次数	30

在实验过程中,访问行为涵盖 3 种类型。由可信区域触发的正确访问、操作失误导致的错误访问,以及由恶意区域发起的访问。针对恶意区域与可信区域,分别实施初始化。根据参数设置,若某区域的综合信任评估值低于 0.3,则重置其信誉值为初始状态;若某区域发生恶意访问达 3 次,则将其信誉值设为 0,并列入黑名单,剥夺其与其他区域交互的权限。实验中,各区域随机选择其他区域实施访问。

2.2 实验过程

利用搭建的基于区块链的智能电网分布式 SDN 管控模型实施实验区域智能电网的管控。部分管控结果如表 2 所示。

符合 ABAC 策略的访问请求都得到了批准,而不符合策略的访问请求都被拒绝。这验证了该管控模型的有效性。为持续优化模型,利用基于深度 Q-learning 实施优化。在优化中,参数设置情况具体如下,经验回放池 D 容量为 10^6 条经验;抽取经验的批量大小为 64;训练回合数

为 1 000;时隙数 T 为 100;探索率 ε 初始值为 1.0,随着训练的进行逐渐减小(每 100 个回合减小 0.1),直至达到一个较小值(0.01);折扣因子为 099;优化器为 SGD。

表 2　管控结果

Tab. 2　Control results

访问主体	访问要求	是否符合 ABAC 策略	访问请求结果
1	读取数据	是	批准
2	写入数据	否	拒绝
3	执行操作 X	是	批准
4	访问特定资源	是	批准
5	删除数据	否	拒绝

2.3　测试项目与结果分析

首先测试深度 Q-learning 算法优化前后的累积奖励,测试结果如图 2 所示。

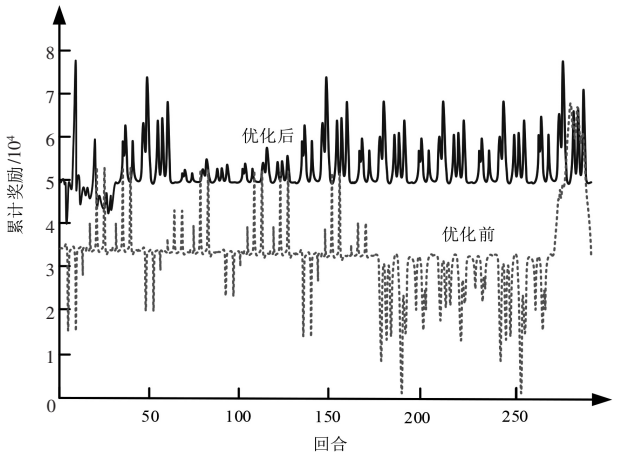

图 2　深度 Q-learning 算法优化前后的累积奖励测试结果

Fig. 2　Cumulative reward test results before and after optimization of deep Q-learning algorithm

由图 2 可知,优化后累积奖励明显大幅增加。这是由于智能电网环境复杂多变,而基于深度 Q-learning 算法优化基于区块链的智能电网分布式 SDN 管控模型后,能够帮助模型处理这种复杂环境,通过训练学习到适应不同情况的策略。因此优化后的模型累积奖励高,具有较强的适应能力,能够在不同场景下保持较好的性能。然而累积奖励高不仅反映了优化后的管控模型在短期内的性能表现,更体现了其长期效益。通过深度 Q-learning 算法的优化,模型有望在长期运行中实现更低的成本、更高的效率和更好的可持续性。

分析设计方法的安全性能,并测试设计方法的命中率。在安全性能与命中率的测试中,将文献[4-6]作为对比方法。四种方法的安全性能测试结果如表 3 所示。

根据表 3 中 4 种方法的安全性能测试结果,设计方法自主身份控制,增加了智能电网系统的灵活性和安全性。设计方法的去中心化设计有助于减少单点故障,提高模型的抗攻击能力。其动态性意味着模型能够适应变化的环境和条件,及时调整策略以应对新的威胁。设计方法的节点安全得到保障,意味着模型能够防止未经授权的访问。其域安全得到保障,确保了不同域之间的数据交换和访问是安全的。其可追溯性使得模型中的所有操作都可以被记录和追踪,有助于审计和调查。综合以上分析,设计方法在多种安全性能方面表现出色,提供了全面的安全保障。而其他 3 种方法在不同程度上存在某些安全性能方面的不足。命中率测试结果如图 3 所示。

表 3　4 种方法的安全性能测试结果

Tab. 3　Safety performance test results of four methods

方法	自主身份控制	去中心化	动态性	节点安全	域安全	可追溯
设计方法	是	是	是	是	是	是
文献[4]	否	否	否	是	是	是
文献[5]	否	否	是	是	是	是
文献[6]	是	是	否	否	是	是

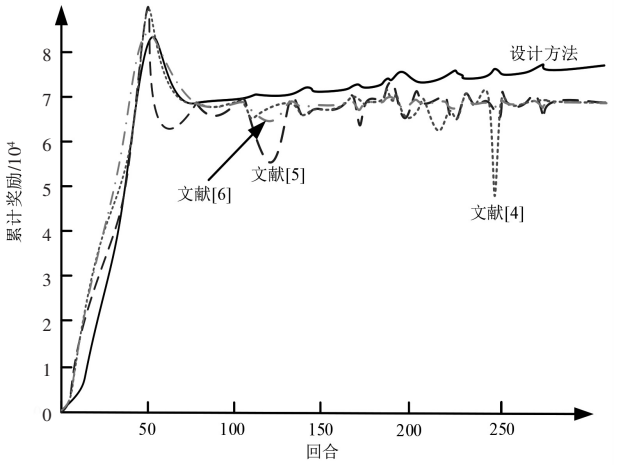

图 3　命中率测试结果

Fig. 3　Hit ratio test results

根据图 3 可知,3 不同的对比方法仅限于网络域间的访问控制来确保信息安全,然而,它们缺乏在域内识别并移除恶意用户节点以及对管理恶意域的能力,导致域内的不良节点能够持续活动,限制了这些策略所能达到的命中率,使其维持在较低水平。设计方法不仅具备识别并清除恶意用户节点及整个恶意域的功能,而且对待访问实体表现出一定的灵活性。尽管该方法在初期可能需要更长的观察期,且命中率的提升相对缓慢,但它能彻底清除恶意域,确保网络环境的安全,从而使命中率稳步提升至一个较高水平并持续保持增长态势。因此,该设计方案相较于文献[4-6]方法,展现出显著优势,实现了更高且稳定的命中率。

3　结论

在探索智能电网管控模型的进程中,深入应用了深度 Q-learning 算法,这不仅革新了传统管控策略,还为电力系统的智能化管理开辟了新路径。通过深度强化学习的强大能力,模型能够自适应地优化调度决策,有效应对电网运行中的复杂多变情况。

(下转第 142 页)