ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 40,No. 3,May 2025,pp. 832－844
DOI:10. 16337/j. 1004-9037. 2025. 03. 021

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

基于攻击流量和漏洞驱动的态势感知评估方法

李　岩，王梓莹，冒佳明，顾智敏，姜海涛

（国网江苏省电力有限公司电力科学研究院，南京 211103）

摘　要：网络安全态势评估在网络防御策略实施环节扮演重要角色。现有的态势评估方法汇聚攻防双方信息构建评估模型，对于攻击检测的准确性和攻击和漏洞利用关系极为敏感。为了应对上述挑战进而提升评估准确性，本文提出了融合攻击和漏洞的态势评估方法。该方法利用多样的攻击数据集训练攻击检测模型，借助于集成学习的思路实现不同模型攻击检测结果的融合。借助于开源的安全大模型提取不同攻击类型与安全漏洞之间的利用关系知识，计算攻击成功利用漏洞的概率，综合攻击危害程度和攻击成功率获得安全态势评估结果。在基准数据集上进行验证，结果表明提出的攻击检测方法提升了攻击检测性能，平均F_1-score达到96.24%，进一步地结合攻击检测结果给出了态势评估应用案例，验证了本文方法的有效性。

关键词：态势评估；攻击检测；安全漏洞；大安全模型

中图分类号：TP391　　**文献标志码：**A

Situation Awareness Assessment Approach Based on Attack Traffic and System Vulnerabilities

LI Yan, WANG Ziying, MAO Jiaming, GU Zhimin, JIANG Haitao

(Electric Power Research Institute, State Grid Jiangsu Electric Power Co., Ltd., Nanjing 211103, China)

Abstract: Network security situation assessment plays an important role in the design and implementation of network defense strategies. The existing situation assessment methods gather the information of both attack and defense to construct an assessment model, which is extremely sensitive to the accuracy of attack detection and the relationship between attack and vulnerability exploitation. To deal with the above challenges and improve the accuracy of assessment, this paper proposes a situation assessment method combining attack and vulnerability. Firstly, various attack data sets are used to train attack detection models, and the attack detection results of different models are fused by the idea of ensemble learning. Secondly, with the help of the open source security model, the exploitation knowledge between different attack types and security vulnerabilities is extracted. Finally, the security situation assessment results are obtained by integrating the degree of attack damage and the probability of vulnerability exploitation calculated using the extracted exploitation knowledge. The results show that the proposed method improves the performance of attack detection, and the average F_1-score reaches 96.24. Furthermore, combined with the attack detection results, a situation assessment application case is given to show the

基金项目：国网江苏省电力有限公司科技项目(J2023180)。
收稿日期：2024-06-24；**修订日期：**2024-08-31

effectiveness of the proposed method.

Key words: situation assessment; attack detection; security vulnerability; large security model

引　言

随着IT系统复杂性的提升以及攻击者攻击能力的加强,快速准确地获知系统的网络安全状况从而提高安全应对的时间提前量成为主动安全防御方向的重要问题之一。借助于网络安全态势感知技术,及时获取并加工分析网络中的各种安全态势要素信息,并综合得出网络的整体安全状况及可能的网络安全变化趋势,是有效解决这一问题的手段[1],因此网络安全态势评估成为网络安全领域的重点关注,仍是目前的研究热点。现有的态势评估方法通常汇聚攻防两方的信息,借助于基于数学模型[2-3]、知识推理[4-5]、机器学习[6-7]和博弈模型[8-9]等评估方法得到综合的态势评价结果。在攻击方信息利用方面,现有的态势评估方法从流量数据、告警日志等出发建立攻击检测模型识别攻击类型,进而量化攻击危害。因此,态势评估方法对于攻击检测模型的准确性是敏感的。然而,现有的攻击检测通常建立在单一数据集上,但单一数据集覆盖的攻击类型有限,而不同数据集的特征维度和数量各不相同,无法构建一个统一的数据集用于攻击检测模型训练,导致攻击检测效果不佳,进而影响了安全态势评估的准确性。此外,采用的检测模型对于特征是敏感的,模型与特征之间的适配性不佳可能导致攻击检测准确性下降,同样也会影响安全态势评估的准确性。另一方面,从防守方资产存在的安全漏洞角度出发,美国基础设施顾问委员会开发的通用安全漏洞评分系统(Common vulnerability scoring system, CVSS)为所有安全漏洞的威胁程度提供一个开放、通用、合理的量化评级标准,较好地解决了评估过程中对于安全漏洞的量化问题。然而,以往的态势评估忽略了攻击对于漏洞利用的可能性,仅根据攻击发生的潜在危害程度来评估态势,这也会影响安全态势评估的准确性。

为了解决上述问题,本文提出了一种融合攻击和漏洞的态势评估方法,旨在提升安全态势评估的准确性。本文的方法利用多样的攻击数据集训练攻击检测模型,在训练过程中通过模型与特征适配性度量选择合适的模型实现攻击检测,并借助于集成学习的思路实现不同模型攻击检测结果的融合,解决攻击检测准确性问题。此外,借助于开源的安全大模型SecGPT(https://github.com/cloud-tera/secgpt)提取不同攻击类型与安全漏洞之间的关系,计算攻击成功利用漏洞的理论概率,综合攻击危害程度和攻击成功率获得安全态势评估结果,解决态势评估准确性低的问题。

1　相关工作

自1999年首次提出网络态势感知(Cyberspace situation awareness, CSA)概念以来,研究人员已经利用来自不同来源的态势相关数据开发了多样化的网络安全态势评估模型。鉴于本文采用攻防对抗环境下综合利用攻击侧和防御侧信息进行目标系统态势评估的思路,重点梳理遵循该思路的现有研究工作进展情况。

从攻击侧信息利用视角出发,可供态势评估利用的信息主要来源于采集的实时网络流量、日志数据。随着机器学习、深度学习等人工智能技术的发展,借助其处理海量数据的优势,基于数据驱动方式的流量挖掘和日志挖掘已经成为了有效利用攻击侧信息进行网络态势评估的主流解决方案[6-7,10-12]。文献[10]构建基于BP神经网络的安全态势评估模型,通过建立完整评价指标体系,实现对电网安全态势评估,但由于BP神经网络存在“过拟合”的固有缺陷,导致评估结果精度不理想。针对这一问题,文献[11]将深度学习算法应用于态势评估中,提高了结果准确性。然而,借助于人工智能技术,虽然处理效率明显提高,但是在缺少样本,尤其是缺少标记样本的情形下,仍会出现评估准确度下降的问题。因

此,在考虑攻击侧信息前提下,综合利用防御侧信息可以进一步提升态势评估的准确性。

从防御侧信息利用视角出发,可供态势评估利用的信息主要来源于目标系统的网络拓扑、漏洞信息和威胁情报等。文献[13-14]提出了一种基于吸收马尔可夫链模型的漏洞评估方法,在建立和完善漏洞评价指标的基础上对漏洞的可利用性得分和影响力进行度量,进而采用图论和马尔科夫链相结合的方法对节点进行评估。该方法虽然提升了节点级态势评估的准确性,但是忽略了目标系统的网络拓扑结构等全局信息,导致从节点级推理目标系统态势值的方法不够合理。针对上述问题,文献[15]提出基于拓扑关系和漏洞信息的网络安全态势感知模型,通过对拓扑漏洞进行分析,建立网络安全态势理解有限状态机,综合计算多种态势组成值计算目标系统的网络安全态势值。文献[16]提出了一种从局部到整体、从微观到宏观的细粒度的网络威胁态势评估方法。综合考虑漏洞利用概率以及目标系统资产价值,从漏洞节点出发逐渐对全网安全态势进行评估。

虽然网络安全态势评估方法已经有了长足进步,但是面对具有复杂关系的、态势数据庞大且多样化的目标系统,单一的从攻击侧或是从防御侧都无法准确地度量和反映出目标系统的安全状态,因此研究综合攻防双方态势数据的评估是可行的改进方法。

2 本文方法

2.1 态势评估方法框架

本文提出的态势评估方法框架如图1所示,分为离线的攻击检测模型训练阶段和实时的攻击检测阶段。该方法从攻击产生的流量和目标系统存在的安全漏洞出发实现对目标系统的态势评估。

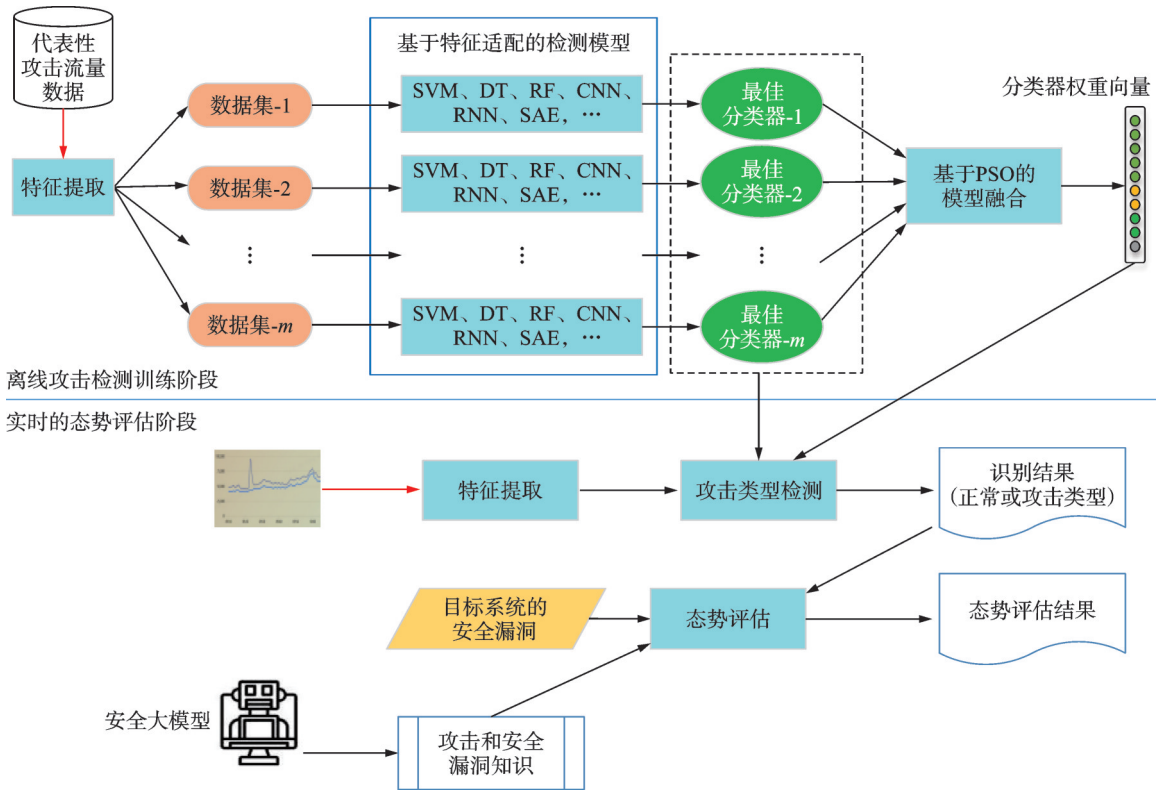

图1 态势评估方法框架

Fig.1 Framework for situation assessment

在离线阶段,利用多样化的攻击或入侵检测数据集训练攻击检测模型,使得能覆盖更多的攻击类型,而不是受限于特定的攻击数据集。攻击或入侵检测数据集不仅提供了关于攻击类型的知识,更重要的是其构建的特征集为原始流量数据特征提取提供了经验知识。同时,通过模型与流量特征的适配性选择Top-k模型进行训练,最后在训练得到的一组最佳攻击检测模型后采用了粒子群优化算法(Particle swarm optimization,PSO)[17]完成攻击检测结果的自动融合,提升攻击检测精度。在实时的态势评估阶段,对特定时间窗口的流量数据进行流量特征提取以适配不同检测模型,通过训练好的一组模型进行攻击类型预测,并集成方式确定最终攻击类型。在此基础上,考虑目标系统存在安全漏洞,结合识别的攻击类型及其危害程度,计算目标系统的安全态势值。

2.2　流量特征提取

在攻击侧,态势评估方法立足于从流量特征角度捕获目标系统正遭受或即将遭受的网络攻击。在离线的攻击检测模型训练阶段,为了提升检测模型性能,采用多个代表性的攻击流量数据进行模型训练。由于攻击流量数据的原始抓包Pcap文件并不都可用,导致无法从原始流量数据中直接进行特征提取。因此,攻击检测模块采用n个已完成特征提取的不同训练数据集$D=\{D_1,D_2,\cdots,D_n\}$进行攻击检测模型训练,这些数据集对应的特征集记为$\{F^1,F^2,\cdots,F^n\}$。然而,对于训练数据集中的任意两个数据集而言,对应的特征集F^x和F^y不完全相同,$1\leqslant x,y\leqslant n$,即$F^x\cap F^y\neq\varnothing$。例如,两个常用的攻击数据集KDD CUP99[18]和UNSW-NB15[19],分别包含41个特征和48个特征,且两者的特征集交集不为空。对于有不同特征的数据集,采用主成分分析方法实现特征降维。在实时态势评估阶段,流量特征提取需要根据训练数据集的流量特征从实时流量数据中提取对应的流量特征信息,为攻击检测模型提供适配后的输入。

2.3　攻击检测

攻击检测旨在从特征化的流量数据中识别特定的攻击类型,本质上是一个多元分类问题。对于特定的训练数据集,不同的检测模型表现出的性能各异。同一个检测模型在不同训练数据集上同样表现不尽相同。为了提升攻击检测的准确性,通过检测模型与特征的适配度量来选取模型,建立面向不同特征集的检测模型。

首先,给出模型与特征集适配性度量。假设存在m个不同的攻击类型检测模型或分类器可以用于从特征化的流量数据中识别攻击类型,记为C。在检测模型训练阶段,仅用数据集的训练样本进行模型训练,一旦训练完成则采用验证数据集中的样本进行测试,通过测试评价模型与特征集适配性。给定任意的数据集$D_i\in D$,有$D_i=D_{i1}\cup D_{i2}$,且$D_{i1}\cap D_{i2}=\varnothing$,这里$D_{i1}$和$D_{i2}$分别表示数据集$D_i$中划分得到的不相交的训练集和验证集。采用多元分类领域最常用的平均macro F_1作为适配性度量,即有

$$F_{1\text{macro}}(c,i)=\frac{1}{q_i}\sum_{j=1}^{q_i}\frac{2\cdot\text{TP}_j}{2\cdot\text{TP}_j+\text{FN}_j+\text{FP}_j}\tag{1}$$

式中:c表示特定的检测模型,$c\in C$;q_i表示数据集D_i提供的攻击类别数,包括Normal类别;TP_j、FP_j、FN_j分别为度量第j个类别的不同识别结果的统计结果。基于上述适配性度量结果,为数据集D_i选择macro F_1最高的K个模型作为最佳检测模型,记为$\{c_{i_1}(D_i),c_{i_2}(D_i),\cdots,c_{i_K}(D_i)\}$,$c_{ij}(D_i)\in C$。然后为$D$中的每个数据集分别选取$K$个最佳检测模型得到最佳模型集,用于攻击类型识别。

2.4　多模型融合的攻击类型

多模型融合模块旨在综合多个模型的攻击类型识别结果,确定最终的攻击类型,用于态势评估。为此,将$m\times K$个攻击类型检测模型的识别结果以置信度加权方式确定最终的攻击类型。为了在模型训练过程中从训练集上搜索这些最佳检测模型集的最优权重,引入了粒子群优化算法对权重进行调整优化。

在本模型的权重计算应用中，将权重计算过程视为一个优化问题，PSO受飞鸟集群活动规律性的启发，将优化问题看作是觅食的鸟群，粒子群中的每一个粒子即为优化问题的潜在解，将潜在解看作是搜索空间中的一只鸟，鸟所寻找的食物即是问题的最优解，粒子群中的每一个粒子的优劣程度由优化问题的适应性函数来衡量。多模型融合的PSO权重计算方法如下：

（1）初始化：初始化大小为$m \times K$的一维权重向量$\boldsymbol{W}$，每个权重分量对应一个检测模型。

（2）更新：通过迭代计算对权重向量进行调整和优化，寻找最优解。每一次迭代，粒子依据个体极值（pbest）和全局极值（gbest）实现更新，个体极值为粒子自身经历过的最佳位置，全局极值为整个粒子群经历过的最佳位置。粒子速度和位置迭代过程的计算分别为

$$V_i(l+1)=\alpha V_i(l)+s_1 r_{1i}(\text{pbest}_i-G_i(l))+s_2 r_{2i}(\text{gbest}-G_i(l)) \tag{2}$$

$$G_i(l+1)=G_i(l)+V_i(l+1) \tag{3}$$

式中：l表示迭代轮次；V_i和G_i分别代表第i个粒子的运动速度和所在位置；α为惯性系数，数值越大表示越不易改变之前的运动路线；s_1为个体加速因子，数值越大表示粒子越倾向于选择自身经历过的最佳位置，s_2为社会加速因子，数值越大表示粒子越倾向于选择整个粒子群经历过的最佳位置，s_1和s_2一般取值为2；r_{1i}和r_{2i}分别代表第i个粒子在寻求个体极值和全局极值方向上的随机性，取值为[0, 1]范围内的随机数；取gbest为迭代结束的最终结果。

（3）反馈：在迭代过程中需要通过适应性函数作为衡量某个粒子所在位置的优劣标准，这里采用分类准确率作为适应性函数，有

$$\text{fitness}=-\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{Z}\left[l_{ij}\log_{10}(s_{ij})+(1-l_{ij})\log_{10}(1-s_{ij})\right] \tag{4}$$

式中：l_{ij}代表第i个样本标签为j的真实值；s_{ij}表示标签j在样本i上的概率；N代表样本数；Z代表标签数。

PSO算法通过不断进行更新和反馈操作优化粒子位置，最后寻找到全局最优值gbest作为模型融合的权重向量$\boldsymbol{W}$。假设在数据集D上的类别数为h，则每个攻击检测模型检测结果可以表示为1个h维向量，所有m个攻击类型检测模型的识别结果表示为1个$m \times h$维的向量$\boldsymbol{Rst}$，因此$\boldsymbol{W} \times \boldsymbol{Rst}$即为最终的检测结果向量，取向量中值最大的类别作为最终的攻击类型。

2.5 基于安全大模型的攻击和安全漏洞知识获取

为了刻画攻击与安全漏洞之间的关系，同时在态势评估中量化攻击成功实施后的危害程度，同时避免以人工方式收集和梳理攻击与漏洞之间的关系，借助于安全大模型蕴含的丰富漏洞信息和攻击信息构建攻击与安全漏洞知识库。

攻击与安全漏洞知识库包含两大类实体：攻击类别和漏洞以及攻击类别对于漏洞的利用关系。若特定的攻击类别可利用特定漏洞，则CVSS V3的评分作为攻击危害程度的度量值，取值为0～10之间。利用大模型获取攻击和漏洞信息并构建攻击与安全漏洞知识库的方法分为3个步骤。首先，编制用于提取安全漏洞信息的提示Prompt模板，如图2(a)所示，提取的信

```
作为一个顶尖的网络安全专家，请以json格式列出你所知的
安全漏洞，格式为：漏洞名称: {漏洞名称}
漏洞描述: {漏洞描述}
洞编号: {漏洞编号，例如 CVE-XXXX-XXXX}
受影响的系统/软件: {受影响的系统或软件名称和版本}
CVVS评分值：{ CVVS V3评分值}
```

(a) Prompt template for extracting system security vulnerabilities

```
作为一个顶尖的网络安全专家，请以json格式列出所知的
网络攻击类型，格式为：攻击类别: {攻击类别}
攻击方法: {攻击方法描述}
知名的攻击工具: {知名的攻击工具}
成功攻陷带来的影响: {成功攻陷带来的影响}
```

(b) Prompt template for extracting network attack types

```
作为一个顶尖的网络安全专家，请问{攻击类别}能否
利用安全漏洞{漏洞编号}
实施攻击，直接输出结果：能或不能。
```

(c) Prompt template for extracting the relationship between network attacks and vulnerability exploitation

图2 攻击与安全漏洞知识库构建方法使用的提示模板

Fig.2 Prompt template for constructing attack and security vulnerability knowledge base

息包括漏洞名称、漏洞描述、漏洞编号、受影响的系统或软件和CVSS评分值;其次,编制用于提取攻击类别的提示Prompt模板,如图2(b)所示,提取的信息包括攻击类别、攻击方法、攻击工具和成功攻陷带来的影响;最后,编制用于提取攻击与安全漏洞利用关系的提示Prompt模板,如图2(c)所示,提取的信息是特定的攻击类型能否利用特定的安全漏洞。若漏洞能利用,则CVSS V3评分值作为攻击类别对漏洞利用关系的强度。

这里,采用网络安全大模型SecGPT来实现自动的攻击与漏洞利用知识抽取。SecGPT是在GitHub上开源的全球首个网络安全大模型,该模型支持网络安全知识问答任务,可以用于回答网络安全领域的问题,为安全从业者提供支持和解释。表1给出了利用安全大模型建立的攻击与安全漏洞的知识实例,在本文工作的态势评估环节利用该知识进行评估。

表1　攻击与安全漏洞利用关系及其危害程度

Table 1　Relationship between attack and vulnerability and their severities

攻击类别	漏洞ID	危害程度	攻击类别	漏洞ID	危害程度
DoS	CVE-2021-20094	7.5	Reconnaissance	CVE-2023-35009	5.3
U2R	CVE-2023-23410	7.8	Shellcode	CVE-2024-0581	5.8
R2L	CVE-2023-46128	6.5	Worm	CVE-2022-23241	8.1
Probe	CVE-2022-41040	8.8	Heartbleed	CVE-2014-0160	7.5
Analysis	CVE-2023-5157	7.5	DDoS	CVE-2023-22422	7.5
Backdoor	CVE-2022-30885	9.8	Web attack	CVE-2021-31760	8.8
Exploit	CVE-2020-6463	8.8	Infiltration	CVE-2017-9805	7.5
Fuzzers	CVE-2022-41911	7.5	Brute Force	CVE-2022-0953	6.1
Generic	CVE-2016-6329	5.9	Bot	CVE-2021-37522	9.8

2.6　态势度量

为了提升网络安全态势评估的准确性,针对目标系统存在安全漏洞,结合识别的攻击类型及其危害程度,计算目标系统的安全态势值(Security situation value,SSV)。

假设目标系统存在的漏洞集为$\mathrm{Vuls}=\{\mathrm{vul}_1,\mathrm{vul}_2,\cdots,\mathrm{vul}_V\}$,$V=|\mathrm{Vuls}|$,而目标系统可能遭受的攻击类型$\mathrm{AtkType}=\{\mathrm{atk}_1,\mathrm{atk}_2,\cdots,\mathrm{atk}_U\}$,$U=|\mathrm{AtkType}|$,$\mathrm{atk}=\langle\mathrm{atkId},\mathrm{threatVal}\rangle$。假设安全态势评估的时间粒度为$\Delta t$,不失一般性,考虑时间段$[t,t+\Delta t]$上的网络安全态势,该时间段上目标系统存在的$v$个漏洞,$\mathrm{vuls}=\{\mathrm{vul}_1,\mathrm{vul}_2,\cdots,\mathrm{vul}_v\}$,$\mathrm{vuls}\subset\mathrm{Vuls}$;同时,目标系统在该时间段上遭受的$u$种攻击,$\mathrm{atks}=\{\mathrm{atk}_1,\mathrm{atk}_2,\cdots,\mathrm{atk}_u\}$,且对于任意的攻击类型$\mathrm{atk}_i$,$\mathrm{atk}_i\in\mathrm{AtkType}$,该类型攻击发生的次数记为$\mathrm{count}(\mathrm{atk}_i)$,该类型攻击利用漏洞$\mathrm{vul}_j$,$\mathrm{vul}_j\in\mathrm{Vuls}$,成功实施攻击的概率记为$p_{ij}$。具体地,类型$\mathrm{atk}_i$的攻击成功概率为

$$p_i=\begin{cases}1 & I(\mathrm{atk}_i,\mathrm{vuls})=1\ \text{or}\ I(\mathrm{atk}_i,\varnothing)=1\\ 0 & \text{其他}\end{cases}\tag{5}$$

式中:$I(\cdot,\cdot)$为一个势函数,取值为1或0。当攻击可以利用vuls中任意漏洞或者攻击不依赖于漏洞可实施时,攻击成功概率为1,即对应$I(\mathrm{atk}_i,\mathrm{vuls})=1$,以及$I(\mathrm{atk}_i,\varnothing)=1$,否则攻击成功概率为0。

目标系统的安全态势值SSV取决于系统所遭受的攻击成功概率及其危害性大小,因此采用式(6)计算SSV。显然,SSV越大,目标系统的安全性越低。

$$\mathrm{SSV}_{[t,t+\Delta t]}=\sum_{i=1}^{u}p_i\times \mathrm{atk}_i.\mathrm{threatVal}\times \mathrm{count}(\mathrm{atk}_i) \tag{6}$$

式中：atk_i.threatVal表示在时间段$[t, t+\Delta t]$攻击检测模型发现的攻击atk_i的危害程度值；count(atk_i)表示某种攻击发生次数。

3 实　　验

3.1 实验设置

采用网络安全领域权威的数据集来训练攻击检测模型，选取的数据集包括：NSL-KDD、UNSW-NB15和CSE-CIC-IDS2018，详细信息如表2所示。这些数据集提供了用于训练和测试的独立样本集，在训练过程中使用完整的训练样本集。为了获得每个数据集上的最佳分类器和最佳分类器的权重向量，从测试样本集中随机选取一定比例的样本作为验证集用于评价分类性能。由于NSL-KDD测试样本集的规模远小于其他两个数据集，因此从NSL-KDD随机选取30%用作验证集，其余70%用作测试集，而对于UNSW-NB15和CSE-CIC-IDS2018，则以随机选取不放回方式选取10%的样本用作验证集，10%的样本用作测试集，且UNSW-NB15和CSE-CIC-IDS2018中取样的样本采用对应的原始流量数据进行验证和测试，而不是直接采用经过特征化的样本，其主要目的是获得基于PSO方法学习的最佳分类器权重。

表2 数据集详细信息

Table 2 Details of the used datasets

序号	名称	特征数/个	攻击类别	训练样本数/个	测试样本数/个
1	NSL-KDD	41	5,{DoS,U2R,R2L,Probe,Normal}	125 973	22 544
2	UNSW-NB15	49	9,{Analysis,Backdoor,DoS,Exploit,Fuzzers,Generic,Reconnaissance,Shellcode,Worm,Normal}	82 332	175 341
3	CSE-CIC-IDS2018	80	9,{Heartbleed,DDoS,DoS,Webattack,Infiltration,Brute Force,Bot,Normal}	1 619 351	404 836

本文采用了5种代表性的方法构建攻击检测模型，包括机器学习方法：高斯贝叶斯分类器Gaussian-NB[20]、随机森林(Random forest, RF)[21]、K最近邻算法(K-nearest neighbor, KNN)[22]；深度学习方法：卷积神经网络(Convolutional neural networks, CNN)[23]、长短期记忆网络(Long short-term memory, LSTM)[24]。选取CNN和LSTM的原因在于：CNN使用卷积核逐层提取复杂特征，能自动学习高维网络流量的特征，而LSTM通过对序列数据建模，学习可变长度输入序列的隐藏的序列特征。选取最好的$K=3$个模型用于融合，为了便于描述，将采用PSO融合后的模型称为PsoDet。同时选取了Griffin[25]和Whisper[26]算法作为对比算法，选取的动机在于它们是最近面向流量的攻击检测方法中的代表性方法。采用多分类方法评价的经典指标p_{macro}、r_{macro}和$F_{1\mathrm{macro}}$从宏观角度来评估攻击检测的结果，即有

$$p_{\mathrm{macro}}=\frac{1}{d}\sum_{i=1}^{d}\mathrm{precision}_i \tag{7}$$

$$r_{\mathrm{macro}}=\frac{1}{d}\sum_{i=1}^{d}\mathrm{recall}_i \tag{8}$$

$$F_{1\,\mathrm{macro}}=\frac{2\times p_{\mathrm{macro}}\times r_{\mathrm{macro}}}{p_{\mathrm{macro}}+r_{\mathrm{macro}}} \tag{9}$$

式中:d表示攻击类别数;$\text{precision}_i=\frac{\text{TP}_i}{\text{TP}_i+\text{FP}_i}$,表示第$i$个类别的精确率;$\text{recall}_i=\frac{\text{TP}_i}{\text{TP}_i+\text{FN}_i}$,表示第$i$个类别的召回率。

使用Python语言实现了态势评估方法,攻击检测模型构建涉及的机器学习算法采用了sklearn中对应的开源实现,通过主成分分析方法进行特征提取。基于CNN模型的网络由输入层、卷积层、池化层、全连接层与输出层按顺序连接构成,其中卷积层和池化层构成隐藏层,隐藏层个数为2,卷积层中采用ReLU函数作为激活函数,学习率为0.005,drop-connect的大小为0.1,第1层和第2层的卷积滤波器的数量分别为32和64,最大池化长度为2。LSTM模型网络由3层结构组成:输入层、隐藏层和输出层,输入层的神经元数量与特征数量保持一致,隐藏层的个数为1~4层,每层包含4~64个记忆块,每一个记忆块包含1个记忆单元,且输入层的神经元与隐藏层的记忆块之间是全连接,输出层的神经元数量与类别数量保持一致。学习率控制在0.01~0.05,epoch次数为300次,batch大小为32,采用ADAM优化器和分类交叉熵作为损失函数。通过对超参数的调整确定了针对不同数据集的最佳模型。所有的实验均在CPU为Intel至强E5-2678v3 @ 2.50 GHz,内存128 GB,外存8 TB的Ubuntu服务器上完成。

3.2　攻击类型检测评价

本文方法实现攻击类型检测实用性从性能和效率两个方面进行评价。首先,评估不同测试数据集下攻击类型检测模型的性能,如表3所示。表3中给出了多分类情形下的平均检测结果,可以得到如下观察。首先,以$F_{1\,\text{macro}}$为评价标准,不同数据集上最佳的K个分类器不尽相同。具体地,RF,LSTM和CNN分别在NSL-KDD、UNSW-NB15和CSE-CIC-IDS2018上表现最佳。其次,融合后的模型在所有数据集上都取得了最佳表现,特别是在数据集NSL-KDD上提升了约5%。上述结果表明融合模型有助于提升攻击检测精度。

表3　离线阶段训练好的检测模型性能

Table 3　Performance of the offline trained detection models

方　法	NSL-KDD			UNSW-NB15			CSE-CIC-IDS2018		
	p_{macro}	r_{macro}	$F_{1\,\text{macro}}$	p_{macro}	r_{macro}	$F_{1\,\text{macro}}$	p_{macro}	r_{macro}	$F_{1\,\text{macro}}$
RF	0.832 6	0.960 4	0.891 9	0.890 4	0.812 5	0.849 7	0.972 4	0.973 1	0.972 7
GNB	0.796 5	0.960 4	0.870 8	0.881 1	0.925 1	0.902 6	0.930 9	0.939 8	0.935 3
KNN	0.740 4	0.973 6	0.841 1	0.930 6	0.837 5	0.881 6	0.985 3	0.985 9	0.985 6
CNN	0.876 8	0.831 8	0.853 7	0.942 1	0.940 6	0.941 3	0.989 6	0.988 9	0.989 2
LSTM	0.887 1	0.868 5	0.877 7	0.944 6	0.944 5	0.944 5	0.939 4	0.948 5	0.943 9
Griffin	0.937 6	0.942 8	**0.940 2**	0.946 9	0.954 2	**0.950 5**	0.987 6	0.990 8	0.989 2
Whisper	0.923 1	0.932 3	0.927 7	0.939 8	0.945 3	0.942 5	0.990 3	0.990 5	**0.990 4**
PsoDet	0.941 2	0.940 0	**0.940 6**	0.956 8	0.954 6	**0.955 7**	0.990 9	0.990 8	**0.990 8**

其次,为了验证PSO对于提升模型性能的有效性,进行了消融实验。采用其他常用的方法替换PSO,考虑了投票法(Voting)、平均法(Averaging)和堆叠法(Stacking),构建了3个变体模型,实验结果如表4所示。从表中可以看出,采用PSO优化集成的检测方法相较于其他3种集成方式有比较明显的优势,在数据集NSL_KDD上F_1值最多能提升9.76%,这表明采用PSO作为模型集成方式的有效性。

最后,评估检测方法在应用于实时分析阶段的效率。将运行时间作为效率评价指标,根据态势评估方法框架可知,实时的检测时间取决于两个部分的时间开销:特征提取时间和检测时间。实验中采用CSE-CIC-IDS2018中2018-2-16的原始流量数据作为实验对象,将流量数据按照1,3,5和10 min粒

表4 评估PSO对于提升检测模型性能的影响

Table 4 Impact of PSO on the proposed detection model

方 法	NSL-KDD			UNSW-NB15			CSE-CIC-IDS2018		
	p_{macro}	r_{macro}	$F_{1\,\text{macro}}$	p_{macro}	r_{macro}	$F_{1\,\text{macro}}$	p_{macro}	r_{macro}	$F_{1\,\text{macro}}$
PsoDet	0.941 2	0.940 0	**0.940 6**	0.956 8	0.954 6	**0.955 7**	0.990 9	0.990 8	**0.990 8**
Voting	0.845 2	0.867 3	0.856 1	0.901 5	0.854 8	0.877 5	0.946 6	0.954 7	0.950 6
Averaging	0.846 1	0.839 9	0.843 0	0.908 1	0.892 4	0.900 2	0.845 8	0.958 9	0.952 3
Stacking	0.891 2	0.902 7	0.896 9	0.937 8	0.881 8	0.908 9	0.987 7	0.991 7	0.989 7

度进行流量划分，而后采用流量分析器提取数据集对应的特征，输入到训练好的检测模型完成模型检测。图3给出了不同粒度下完成原始流量解析、攻击检测的时间开销。图中的结果是不同粒度下的最大时间开销，主要原因在于不同时间段内的流量大小不尽相同，因此选取最糟糕的情形度量其效率。从图中可以看出，不同粒度下的总的时间开销远小于设定的时间粒度，能满足态势评估中实时检测要求。

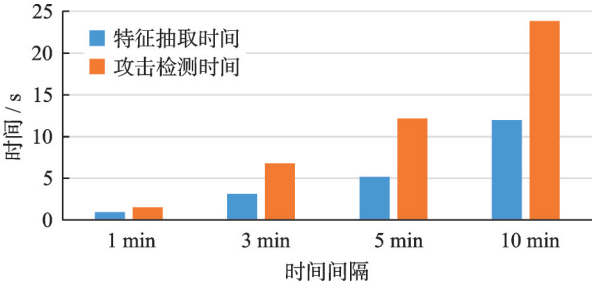

图3 实时检测的时间开销

Fig.3 Time cost for online detection

4 态势分析应用与评估

4.1 应用场景

为了验证态势感知评估模型的合理性，在给定的目标系统和攻击场景下进行了应用。假设目标系统由4个计算节点、1个态势感知节点和1台路由器构成，其网络拓扑图如图4所示。目标系统部署了常用的服务，包括web服务、文件服务、邮件服务和数据库服务，态势感知服务器通过镜像方式获取流量数据，应用训练好的攻击检测模型识别攻击类型，综合攻击和漏洞评估安全态势。

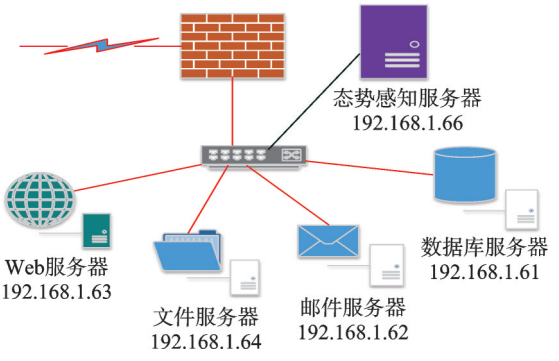

图4 目标系统的拓扑结构

Fig.4 Topology of the target system

根据表1设定目标系统包含所有漏洞，采用TCPReplay分别将CSE-CIC-IDS2018中3天(2018-2-14～2018-2-16)原始的数据包(Pcap文件)进行流量重放，通过路由器将流量随机重定向到任意一个计算节点，选取的流量持续24 h，在这段时间内，模拟攻击者不间断地对服务器进行各种类型的攻击，具体的攻击类型和持续时间如表5所示。态势评估的时间间隔设定为5 min，通过攻击检测模型识别攻击类型，并计算每一种攻击类型的攻击次数，通过式(6)得到安全态势值。

4.2 态势评估结果

图5给出了5 min粒度的目标系统的安全态势SSV随着时间变化曲线。结合表5中的攻击信息，可以看出，随着攻击包的出现和消失，目标系统的SSV发生了对应的改变，攻击出现时SSV值增大，反之则减少，反映出SSV能捕获目标系统遭受外部攻击。产生这一结果的主要原因在于目标系统的漏洞设

表5　流量中的攻击信息

Table 5　Attack information in traffics

序号	日期	攻击类别	开始时间	结束时间
1	2018-02-14	FTP-BruteForce	10:32	12:09
		SSH-Bruteforce	14:01	15:31
2	2018-02-15	DoS-GoldenEye	9:06	10:09
		DoS-Slowloris	10:59	11:40
3	2018-02-16	DoS-SlowHTTPTest	10:12	11:08
		DoS-Hulk	13:45	14:19

定采取了静态模式。进一步,目标系统的安全态势分解到构成目标系统的各个计算节点上。图6~8分别给出了不同时间段各个目标系统的安全态势SSV随着时间变化曲线。从图上可以得到相同的观察结果,这里不再重复阐述。

由于安全态势评估受到攻击检测精度的影响,因此进行了一组实验以评估攻击不同检测算法对于SSV的影响,其结果如图9所示。图中Baseline表示2018年2月16日的真实安全态势值,不同的攻击检测算法产生了不同的安全态势值,其中与Baseline方差最小的是CNN算法。不同检测算法的精度差异导致攻击类型预测结果的不同,而不同的攻击类型对应的危害程度不同进而导致了安全态势值的变化。

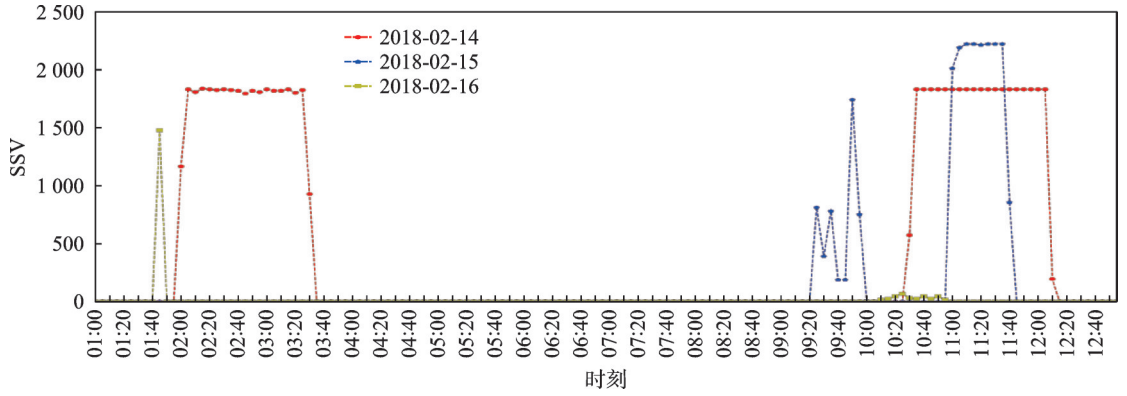

图5　目标系统的网络安全态势变化曲线

Fig.5　Curves of network security situation assessment for the target system

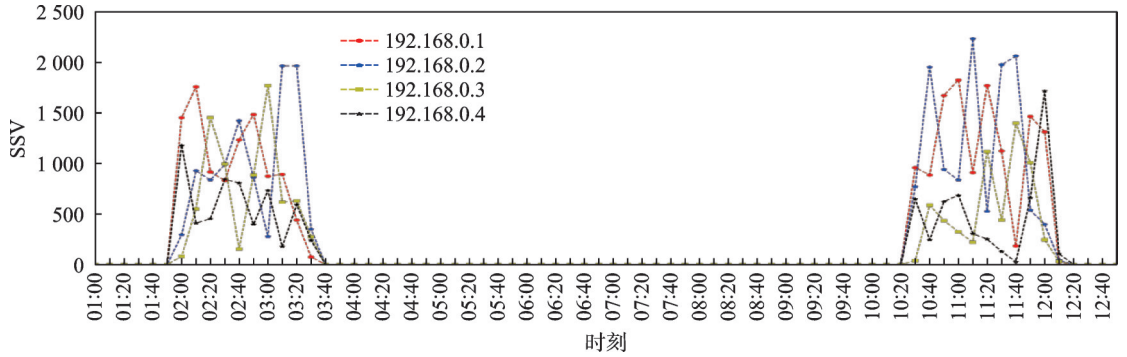

图6　2018年2月14日不同节点的网络安全态势变化曲线

Fig.6　Curves of network security situation assessment of multiple nodes on 2018/2/14

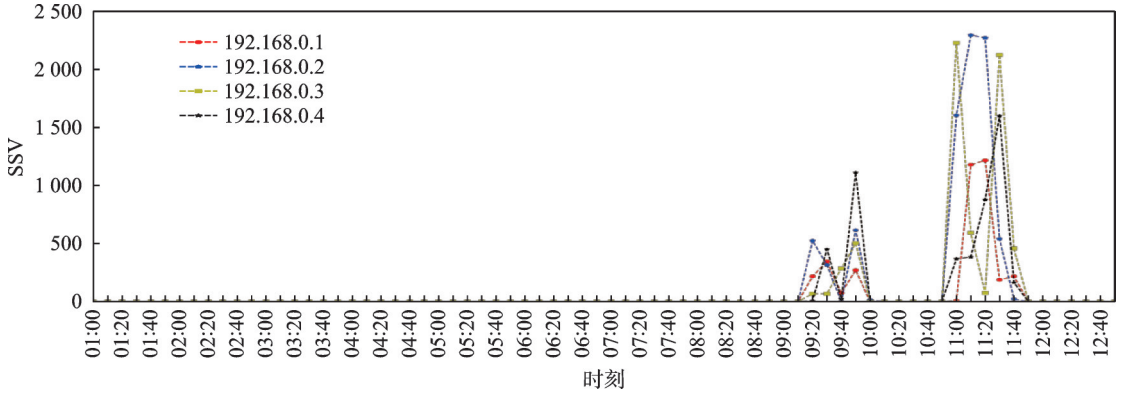

图7 2018年2月15日不同节点的网络安全态势变化曲线

Fig.7 Curves of network security situation assessment of multiple nodes on 2018/2/15

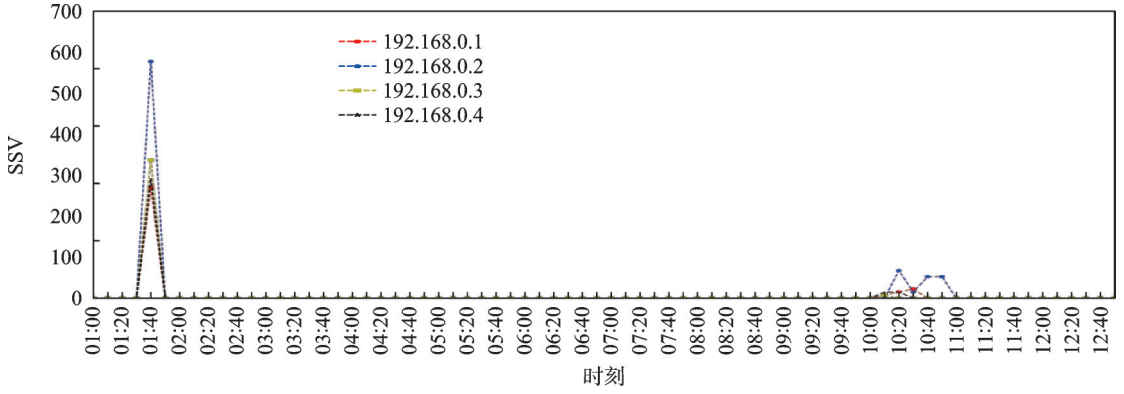

图8 2018年2月16日不同节点的网络安全态势变化曲线

Fig.8 Curves of network security situation assessment of multiple nodes on 2018/2/16

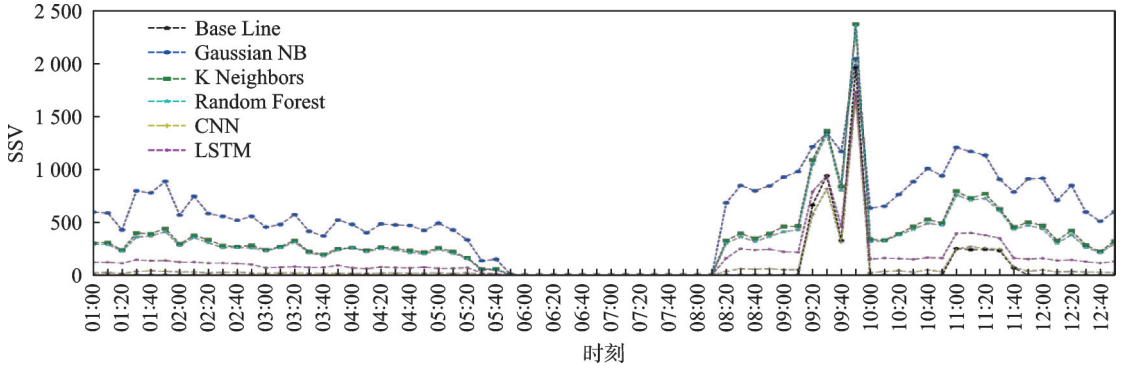

图9 2018年2月16日不同方法得到的网络安全态势变化曲线

Fig.9 Curves of network security situation assessment from various detection models on 2018/2/16

5 结束语

为了准确量化网络攻击导致的安全态势的变化，从网络攻击和安全漏洞两个角度入手，提出了一种综合网络攻击和安全漏洞信息的态势评估方法。一方面，本文提出的方法采用面向网络流量的攻击检测方法实现对网络攻击行为的感知，并区分网络攻击行为的不同类别，采用了包含多种攻击类型的

数据集进行模型训练提升了攻击类别的覆盖率,采用基于PSO的集成学习方法融合多个基攻击检测模型提升了攻击检测精度,使得平均F_1-score达到了约96%。另一方面,本文提出了态势计算方法借助于安全大模型提取攻击类别与安全漏洞相关的领域知识,进而利用领域知识计算攻击成功率和攻击危害程度,得到最终的安全态势值,准确地量化了安全态势随着网络攻击行为出现和消失的变化情况,且态势评估时间开销远低于评估间隔,能满足实用性要求。

参考文献:

[1] 龚俭,臧小东,苏琪,等. 网络安全态势感知综述[J]. 软件学报,2017,28(4): 1010-1026.
GONG Jian, ZANG Xiaodong, SU Qi, et al. Survey of network security situation awareness[J]. Journal of Software, 2017, 28(4): 1010-1026.

[2] 徐植, 陈俊,张智勇,等. 新型电力系统中基于人工免疫和隐马尔可夫的网络安全态势评估[J]. 华东师范大学学报(自然科学版),2023(5): 182-192.
XU Zhi, CHEN Jun, ZHANG Zhiyong, et al. Network security assessment based on hidden Markov and artificial immunization in new power systems[J]. Journal of East China Normal University (Natural Science), 2023(5): 182-192.

[3] 苏鹏涛,吴贶,陈孟婕,等. 基于隐马尔可夫模型的电力信息系统动态威胁定量分析[J]. 上海理工大学学报,2022,44(4): 388-396,416.
SU Pengtao, WU Kuang, CHEN Mengjie, et al. Dynamic threat quantitative analysis of power information system based on hidden Markov model[J]. Journal of University of Shanghai for Science and Technology, 2022,44(4): 388-396,416.

[4] PAPAMARTZIVANOS D, GOMEZ MARMOL F, KAMBOURAKIS G. Dendron: Genetic trees driven rule induction for network intrusion detection systems[J]. Future generations computer systems, 2018,79(2): 558-574.

[5] 徐兴硕,李世明. 基于证据推理算法的网络安全评估[J]. 哈尔滨师范大学自然科学学报,2023,39(1): 83-88.
XU Xingshuo, LI Shiming. Network security assessment based on evidential reasoning algorithm[J]. Natural Science Journal of Harbin Normal University, 2023,39(1): 83-88.

[6] 骆公志,侯若娴. 基于邻域量化容差条件熵增量式更新的网络入侵检测方法[J]. 数据采集与处理,2024,39(1): 181-192.
LUO Gongzhi, HOU Ruoxian. Network intrusion detection method based on incremental updating of neighbor-hood valued tolerance condition entropy[J]. Journal of Data Acquisition and Processing, 2024,39(1): 181-192.

[7] WU J, OTA K, DONG M, et al. Big data analysis-based security situational awareness for smart grid[J]. IEEE Transactions on Big Data, 2018,4(3): 408-417.

[8] 张勇,谭小彬,崔孝林,等. 基于Markov博弈模型的网络安全态势感知方法[J]. 软件学报, 2011,22(3): 495-508.
ZHANG Yong, TAN Xiaobin, CUI Xiaolin, et al. Network security situation awareness approach based on Markov game model[J]. Journal of Software, 2011,22(3): 495-508.

[9] 张红斌,尹彦,赵冬梅,等. 基于威胁情报的网络安全态势感知模型[J].通信学报, 2021,42(6): 182-194.
ZHANG Hongbin, YIN Yan, ZHAO Dongmei, et al. Network security situational awareness model based on threat intelligence[J]. Journal on Communications, 2021,42(6): 182-194.

[10] 陈维鹏, 敖志刚, 郭杰,等. 基于改进的BP神经网络的网络空间态势感知系统安全评估[J]. 计算机科学, 2018, 45(S2): 335-337,341.
CHEN Weipeng, AO Zhigang, GUO Jie, et al. Research on cyberspace situation awareness security assessment based on improved BP neural network[J]. Computer Science, 2018, 45(S2): 335-337,341.

[11] 于群,李浩,屈玉清. 基于深度学习的电网安全态势感知[J].科学技术与工程. 2019,19(35): 273-278.
YU Qun, LI Hao, QU Yuqing. Security situational awareness of power grid based on deep learning[J]. Science Technology and Engineering, 2019,19(35): 273-278.

[12] 杜静,敖富江,李鹏飞,等. 网络空间态势信息的特点及其知识表示方法[J]. 数据采集与处理,2019,34(3): 500-508.
DU Jing, AO Fujiang, LI Pengfei, et al. Characteristics and knowledge representation of cyberspace situation information[J]. Journal of Data Acquisition and Processing, 2019,34(3): 500-508.

[13] 胡浩,刘玉岭,张红旗,等. 基于吸收Markov链的网络入侵路径预测方法[J]. 计算机研究与发展,2018,55(4): 831-845.
HU Hao, LIU Yuling, ZHANG Hongqi, et al. Route prediction method for network intrusion using absorbing Markov chain[J]. Journal of Computer Research and Development, 2018,55(4): 831-845.

[14] GAO H, WANG S, ZHANG H, et al. Network security situation assessment method based on absorbing Markov chain[C]// Proceedings of International Conference on Networking and Network Applications. Urumqi, China: IEEE, 2022: 556-561.

[15] 李腾飞,李强,余祥,等. 基于拓扑漏洞分析的网络安全态势感知模型[J]. 计算机应用,2018,38(Z2): 157-163,169.

LI Tengfei, LI Qiang, YU Xiang, et al. Network security situational awareness model based on topological vulnerability analysis[J]. Journal of Computer Applications. 2018,38(Z2): 157-163,169.

[16] 刘世文,司成,张红旗. 一种细粒度的网络威胁态势评估方法[J]. 计算机工程与应用, 2018,54(10): 149-153.
LIU Shiwen, SI Cheng, ZHANG Hongqi. Fine-grained network threat situation assessment method[J]. Computer Engineering and Applications, 2018,54(10): 149-153.

[17] 郭春晓,苏旸. 一种新的基于量子进化策略的网络安全态势优化预测算法[J]. 小型微型计算机系统,2014,35(9): 2083-2087.
GUO Chunxiao, SU Yang. A new optimized algorithm based on quantum evolutionary strategy for network security situation prediction[J]. Journal of Chinese Computer Systems, 2014,35(9): 2083-2087.

[18] TAVALLAEE M, BAGHERI E, LU W, et al. A detailed analysis of the KDD CUP 99 data set[C]//Proceedings of IEEE Symposium on Computational Intelligence for Security and Defense Applications. Ottawa, ON, Canada: IEEE, 2009: 1-6.

[19] MOUSTAFA N, JILL S. UNSW-NB15: A comprehensive data set for network intrusion detection systems[C]// Proceedings of Military Communications and Information Systems Conference. Canberra, ACT, Australia: [s.n.], 2015.

[20] 顾兆军,刘婷婷,高冰,等. 基于GAN-Cross的工控系统类不平衡数据异常检测[J]. 信息网络安全,2022(8): 81-89.
GU Zhaojun, LIU Tingting, GAO Bing, et al. Anomaly detection of imbalanced data in industrial control system based on GAN-Cross[J]. Netinfo Security, 2022(8): 81-89.

[21] 刘振鹏,苏楠,秦益文,等. FS-CRF: 基于特征切分与级联随机森林的异常点检测模型[J]. 计算机科学, 2020,47(8): 185-188.
LIU Zhenpeng, SU Nan, QIN Yiwen, et al. FS-CRF: Outlier detection model based on feature segmentation and cascaded random forest[J]. Computer Science, 2020,47(8): 185-188.

[22] 吕成戍,王维国. 基于SVM-KNN的半监督托攻击检测方法[J]. 计算机工程与应用,2013(22): 7-10.
LYU Chengshu, WANG Weiguo. Semi-supervised shilling attacks detection method based on SVM-KNN[J]. Computer Engineering and Applications, 2013(22): 7-10.

[23] 王运兵, 姬少培, 查成超. 基于CNN与WRGRU的网络入侵检测模型[J]. 通信技术,2022,55(4): 486-492.
WANG Yunbing, JI Shaopei, ZHA Chengchao. Network intrusion detection model based on CNN and WRGRU[J]. Communication Technology, 2022,55(4): 486-492.

[24] CHAIBI N, ATMANI B, MOKADDEM M. Deep learning approaches to intrusion detection: A new performance of ANN and RNN on NSL-KDD[C]//Proceedings of the 1st International Conference on Intelligent Systems and Pattern Recognition. New York, USA: Association for Computing Machinery, 2020: 45-49.

[25] YANG Liyan, SONG Yubo, GAO Shang, et al. Griffin: Real-time network intrusion detection system via ensemble of autoencoder in SDN[J]. IEEE Transactions on Network and Service Management,2022, 19(3): 2269-2281.

[26] FU Chuanpu, LI Qi, SHEN Meng, et al. Frequency domain feature based robust malicious traffic detection[J]. IEEE/ACM Transactions on Networking, 2023,31(1): 452-467.

作者简介:

李岩(1990-),**通信作者**,男,高级工程师,研究方向:网络安全,E-mail: sdqfnqly2008@163.com。

王梓莹(1991-),女,工程师,研究方向:网络安全,E-mail:wangziying1008@126.com。

冒佳明(1991-),男,高级工程师,研究方向:网络安全,E-mail: hufimao@163.com。

顾智敏(1993-),女,工程师,研究方向:电力系统网络安全,E-mail: guzmdw@163.com。

姜海涛(1985-),男,高级工程师,博士,研究方向:电力系统网络安全,E-mail: jianghaitaoxin@163.com。

(编辑:刘彦东)