网络安全垂域大模型在安全左移中的应用

姜校一*
杭州安恒信息技术股份有限公司　浙江　杭州　310000

摘　要：网络安全威胁升级、攻击手段复杂，传统安全防护模式依赖事后响应，难适应多变安全需求，安全左移将安全措施前置到开发设计阶段，是提升防御能力关键，网络安全领域大模型有自然语言处理和知识推理能力，为安全左移带来创新方案。本文剖析网络安全领域大模型技术特性，结合安全左移核心需求，探讨大模型在漏洞挖掘、安全策略制定、代码审计等场景应用，构建基于大模型的安全左移体系，可早期识别安全风险、主动防御，降低安全事件概率与处置成本。实践显示该模型应用让安全漏洞发现效率超40%，增强企业安全防护前瞻性与有效性。

关键词：网络安全垂域大模型；安全左移；漏洞挖掘；代码审计；主动防御

引言

近年数字化转型加快，企业网络环境复杂，勒索软件、数据泄露等安全事件不断出现，传统安全防护侧重网络边界防御与事后响应，被动防御模式有响应慢、修复成本高问题。安全左移把安全流程提前到软件开发等早期阶段，在需求分析等环节嵌入安全机制，从根源降低安全风险，但传统安全左移实践有不少难题：人工分析难涵盖复杂代码逻辑，安全策略制定不精确，漏洞识别效率低且误报多。网络安全垂域大模型是人工智能在网络安全领域的深度应用，可处理海量安全数据、理解复杂知识、进行智能决策，能有效补足传统安全左移手段缺陷，探究大模型在安全左移中的应用，对提升企业安全防护水平、构建主动安全防御体系意义重大。

1 网络安全垂域大模型与安全左移理论基础

1.1 网络安全垂域大模型技术特性

数据层作为大模型运行根基，承担多源异构安全数据整合重任，实际作业时，它采集众多历史漏洞报告，记载软件、系统过往漏洞详情，像漏洞类别、影响程度、修复措施等，CVE数据库便存有全球海量漏洞信息。安全日志同样关键，完整记录网络系统用户登录、数据访问、系统配置变更等各类操作事件，为模型分析供应丰富信息，网络攻击案例不可或缺，分析真实攻击可让模型掌握攻击者手法策略。各类权威安全标准规范，如ISO 27001、NIST网络安全框架等，为模型提供可靠规则指引，有大型企业的数据层整合内部多年积累的网络入侵、数据泄露等安全事件记录，融合公开行业数据，打造出庞大完备数据集，筑牢模型训练基础。

模型层采用Transformer架构，核心是自注意力机制，能让模型在处理安全文本数据时，动态留意不同位置信息，实现深度语义理解，训练时先靠海量安全语料进行无监督预训练，此时模型学习大量未标注的安全文本数据，自动提取通用特征与模式，掌握网络安全领域基础语言表述和知识架构。随后进入有监督微调阶段，针对漏洞分类、攻击意图识别等特定安全任务，利用标注好的数据集优化模型参数，就像训练漏洞分类模型，将已知漏洞数据按SQL注入、跨站脚本攻击等不同类型标注，使模型学会精准判断漏洞类别，提升在实际应用中的准确度。应用层据模型层安全分析结果为用户提供智能决策支持，网络安全垂域大模型三大核心能力，安全知识推理依已知漏洞信息构建图谱与算法推导风险，像分析某软件组件历史漏洞，结合功能、环境及趋势预测同类新漏洞；自然语言交互让用户以自然语言对话，问“如何防范DDoS攻击”模型给防护策略；多模态处理可分析文本、代码、流量等数据，分析代码读片段识漏洞，分析流量判DDoS或恶意软件传播等异常模式。

1.2 安全左移的核心内涵与目标

需求分析阶段，安全左移要对业务需求做全面安全风险识别，像在线支付系统，安全团队和业务团队合作明确数据保护、身份验证、交易安全等安全需求，确定数据加密方式与强度，选择多因素认证等合适身份认证机制，制定交易安全规则防欺诈盗刷。这些安全需求写入需求规格说明书，作后续设计开发依据，架构设计阶段，威胁建模是安全左移重要实践，分析系统架构识别潜在攻击面与风险点，优化架构设计。微服务架构设计里，为降低系统内横向渗透风险，隔离关键服务，用网络隔离技术把核心业务服务部署在独立安全区域，设立严格访问控制策略，保障授权服务和用户才可访问，还合理规划服务调用关系，避免权限集中，提升系统整体安全性。

用静态代码分析工具自动检查已编写代码，识别缓

冲区溢出、SQL注入、跨站脚本攻击等潜在安全漏洞，动态代码检测在代码运行时模拟输入场景，监测运行行为，发现内存泄漏、资源竞争等运行时安全漏洞。编码阶段及时发现修复漏洞，可大幅降低系统上线后安全风险，测试阶段引入渗透测试、模糊测试模拟真实攻击，验证系统安全防护能力，渗透测试由专业人员用攻击工具技术突破系统防线，发现漏洞薄弱环节[1]。模糊测试向系统输入大量随机畸形数据，观察响应检测异常输入下的安全问题，安全左移目标有两方面，一是上线前尽量识别消除安全隐患，降低安全事件发生概率，研究显示开发早期修复安全漏洞成本是上线后1/100，凸显安全左移降成本优势，二是早期风险识别修复减少后期安全运维工作量与成本，让运维人员将更多精力用于监控系统、应对新威胁，提升安全管理效率效益。

2 网络安全垂域大模型在安全左移中的应用场景

2.1 智能漏洞挖掘与风险预测

拿Python来说，其动态类型和灵活语法在开发中易现安全漏洞，网络安全垂域大模型分析Python代码语法结构、变量使用、函数调用等，结合安全编码规则库，可自动检测潜在SQL注入漏洞。代码若将用户输入直接拼接到SQL查询语句，模型能识别这种不安全编码并提示SQL注入风险，对于Java，大模型能识别不安全序列化操作，Java对象序列化时，反序列化逻辑处理不当，攻击者可用特制序列化数据执行任意代码，大模型分析Java代码中序列化相关类和方法，能发现这类隐患。和传统静态代码分析工具比，网络安全垂域大模型优势明显。传统工具基于简单规则匹配，易产生大量误报，大模型靠对代码上下文语义的理解，能更准确判断代码是否存在安全漏洞，某企业引入大模型审计代码后，漏洞误报率从35%降至12%，提高了审计效率准确性，节省了安全人员排查误报时间精力。

在风险预测上，网络安全垂域大模型依据历史漏洞数据和当下系统配置信息，经复杂关联分析算法，能预测将来可能出现的安全风险。模型结合系统所用开源组件版本信息，分析该版本有无已知漏洞，以及漏洞在当前系统运行环境下被利用的可能性[2]。不少开源软件发布新版本会修复旧版本安全漏洞，可部分企业因各种原因未及时更新到最新版，进而面临安全风险，大模型能根据开源组件更新日志、安全公告和企业系统实际运行环境参数，如操作系统版本、网络拓扑结构等，评估企业继续用旧版本开源组件的风险程度，模型还留意网络攻击趋势和行业安全情报。分析全球网络攻击数据，知晓当前流行攻击手段技术以及对企业系统的威胁，若当前网络出现针对某类特定应用程序的新型攻击趋势，大模型结合企业内部系统有无相关应用程序，以及应用程序配置使用情况，预判企业系统受此类攻击的可能性并及时预警，某金融机构用网络安全垂域大模型做风险预测，分析其核心业务系统所用开源中间件版本，预测到中间件存在未公开但已被部分黑客组织关注的潜在漏洞。基于模型预测结果，该金融机构提前采取防护措施，如加固中间件、设置额外访问控制策略等，成功避免可能致客户数据泄露的严重攻击事件，保护了金融机构声誉和客户利益。

2.2 安全策略自动化生成与优化

在访问控制策略制定时，模型分析用户角色、资源属性和操作行为，生成符合最小权限原则的策略，企业内部系统里，不同用户角色工作职责和权限需求不同。普通员工只需访问工作相关文件和应用程序，像销售部门员工要访问CRM系统客户信息和销售数据，无需碰财务部门敏感财务数据，模型分析用户角色会考虑其在组织架构位置、业务功能及历史操作记录等。对于资源属性，模型识别资源类型（文件、数据库、网络服务等）、敏感程度（公开、内部、机密等）和所属业务模块，综合分析用户角色和资源属性，模型为每个用户角色生成权限列表，保证用户只访问工作所需资源，执行相符操作，大型企业办公自动化系统中，网络安全垂域大模型为不同部门员工生成个性化访问控制策略[3]。研发部门员工能访问代码仓库、开发工具等资源进行编写、测试操作；人力资源部门员工只能访问员工信息管理系统做档案管理、考勤记录等操作。实施模型生成的访问控制策略，企业有效降低权限滥用引发的安全风险。

在加密策略制定上，模型依据数据敏感级别和传输环境特点，推荐最适配的加密算法与密钥管理方案，金融交易数据含客户资金信息、交易记录等高度敏感内容，一旦泄露会给客户和金融机构带来巨大损失。网络安全垂域大模型考虑金融交易数据的极高敏感程度，以及互联网环境传输时面临网络窃听、中间人攻击等安全威胁，基于这些模型建议采用AES-256加密算法，此算法加密性能强，能有效保护数据传输和存储时的保密性。为保障密钥安全管理，模型推荐借助硬件安全模块（HSM）进行密钥生成、存储与管理，HSM提供物理安全防护，避免密钥被窃取或篡改。

大模型能优化现有安全策略，模拟攻击场景评估策略有效性，找出策略冲突或防护薄弱之处，某大型互联网企业用大模型优化防火墙策略，模型先全面分析企业现有的防火墙规则，接着模拟端口扫描、恶意IP访问等多

种攻击场景，检测现有规则抵御这些攻击的防护能力[4]。经分析发现，企业防火墙规则存在大量冗余和冲突的规则，这既增加了规则管理的复杂度，还可能影响防火墙性能。依据模型分析结果，企业优化了防火墙策略，删除冗余规则，调整冲突规则，让策略规则数量减少30%，同时安全防护效果大幅提升，有效降低了策略管理复杂度与潜在风险。

3 网络安全垂域大模型应用的挑战与应对策略

3.1 技术实现与数据安全挑战

网络安全垂域大模型应用存在技术与数据安全两方面挑战，技术上模型训练要大量计算资源和高质量数据，网络安全领域数据专业性强、更新快，数据标注难且成本高。网络安全数据含众多专业术语和复杂技术细节，对标注人员专业知识要求高，准确标注需耗费大量时间精力。而且模型的可解释性问题限制其在安全决策中的运用，大模型多基于复杂神经网络结构，内部决策过程难直观理解，用户不明白模型输出结果的依据，关键安全决策场景中，用户对模型信任度受影响，怕模型判断失误。

数据安全领域，训练数据含诸多敏感信息，如企业内部漏洞数据、用户隐私信息等，存在泄露风险，敏感数据泄露，企业会遭受经济损失，声誉受损。模型也易遭攻击，恶意攻击者利用投毒、对抗样本等方式操控模型输出，引发错误安全决策。投毒攻击攻击者向训练数据注入恶意数据，使模型习得错误模式；对抗样本攻击构造特殊输入数据，让模型输出错误结果实现攻击。

3.2 应用落地与管理优化策略

技术上运用联邦学习、迁移学习减少大规模标注数据依赖，实现跨机构协同训练，联邦学习让多机构不共享原始数据却能共同训练模型，本地训练后上传参数聚合，既保护隐私又利用数据提升泛化能力，迁移学习将一任务领域知识用于相关任务，降低标注数据需求[5]。引入可解释技术增强模型决策透明度，基于规则推理把决策过程转化为规则助用户理解结论得出方式，可视化分析用图表直观展示决策依据，数据安全方面，用同态加密、差分隐私保护训练数据隐私，同态加密实现密文计算保护隐私，差分隐私加噪声保护个体信息。还需对模型安全加固，借对抗训练提升抵御恶意攻击能力，管理上建立严格数据访问控制与审计机制，明确权限记录使用确保合规，制定模型评估标准和安全规范，界定应用边界明确责任防滥用，开展人员安全培训，提升员工对大模型安全风险认知与防范能力，保障正确使用及时应对潜在问题。

结语

网络安全垂域大模型为安全左移带来创新技术路径，借智能漏洞挖掘、安全策略自动生成等应用，大幅提升安全风险早期识别与主动防御水平。虽技术实现及数据安全存在难题，不过伴随技术演进和管理机制健全，大模型于安全左移中价值将愈发凸显。日后，网络安全垂域大模型要进一步整合多模态数据处理、强化学习等技术，增强对复杂安全场景的适配能力。

参考文献

[1]林云志.基于垂域大模型的轨道交通智能化初探[J].电气化铁道,2024,35(06):59-63.

[2]叶剑飞,王晓周,邓攀科,等.落实安全左移,应关注软件设计过程[J].中国信息安全,2024,(07):38-41.

[3]江苏海门聚力打造县域网络安全“四位一体”防护体系[J].中国网信,2024,(06):88-89.

[4]王荣海,蔡水捷.从安全左移到安全内生,中国银联安全研发体系建设实践[J].中国金融电脑,2023,(07):41-44.

[5]陈克豪.安全左移场景下的软件成分分析评估与改进方法研究[D].浙江理工大学,2023.