计算机科学
COMPUTER SCIENCE

基于字符特征的 DGA 域名检测方法研究综述

王宇, 王祖朝, 潘瑞

引用本文

王宇, 王祖朝, 潘瑞.基于字符特征的 DGA 域名检测方法研究综述[J]. 计算机科学, 2023, 50(8): 251-259.

WANG Yu, WANG Zuchao, PAN Rui. Survey of DGA Domain Name Detection Based on Character Feature [J]. Computer Science, 2023, 50(8): 251-259.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.220700277

基于字符特征的 DGA 域名检测方法研究综述

王　宇[1]　王祖朝[1]　潘　瑞[2]
1 中国地质大学(北京)数理学院　北京 100083
2 中国信息通信研究院　北京 100191
(18847163202@163.com)

摘　要　利用域名生成算法(Domain Generation Algorithm,DGA)可以生成大量的随机域名,近年来僵尸网络普遍使用 DGA 域名来增强隐蔽性。高效的检测 DGA 域名,对发现僵尸网络和保障网络信息安全具有重要意义。基于字符特征的 DGA 域名检测指仅利用域名的字符串完成检测,是一种实时检测方法,也是近年来对 DGA 域名检测研究的热点。对此类方法进行研究发现,使用传统机器学习和深度学习算法能够有效地检测 DGA 域名。但是对基于单词表的 DGA 域名、长度较短的 DGA 域名和新型 DGA 域名,还需要通过改进词嵌入方式、引入注意力机制或加入对抗样本等方法,来提高检测能力。最后对基于字符特征的 DGA 域名检测方法进行总结,分析不同检测方法的优点和存在的问题,提出了未来的研究方向和研究中需要解决的关键问题。

关键词:网络安全;DGA 域名检测;机器学习;深度学习;词嵌入;注意力机制;对抗样本
中图法分类号　TP393.08

Survey of DGA Domain Name Detection Based on Character Feature

WANG Yu[1],WANG Zuchao[1] and PAN Rui[2]
1 School of Science,China University of Geosciences(Beijing),Beijing 100083,China
2 China Academy of Information and Communications Technology,Beijing 100191,China

Abstract　Recent years have seen extensive adoption of domain generation algorithms(DGA) by botnets. Efficient detection of DGA domain name is of great significance for discovering botnets and ensuring cyber security. DGA domain name detection method based on character feature can complete the detection only by using the domain name string. It is a real-time detection method,and has become a hot spot in the research on DGA domain name detection. Research on such methods shows DGA domain name can be effectively detected by using traditional machine learning or deep learning. However,for wordlist-based DGA domain name,shorter-length DGA domain name,or new variant DGA domain name,it is still necessary to improve the detection ability by improving word embedding method,introducing attention mechanisms,or joining adversarial samples,etc. Finally,this paper summarizes the above methods,analyzes their advantages and existing problems,and proposes future research directions and key issues that need to be addressed for DGA domain name detection.

Keywords　Cyber security,DGA domain name detection,Machine learning,Deep learning,Word embedding,Attention mechanism,Adversarial example

1　引言

僵尸网络(Botnet)严重威胁到了网络安全[1]。僵尸网络指攻击者利用软件漏洞等方式,将恶意僵尸程序、蠕虫或病毒等植入目标主机(僵尸主机),并使用一对多的命令与控制(Command and Control,C&C)服务器,远程控制被感染的僵尸主机。僵尸主机通过 C&C 服务器获取到控制命令后,可进行分布式拒绝服务攻击(Distributed Denial of Service,DDoS)、发送垃圾邮件、特洛伊木马病毒或进行其他非法活动[2-3]。

Domain-Flux 机制抗干扰性强,近年来已经成为僵尸网络通信机制的主流发展方向。这种机制的原理是:攻击者使用域名生成算法(Domain Generation Algorithm,DGA)生成大量的随机域名,并注册一部分作为 C&C 服务器的域名。之后,僵尸主机对这些域名进行访问,逐个向域名系统(Domain Name System,DNS)服务器发送解析请求,只要其中一个或几个域名解析成功,僵尸主机便与 C&C 服务器通信成功。一旦该域名被发现,攻击者便注册下一个域名,保证 C&C 服务器对应的域名仍能解析成功。使用 Domain-Flux

到稿日期:2022-07-28　返修日期:2022-11-24
基金项目:国家自然科学基金(62071152)
This work was supported by the National Natural Science Foundation of China(62071152).
通信作者:潘瑞(panrui@caict.ac.cn)

机制，可以保持 C&C 服务器与僵尸主机的通信，也可以防止被安全系统发现和拦截，增强了僵尸网络的隐蔽性[4]。

在 Domain-Flux 机制中，利用 DGA 生成的域名称为 DGA 域名。文献[5]对 43 种僵尸网络进行了研究，其中有 23 种采用了 Domain-Flux 机制，这也表明了僵尸网络使用 DGA 域名的普遍性。因此，对 DGA 域名的有效检测，可以及时准确地发现僵尸网络，并支撑对僵尸网络的追踪和溯源，对保障网络信息安全具有重要意义。

目前 DGA 检测方法主要分为两类，一类是基于关联特征进行检测，这类方法需要获取域名系统的流量、IP 等信息，检测时需要耗费一定的资源和时间[6]；另一类是基于域名字符特征进行检测，这类方法仅需要输入域名字符，检测依赖少，能实时检测，且易于应用。因此，本文对基于字符特征的 DGA 域名检测方法进行了综述研究。

首先对 DGA 域名的生成方式、特点、种类和研究中使用的数据集进行了介绍。然后，对基于字符特征的 DGA 域名检测方法进行了研究。根据使用算法的不同，将其分为 3 类：基于传统机器学习的检测方法、基于深度学习的检测方法和基于附加机制的检测方法。最后，总结不同检测方法的特点，并提出未来的研究方向。

2 DGA 域名

DGA 域名具有存活时间短和种类繁多的特点，大部分 DGA 域名存活时间仅为 1～7 天。

DGA 使用公开获取到的随机种子作为输入，利用不同的算法生成二级域名，再和顶级域名拼接组成 DGA 域名，过程如图 1 所示。

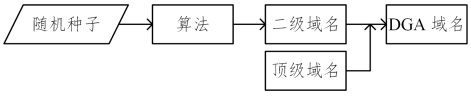


图 1 DGA 域名生成示意图

Fig. 1 Diagram of DGA domains generation

常用的随机种子很多，如时间、网络热词等，其分类如下[7]。

1)时间相关性(Time dependence)。与时间相关，如主机的系统时间、http 响应时间。

2)确定性(Determinism)。主流种子是确定的，基于确定种子生成的 DGA 域名是可预测的。但也有一些种子是不确定的，例如 Torpig 家族用 twitter 的关键词作为种子，这类 DGA 域名是不可预测的。

根据随机种子的不同，DGA 域名的分类如表 1 所列。

表 1 根据随机种子分类的 DGA 域名

Table 1 DGA domains sorted by random seed

DGA 域名分类	家族举例
静态可预测 DGA 域名	Kraken
动态可预测 DGA 域名	Conficker
动态不可预测 DGA 域名	Torpig
静态不可预测 DGA 域名	暂未发现

根据生成算法的不同，DGA 域名的分类如表 2 所列。

表 2 根据算法分类的 DGA 域名

Table 2 DGA domains sorted by algorithm

DGA 域名分类	算法特点	家族举例
基于算术的 DGA 域名	通过算术运算得到数值序列，可以根据 ASCII 码直接表示成域名，或者使用这些数值作为偏移量，指向 DGA 硬编码的字符表中的字符	banjori，conficker
基于哈希的 DGA 域名	通过哈希算法(MD5，SHA256 等)得到哈希值，用哈希值的十六进制表示域名	bamital，dyre
基于单词表 DGA 域名	从单词表中选取单词进行连接	matsnu，suppobox
基于置换的 DGA 域名	对初始域名置换得出所有可能的域名	volatilecedar

检测不同随机种子和算法生成的 DGA 域名家族，需要有相应的数据集支撑算法模型的训练，常用的数据集包括 DGA 域名数据集和正常域名数据集。数据集的质量和样本的平衡性对模型的训练结果和 DGA 域名家族的检测结果有很大的影响。

DGA 域名数据集目前主要有 3 个来源：1)美国 Bambenek Consulting 公司提供的 OSINT DGA 数据集，包含了 50 多个 DGA 域名家族，80 多万个恶意域名，需申请使用；2)德国 Fraunhofer FKIE 研究所提供的 DGArchive 数据集，包含了 60 多个 DGA 域名家族，需申请使用；3)我国 360 网络安全实验室提供的 DGA 数据集，包含 50 多个 DGA 域名家族，100 多万个域名，无需申请即可使用。

正常域名数据集中，最著名且研究中使用最多的是美国的 Alexa 数据集，Alexa 网站对全球域名的活跃情况进行了分析。研究者一般使用 Alexa 的 Top 100 万域名作为正常域名数据集，但是 Alexa 已在 2022 年 5 月 1 日关闭服务。此外，欧盟的 Tranco、美国的 Cisco Umbrella、我国的奇安信等都提供了百万级的活跃域名，它们都可以作为正常域名数据集。

3 基于传统机器学习的检测方法

基于传统机器学习的检测方法是将域名输入后，先做特征工程，再利用传统机器学习，包括无监督学习的算法和监督学习的算法，来完成检测。检测流程如图 2 所示。

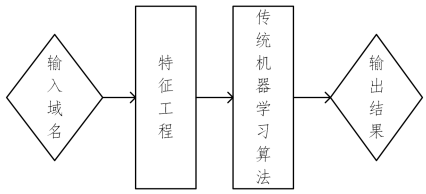


图 2 基于传统机器学习的检测方法的流程图

Fig. 2 Flow chart of detection method based on traditional machine learning

该检测方法中的特征工程主要利用领域知识来手工构建特征。手工特征的可解释性强，对检测结果与具体特征间的关联有较为直观的展现。

由于 DGA 域名是利用特定算法生成的，与正常域名相比，在结构、长度、随机性等方面存在差别。手工特征是根据这些差别设计的一系列特征，DGA 域名的检测效果也依赖于这些手工特征的设计。因此，在基于传统机器学习的检测方法中，域名手工特征的设计是关键，常用的手工特征如表 3 所列。

表 3　常用的手工特征

Table 3　Common handcrafted features

类别	特征名称
结构特征	字母个数
	数字个数
	域名长度
	特殊字符个数
	重复字符占比
语言特征	元音占比
	辅音占比
统计学特征	*N*-gram
	字符频率
	字符随机性

3.1　基于无监督学习的检测方法

基于无监督学习的检测方法以聚类算法为核心，设计手工特征，并根据距离度量将大量的域名划分为多个域名簇。检测时，计算待检测域名与各域名簇中心的距离，将其划分到距离最短的域名簇中，域名簇的类别就是该域名的类别。

Yadav 等[8-9]根据域名单字符与双字符的统计特征，使用 3 种距离度量：Kullback-Leibler 距离（KL）、编辑距离（ED）和 Jaccard 指数（JI）。Antonakakis 等设计了 Pleiades 系统[10]，先用聚类算法判定 DGA 域名，后用隐马尔可夫算法（Hidden Markov Model，HMM）进行分类。

3.2　基于监督学习的检测方法

监督学习算法需要使用标记好的样本，即输入正常域名和 DGA 域名。支持向量机（Support Vector Machine，SVM）的原理是寻找样本中最大间隔的超平面作为决策边界，将特征向量映射到空间形成样本点，找到一条线对样本进行分类，是一种应用广泛的分类算法。

文献［11］较为全面地比较了 SVM、C4.5 决策树、极限学习机（Extreme Learning Machine）、HMM、循环 SVM，在相同的 DGA 域名数据集上对检测效果进行了评估，结果表明循环 SVM（Recurrent SVM）取得了最好的效果。

但是，基于 SVM 的方法在训练时速度较慢，Huang 等[12]提出了一种基于灰狼优化算法（Grey Wolf Optimizer，GWO）的 SVM 模型，可以快速计算参数的最优解，进而提高检测速度。

随机森林的本质是对训练集样本和特征进行多角度筛选，因此有较强的抗干扰性和泛化能力。Lison 等[13]设计了域名的 28 个手工特征，使用随机森林算法，在 DGA 域名的二分类和多分类检测中，都有着不错的检测结果。

最近，文献［14］使用域名的 *N*-gram 特征结合几种机器学习算法对 DGA 域名进行分类。通过比较不同组合的结果发现，使用域名的 2-Gram 特征结合多层感知机算法（Multi-layer Perceptron，MLP）的检测效果优于使用 SVM 算法和 XGBoost 算法的检测方法。

表 4 列出了基于传统机器学习的检测方法的比较结果。基于无监督学习的方法不需要输入标记的数据集，能够实现自动分类，并且简单直观。基于有监督学习的方法中，手工特征结合 SVM 算法、随机森林和 MLP 算法是准确性较高的检测方法。但是，机器学习算法在训练时占用的内存较大且计算时间较长，使得系统维护困难，导致难以应用于大规模样本训练[15]。此外，做特征工程时，手工特征也对实验结果具有重要的影响，因此设计手工特征也需要一定的知识经验。

表 4　传统机器学习检测方法的比较

Table 4　Comparison of traditional machine learning detection methods

方法	文献	数据集	主要算法	关键结论
基于无监督学习的方法	［8-9］	从网络中自行抓取	聚类算法	不同距离度量的检测结果差异不大，最高可达 100% 的检测率和 2.5% 的误报率
	［10］	从 DNS 流量中过滤	聚类算法	15 个月中发现了 12 个 DGA 家族，一半是新发现的 DGA 域名
基于监督学习的方法	［11-13］	OSINT DGA 数据集、Alexa Top 100 万域名	随机森林、SVM、决策树、XGBoost、极限学习机	SVM、随机森林的检测准确率较高；基于灰狼优化算法的 SVM 模型检测结果准确率提高了 3.46%
	［14］	360 DGA 数据集、Alexa Top 100 万域名	SVM，MLP，XGBoost	2-Gram 特征结合 MLP 的检测效果最好，准确率为 95%，对未知 DGA 域名的检测准确率为 88.5%

4　基于深度学习的检测方法

DGA 域名也是一种短文本，因此可以将深度学习应用到 DGA 域名检测中。检测的过程如下：将域名字符输入后，经过词嵌入将文本字符转化为词向量，接下来用深度神经网络进行特征的提取，最后输出分类结果，这种方法的特点如图 3 所示。

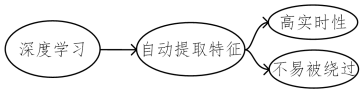


图 3　深度学习自动提取特征的特点

Fig. 3　Features of deep learning automatic feature extraction

1）不易被绕过：自动提取域名隐含的特征，不易被僵尸网络攻击者针对性设计。

2）高实时性：待测样本不需要做特征工程，预处理时间较短，检测速度比基于传统机器学习的方法更快，实时性更高。

4.1　基于循环神经网络的模型

循环神经网络（Recurrent Neural Network，RNN）擅长提取文本序列数据的信息，但是 RNN 会发生梯度爆炸或者消失，使它在学习时不稳定，难以捕捉到长期记忆，因此直接使用 RNN 检测 DGA 域名的效果不佳。

长短期记忆网络（Long Short-Term Memory，LSTM）是一种特殊的 RNN，引入门（Gate）机制用于控制特征的流入和损失，将短期记忆与长期记忆结合，一定程度解决了梯度消失的问题，在 DGA 域名检测中被广泛应用。

2016 年，Woodbridge 等[16]使用深度学习方法来进行

DGA 域名的检测。检测过程如图 4 所示，先将域名字符作为输入，经过词嵌入层生成字符级的词向量，之后由 LSTM 进行特征提取，最后使用逻辑回归进行分类。

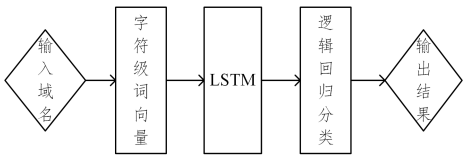


图 4 LSTM 检测的流程图

Fig. 4 Flow chart of LSTM

门控循环单元(Gated Recurrent Unit，GRU) 也是 RNN 的一种，是 LSTM 的一种变体。它的结构比 LSTM 更加简单，在训练时能节省很多时间。文献 [17] 提出了一种基于 GRU 的检测模型，并与基于 SVM、逻辑回归的检测方法进行了对比，基于 GRU 的检测模型在各项分类性能指标上都表现优异，而且模型收敛速度快，收敛过程平稳。

双向长短期记忆网络(Bi-directional LSTM) 是双向 LSTM 的组合，能够同时提取前后两个方向上的语义依赖关系，是对 LSTM 的改进。文献 [18]比较了几个不同 RNN 模型对 DGA 域名的检测效果。基于 Bi-LSTM 的模型准确率略高于基于 LSTM 的模型，基于 GRU 的模型虽然结构更简单，训练时间最短，但检测效果不如前两种。总体来说，3 个模型的检测能力差异不大。

在实际网络环境中，不同 DGA 域名家族的数量差距较大，因此许多 DGA 域名在训练数据集中的数量很少，容易出现样本不平衡问题。Tran 等[19] 为了缓解这一问题，引入代价敏感损失函数，提出了 LSTM-MI 模型。Chen 等[20] 提出了基于 LSTM 的样本重采样比例优化的方法，称为 LSTM-PQDO 模型。该模型在综合考虑样本特征和原始数量的基础上，围绕着初始解，在正确的方向上启发式寻找更好的解，用最优解迭代样本的重采样比例，从而实现了重采样比例的动态优化。LSTM-PQDO 模型减少了数据集不平衡带来的影响，与文献 [19]中的 LSTM-MI 模型相比，多种分类性能指标都有提升，检测效果更好。

4.2 基于卷积神经网络的模型

卷积神经网络(Convolutional Neural Network，CNN) 常用在自然语言处理、计算机视觉等领域[21]。Zhang 等[22] 在文本分类上进行了探索，证明了字符级词嵌入的 CNN 有良好的表现。这项研究对使用字符级词嵌入的 CNN 对 DGA 域名进行检测提供了研究方向。

Saxe 等[23] 提出了 eXpose 模型，使用字符级词嵌入，并利用 CNN 自动提取特征。Yu 等[24] 设计了 PCNN 模型，使用 3 个并行 CNN 同时提取 DGA 域名的 2-gram，3-gram，4-gram 的特征，检测效果较只用一个 CNN 的模型有所提升。

但是，CNN 只提取域名字符的局部特征这一特点，使整体模型欠缺全局特征的提取，后期研究在 CNN 的基础上进行改进。Zhou 等[25] 设计了一个基于时间的 CNN 模型提取域名的隐含特征，除提取域名局部特征之外，还加入了时间特征，提升了 DGA 域名的检测效果。

Yang 等[26] 在 CNN 上增加了提取更深层字符级特征的卷积分支，同时将提取域名的浅层和深层字符级特征融合，还引入了一种聚焦损失函数以解决样本不平衡导致检测率低的问题。最终提高了对 DGA 域名的检测准确率，尤其能够显著提升对部分样本较少的域名的检测准确率。

4.3 基于组合神经网络的模型

一些研究将 CNN 擅于提取局部特征的优势和 LSTM，GRU 擅于处理长期时序依赖信息的优势相结合。Zhou 等[27] 提出了 C-LSTM 模型，先用 CNN 提取局部特征，再用 LSTM 获取全局特征，证明了组合模型在文本特征提取上的有效性。

在 DGA 域名检测中，文献 [28-29]都用实验证明了使用 CNN-LSTM 模型进行域名字符特征的提取融合，检测效果比单独使用 CNN 和 LSTM 更好。文献 [30]分析比较了多种模型，发现采用 CNN 和 Bi-GRU 组合构建的 DGA 域名检测模型可获得较优的检测效果。

表 5 列出了基于深度学习的检测方法的比较结果。

表 5 深度学习检测方法的比较

Table 5 Comparison of deep learning detection methods

模型	文献	数据集	神经网络模型	关键结论
基于循环神经网络的模型	[16-18]	文献[16-24] 使用的数据集相同；OSINT DGA 数据集、Alexa Top 100 万域名	LSTM，GRU，Bi-LSTM	LSTM 模型对 90% 的 DGA 域名分类的假阳性率为 0.01；GRU 模型的检测效果略逊于 LSTM 模型，Bi-LSTM 模型的准确率略高于 LSTM 模型
	[19]		LSTM-MI	在多分类检测中的宏平均精确率和召回率比使用 LSTM 的模型提高了 7%
	[20]		LSTM-PQDO	宏观平均 F1 值增加了 4.58%～9.09%
基于卷积神经网络的模型	[23-24]		CNN，PCNN	CNN 模型对 DGA 域名的检测率假阳性率为 0.1%，PCNN 模型的检测效果优于 CNN 模型
	[25]	DGArchive 数据集、Alexa Top 100 万域名	基于时间的 CNN	加入时间特征的 CNN 模型 AUC 为 0.996 2
	[26]	360 DGA 数据集、Cisco Top 100 万域名	增加了更深层次的卷积分支的 CNN	对 20 种恶意域名的平均检测准确率为 97.62%，与 CNN 模型相比提高了 0.94%
基于组合神经网络的模型	[28-30]	360 DGA 数据集、Alexa Top 100 万域名	CNN-LSTM	检测准确率优于单独使用 CNN 和 LSTM
	[31]	OSINT DGA 数据集、Alexa Top 100 万域名	CapsNet	取得了与 CNN-LSTM 模型同样好的准确率，都在 99%以上，同时速度提升了一个量级

CNN 和 RNN 是两类常用的深度神经网络，在 DGA 域名的二分类和多分类检测上有着很高的准确率。一些文献

在此基础上提出了改进的模型，有效提高了 DGA 域名的检测效果。但是，仍存在样本不平衡问题和模型泛化能力弱等问题，文献[26]引入的聚焦损失函数对样本不平衡导致检测率低的问题有一定的改善。

组合神经网络结合了不同网络的优势，但是训练时间大幅度增加，如何加速训练过程是一个需要解决的问题。Berman 等[31]验证了基于胶囊网络(CapsNet) 的模型在 DGA 检测时取得了和 CNN-LSTM 模型同样好的准确性，同时速度提升了一个量级。这项研究为 CapsNet 在 DGA 检测中的应用奠定了基础。

5　基于附加机制的检测方法

虽然基于深度学习的检测方法已经取得不错的检测效果，但是对于一些特定 DGA 域名的检测效果不佳，主要包含以下 3 种，如图 5 所示。

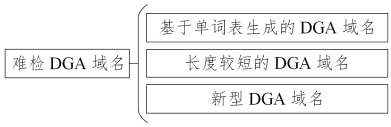


图 5　难检 DGA 域名

Fig. 5　Difficult detecting DGA domains

1)基于单词表生成的 DGA 域名

这类域名是从单词表中选取单词进行连接生成的，具有很高的可读性，与正常域名非常接近，因此很难识别。例如，Matsnu 家族包含动词列表及名词列表，在生成域名时，先从动词列表中随机抽取单词，然后在名词列表中随机抽取单词，组合后生成域名，如“charactermarry. com”。此外，常见的基于单词表生成的 DGA 域名还有 nymaim 家族、pizd 家族、suppobox 家族等。

在自然语言处理中，单词一般先经过 Word2Vec[32]，Glove[33]，Elmo[34]，GPT-3[35]，BERT[36] 等预训练生成词向量。域名字符是一种短文本，在 DGA 域名检测中，大多采用的是字符级词嵌入，由于没有经过单词的预训练，限制了域名字符蕴含的单词语义信息，因此对基于单词表生成的 DGA 域名的检测率低。

2)长度较短的 DGA 域名

Fu 等[37]在 2017 年提出了两种较难检测的 DGA 域名，被称为 SDGA 域名。它们使用 HMM 和概率上下文无关语法(Probabilistic Context-Free Grammar，PCFG) 生成，与正常域名相比，长度很短，特征不太明显，基于深度学习检测的方法在这种情况下可能会失去作用。而且，正常域名数量和 SDGA 域名数量差距过大，前者数量可以达到数千万，而后者不到十万，样本的不平衡也增加了分类难度。

3)新型 DGA 域名

传统的机器学习和深度学习在训练模型时会出现数据的时效性不足的问题。因为使用的数据都是已有的 DGA 域名，而之后出现的新型 DGA 域名在训练时缺乏相关的样本，在检测时较为困难。

此外，攻击者也会利用检测模型的一些弱点，有针对性地生成难以检测的域名，新型 DGA 域名经常出现在大规模的网络安全事件中。因此，提高对新型 DGA 域名的检测能力十分重要。

为了提升对难检 DGA 域名的检测能力，相关研究使用一些附加机制来提高检测准确率，并且已经取得不错的效果，主要有以下 3 种方法。

5.1　改进词嵌入方式

基于字符级词嵌入的 CNN 模型和 LSTM 模型对基于单词表生成的 DGA 域名表现不佳，为了解决这个问题，研究者改进了词嵌入方式，之后再利用深度神经网络进行特征提取和分类。

文献 [38] 先将域名在一个不相关的大型语料库上进行预训练，然后使用单词级词嵌入；文献 [39]使用基于字符和双字母组级别的混合词嵌入，以提高域名字符的信息利用度；文献 [40] 将人工特征与深度特征共同作为词嵌入；文献 [41] 增加了域名的语义信息，将原有字符特征和语义特征拼接后作为词嵌入。这些方式增强了对域名字符蕴含的单词语义信息的提取，对基于单词表生成的 DGA 域名的检测起到了良好效果。但是对于长度较短和新型 DGA 域名的检测，效果并不是很明显。

文献 [42] 从域名的元素组成和语义分析的角度提出了一种自适应词嵌入，将嵌入结果输入并行 CNN 进行特征提取，不仅提高了对基于单词表生成的 DGA 域名的检测效果，而且对长度较短的 SDGA 域名的检测效果也有所提高。

5.2　结合注意力机制

注意力机制可以理解为对不同部分分配权重，并给予不同的关注，被广泛应用于深度学习的各项任务中。Bahdanau 等[43]首次将注意力机制应用到自然语言处理领域中，是一项开创性进展。

对于 DGA 域名来说，不同位置字符的重要性也不同，例如，在 banjori 家族中，随机种子只改变域名的前 4 个字母(例如：earnestnessbiophysicalohax. com，kwtoestnessbiophysicalohax. com)，域名后面部分不变。因此，在检测时可以只关注后半部分，不需要关注整个域名。

文献 [44-45] 都利用 LSTM 结合注意力机制构建检测模型，利用 LSTM 提取域名特征的同时也考虑了 DGA 域名中不同位置的不同字符的权重，该模型具有良好的分类效果，比简单 LSTM 的模型有更高的分类精度。

文献 [46] 提出了一种注意力机制和 CNN 结合的 D3-SACNN 模型，使用 CNN 提取域名的局部特征，利用注意力机制获取字符和隐含特征之间的依赖性。文献 [47] 提出了 fast3DS 模型，使用并行 CNN 代替标准卷积层，利用轻量级全局平均池连接架构代替全连接层，可有效减少参数和计算时间。为了弥补模型轻量化导致的准确性下降，结合注意力机制来提高模型检测的准确性。

Ren 等[48]提出了一种 ATT-CNN-BiLSTM 模型，如图 6 所示。域名字符经过词嵌入层后，依次输入到 CNN 层和 Bi-LSTM 层提取特征。然后设置注意力层获取字符特征的相应权重，最后输出检测结果。该模型在常规 DGA 域名和基于词表生成的 DGA 域名的检测结果上都有着不错的表现。

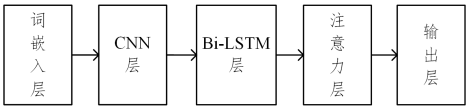


图 6 ATT-CNN-BiLSTM 模型中的神经网络结构

Fig. 6 Neural network structure in ATT-CNN-BiLSTM

Yang 等[49]提出了一个 HDNN 模型，使用多个 CNN 并行提取域名的局部特征，接着使用一个包含注意力机制的双向长期短期记忆网络(SA-Bi-LSTM) 提取双向的全局特征，最后引入损失函数来缓解训练中的样本不平衡的问题。与文献 [16] 中的基于 LSTM 的模型、文献 [19] 中的 LSTM-IM 模型、文献 [23] 中的 eXpose 模型、文献 [24] 中的 PCNN 模型和文献 [44] 中的 LSTM 结合注意力机制的模型相比，HDNN 模型不仅在常规 DGA 域名的二分类和多分类检测中取得了较好的效果，并在长度较短的 SDGA 域名检测上取得了较好的结果。

Namgung 等[50]使用基于注意力的 Bi-LSTM 和并行 CNN 分别进行域名特征的提取，类似于集成学习的方法，最后经过全连接层进行特征融合与分类，如图 7 所示。该方法也有效提高了 DGA 域名的检测效果，尤其是对于数据量很少的 DGA 域名家族，该模型也能进行很好的预测。

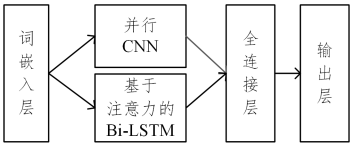


图 7 集成模型中的神经网络结构

Fig. 7 Neural network structure in ensemble models

最近，文献 [51] 从 DGA 域名的长度出发，提出了适应域名长度的特征提取方法。对于超短域名，使用基于注意力机制的方法提取特征；对于中等长度的域名，构建一个二维结构，使域名呈现出类似于图像的明显特征；对于超长域名，构建手工特征来实现域名的有效分类。最后设计不同的检测结构，形成异构 DGA 域名检测模型，提高了检测效果。

5.3 加入对抗样本

深度学习是针对现有数据集进行训练的，可能会被添加了细微扰动所形成的对抗样本欺骗[52]，因此可以在训练时加入一些具有干扰性的样本来改善这个问题。

生成对抗网络(Generative Adversarial Network，GAN) 是基于对抗思想诞生的新型神经网络，它包含一个生成网络和一个判别网络。判别网络要判别样本是来自生成网络的虚假数据还是真实的数据集，而生成网络则不断提升自己的生成虚假样本的能力，生成欺骗判别网络的对抗样本，两者在对抗过程中采取交替迭代的方法不断优化自身网络。利用 GAN 生成 DGA 域名对抗样本的过程如图 8 所示。经过训练的生成网络最大程度生成了虚假的对抗样本。

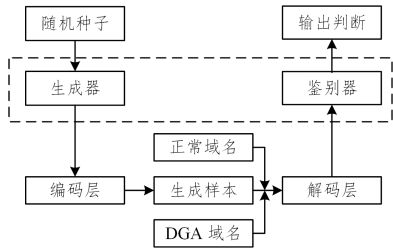


图 8 GAN 生成对抗样本

Fig. 8 GAN generates adversarial examples

Anderson 等[53]利用 GAN，构建了域名生成模型 DeepDGA，生成难以检测的域名作为对抗样本。实验证明，通过增加带有对抗样本的训练集增强了随机森林对 DGA 域名的检测效果。

文献 [54] 构建了一个 CharBot 模型，能够生成大量未注册的 DGA 域名的对抗样本，成功地欺骗了基于随机森林的检测模型和文献 [19] 中的 LSTM-MI 模型，逃避了检测。Liu 等[55]设计了一种融合字符级滑动窗口和深度残差网络的模型，在 DeepDGA 和 CharBot 生成的对抗样本中取得了不错的检测效果。

表 6 列出了基于附加机制的检测方法的比较结果。3 种基于附加机制的检测方能够有效提高对难检 DGA 域名的检测效果。其中，改进词嵌入方式和引入注意力机制有效提高了基于单词表生成的 DGA 域名和长度较短的 SDGA 域名的检测效果；利用对抗样本对现有数据集进行扩充，提升了检测模型的性能，提高了对样本数量较少的 DGA 域名和新型 DGA 域名的检测能力。

表 6 附加机制检测方法的比较

Table 6 Comparison of additional mechanism detection methods

方法	文献	数据集	改进方式	主要结论
改进词嵌入方式	[38]	Andrey Abakumov DGA 数据集、Johannes Bader DGA 数据集、OpenDNS 公共域名	改进单词级词嵌入方法，提高域名字符信息利用度	文献[38-42]通过对词嵌入方法进行优化，来提高模型的泛化能力，降低识别的误报率，提高召回率，模型的整体性能更优
	[39-42]	360 DGA 数据集、Alexa Top 100 万域名	改进字符级词嵌入方法，拓展域名字符特征	
结合注意力机制	[44-48,50]	DGArchive 域名、360 DGA 数据集、Alexa Top 100 万域名	文献[44-50]改变神经网络结构，将注意力机制和 LSTM、CNN 结合	模型的 F1 值最高可达 96.66%。加入注意力机制在大多数 DGA 域名类别检测中实现了最佳性能
	[49]	文献[37]的 SDGA 域名、Cisco Top 100 万域名		
	[51]	360 DGA 数据集、DGArchive 数据集、Alexa Top 100 万域名、Majestic Top 100 万域名	使用异构模型对不同长度的域名进行针对性检测；使用基于注意力机制的方法提取超短域名的特征	当域名长度超过 10 字节时，检测率超过 90%
加入对抗样本	[53-55]	360 DGA 数据集、Alexa Top 100 万域名	用对抗样本扩充数据集，改善了训练数据集	检测时的伪阴性率最低为 0.1%

6　未来研究趋势

基于域名字符特征的 DGA 域名检测依赖少，检测实时性高，未来仍将会是 DGA 域名检测的热点方向之一。下面对未来值得进一步研究的工作进行阐述。

1）传统机器学习检测方法的优化

根据域名字符特点设计手工特征，然后利用传统机器学习算法完成检测。这类方法简单有效，并且部署简单，适合小样本数据集。但是，也有耗时较长、手工特征选取依赖经验、容易被攻击者绕过的缺点。

未来可以在算法优化上进行研究，加快训练和检测速度，减少系统整体耗时。在特征工程中，如何设计更加有效的手工特征也值得继续研究。

2）深度学习检测方法的优化

基于深度学习的检测方法可以自动提取域名字符特征，一定程度上避免了特征工程的缺点，不易被攻击者绕过，而且实时性更高，在大数据集上有很好的检测效果，但是存在样本不平衡、模型泛化能力弱等问题。

如何对深度神经网络做进一步优化，通过改进模型来提高模型的泛化能力和解决样本不平衡的问题，进一步提高检测准确率，是未来的研究方向。

3）难检 DGA 域名检测的进一步研究

为了提高隐蔽性，DGA 域名生成机制也变得更加复杂，DGA 域名也在不断更新，出现了一些较难检测的 DGA 域名。传统的机器学习和深度学习算法对它们的检测效果不佳。基于附加机制的检测方法能够提高对难检 DGA 域名的检测效果。

未来的研究可以针对文中提出的 3 种基于附加机制的检测方法，做进一步改进和提升，或者在此基础上引入其他有效机制。

此外，数据集是各种检测方法的基础，无论是在机器学习还是深度学习中，数据的获取和数据集的构建都十分重要。但在大多数研究中，使用的 DGA 域名数据集都是国外数据集，因此构建我国权威的域名数据集也尤为重要。

结束语　提升 DGA 域名的检测能力对保障网络安全具有重要意义，本文梳理和总结了基于字符特征的 DGA 域名检测方法，同时对未来可能的研究热点进行了介绍，为进一步研究提供了参考。

参 考 文 献

[1] NIU W N，JIANG T Y，ZHANG X S，et al. Fast-flux botnet detection method based on spatiotemporal feature of network traffic[J]. Journal of Electronics & Information Technology，2020，42(8)：1872-1880.

[2] ZOU F，TAN Y，WANG L，et al. Botnet detection based on generative adversarial network[J]. Journal on Communications，2021，42(7)：95-106.

[3] DEHKORDI M J，SADEGHIYAN B. Reconstruction of C&C channel for P2P botnet[J]. IET Communications，2020，14(8)：1318-1326.

[4] WANG Z，GUO Y. Neural networks based domain name generation[J/OL]. Journal of Information Security and Applications，2021，61：102948. https://doi.org/10.1016/j.jisa.2021.102948.

[5] PLOHMANN D，YAKDAN K，KLATT M A，et al. A comprehensive measurement study of domain generating malware[C]// 25th USENIX Security Symposium. Austin，TX，USA：USENIX Association，2016：263-278.

[6] ALMASHHADANI A O，KAIIALI M，CARLIN D，et al. Maldom Detector：A system for detecting algorithmically generated domain names with machine learning[J/OL]. Computers & Security，2020，93：101787. https://doi.org/10.1016/j.redox.2020.101787.

[7] BARABOSCH T，WICHMANN A，LEDER F，et al. Automatic extraction of domain name generation algorithms from current malware[C]// NATO Symposium IST-111 on Information Assurance and Cyber Defense. Koblenz，2012.

[8] YADAV S，REDDY A K K，REDDY A L，et al. Detecting algorithmically generated malicious domain names[C]// Proceedings of the 10th ACM SIGCOMM conference on Internet measurement. Melbourne，Australia，2010：48-61.

[9] YADAV S，REDDY K，REDDY N，et al. Detecting algorithmically generated domain-flux attacks with DNS traffic analysis [J]. IEEE/ACM Transactions on Networking，2012，20(5)：1663-1677.

[10] ANTONAKAKIS M，PERDISCI R，NADJI Y，et al. From {throw-away} traffic to bots：detecting the rise of {DGA-based} malware[C]// 21st USENIX Security Symposium(USENIX Security 12). Bellevue，WA，2012：491-506.

[11] MAC H，TRAN D，TONG V，et al. DGA botnet detection using supervised learning methods[C]// Proceedings of the Eighth International Symposium on Information and Communication Technology. Nha Trang City，Viet Nam，2017：211-218.

[12] HUANG J，ZHANG G，SHEN Y. DGA domain name detection based on SVM under grey wolf optimization algorithm[C]// 2019 IEEE 10th International Conference on Software Engineering and Service Science(ICSESS). Newyork：IEEE Press，2019：245-248.

[13] LISON P，MAVROEIDIS V. Automatic detection of malware-generated domains with recurrent neural models[J]. arXiv：1709.07102，2017.

[14] MU Z C. Predicting Domain generation algorithms with N-Gram models[C]// 2022 International Conference on Big Data，Information and Computer Network(BDICN). Newyork：IEEE Press，2022：31-38.

[15] WANG H. Botnet detection via machine learning techniques [C]// 2022 International Conference on Big Data，Information and Computer Network(BDICN). IEEE，2022：831-836.

[16] WOODBRIDGE J，ANDERSON H S，AHUJA A，et al. Predicting domain generation algorithms with long short-term memory networks[J]. arXiv：1611.00791，2016.

[17] CHEN L G，ZHANG Y D，GENG G G，et al. Detection of ran-

dom generated names using recurrent neural network with gated recurrent unit[J]. Computer Systems & Applications, 2018, 27(8):198-202.

[18] SHAHZAD H, SATTAR A R, SKANDARANIYAM J. DGA domain detection using deep learning[C]// 2021 IEEE 5th International Conference on Cryptography, Security and Privacy (CSP). Newyork:IEEE Press, 2021:139-143.

[19] TRAN D, MAC H, TONG V, et al. A LSTM based framework for handling multiclass imbalance in DGA botnet detection[J]. Neurocomputing, 2018, 275:2401-2413.

[20] CHEN Y, PANG B, SHAO G, et al. DGA-based botnet detection toward imbalanced multiclass learning[J]. Tsinghua Science and Technology, 2021, 26(4):387-402.

[21] KIM Y. Convolutional neural networks for sentence classification[C]// The 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar, 2014: 1746-1751.

[22] ZHANG X, ZHAO J, LECUN Y. Character-level convolutional networks for text classification[J/OL]. Advances in Neural Information Processing Systems, 2015, 28. https://doi.org/10.48550/arXiv.1509.01626.

[23] SAXE J, BERLIN K. eXpose: A character-level convolutional neural network with embeddings for detecting malicious URLs, file paths and registry keys[J]. arXiv:1702.08568, 2017.

[24] YU B, PAN J, HU J, et al. Character level based detection of DGA domain names[C]// 2018 International Joint Conference on Neural Networks. Rio de Janeiro, Brazil, 2018:1-8.

[25] ZHOU S, LIN L, YUAN J, et al. CNN-based DGA detection with high coverage[C]// 2019 IEEE International Conference on Intelligence and Security Informatics(ISI). New York:IEEE Press, 2019:62-67.

[26] YANG L H, LIU G J, ZHAI J T, et al. Improved algorithm for detection of the malicious domain name based on the convolutional neural network[J]. Journal of Xidian University, 2020, 47(1):37-43.

[27] ZHOU C, SUN C, LIU Z, et al. A C-LSTM neural network for text classification [J]. Expert Systems with Applications, ELSEVIER, 2017, 72:221-230.

[28] ZHANG B, LIAO R J. Malicious domain name detection model based on CNN and LSTM[J]. Journal of Electronics & Information Technology, 2021, 43(10):2944-2951.

[29] XU G T, SHENG Z W. DGA malicious domain name detection method based on fusion of CNN and LSTM[J]. Netinfo Security, 2021, 21(10):41-47.

[30] PEI L Z, ZHAO Y J, WANG Z, et al. Comparison of DGA Domain Detection Models Using Deep Learning [J]. Computer Science, 2019, 46(5):111-115.

[31] BERMAN D S. DGA CapsNet: 1D application of capsule networks to DGA detection[J]. Information, 2019, 10(5):157.

[32] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality [C]// The 27th Advances in Neural Information Processing Systems. Stateline, USA, 2013:3111-3119.

[33] PENNINGTON J, SOCHER R, MANNING C. Glove: global vectors for word representation[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha. Qatar, 2014:1532-1543.

[34] PETERS M E, NEUMANN M, IYYER M, et al. Deep contextualized word representations[C]// Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics. New Orleans, 2018:2227-2237.

[35] BROWN T, MANN B, RYDER N, et al. Language models are few-shot learners[J]. Advances in Neural Information Processing Systems, 2020, 33:1877-1901.

[36] HOWARD J, RUDER S. Universal language model fine-tuning for text classification [C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne, Australia, 2018:328-339.

[37] FU Y, YU L, HAMBOLU O, et al. Stealthy domain generation algorithms[J]. IEEE Transactions on Information Forensics and Security, 2017, 12(6):1430-1443.

[38] KOH J J, RHODES B. Inline detection of domain generation algorithms with context-sensitive word embeddings[C]// 2018 IEEE International Conference on Big Data (Big Data). New York:IEEE Press, 2018:2966-2971.

[39] DU P, DING S F. A DGA domain name detection method based on deep learning models with mixed word embedding[J]. Journal of Computer Research and Development, 2020, 57(2):433-446.

[40] HU P C, DIAO L L, YE H, et al. DGA domains detection based on artificial and depth features[J]. Computer Science, 2020, 47(9):311-317.

[41] PAN R, CHEN J, MA H Y, et al. Using extended character feature in Bi-LSTM for DGA domain name detection[C]// 2022 IEEE/ACIS 22nd International Conference on Computer and Information Science(ICIS). New York:IEEE Press, 2022:115-118.

[42] YANG L, LIU G, LIU W, et al. Detecting multielement algorithmically generated domain names based on adaptive embedding model[J]. Security and Communication Networks, 2021, 2021(6):1-20.

[43] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. arXiv:1409.0473, 2014.

[44] QIAO Y, ZHANG B, ZHANG W, et al. DGA domain name classification method based on long short-term memory with attention mechanism[J]. Applied Sciences, 2019, 9(20):4205.

[45] TUAN T A, LONG H V, TANIAR D. On Detecting and Classifying DGA Botnets and their Families[J/OL]. Computers & Security, 2022, 113: 102549. https://doi.org/10.1016/j.cose.2021.102549.

[46] ZHAO K, GUO W, QIN F, et al. D3-SACNN: DGA domain detection with self-Attention convolutional network[J]. IEEE Ac-

cess，2021，10：69250-69263.

[47] YANG L，LIU G，WANG J，et al. Fast3DS：A real-time full-convolutional malicious domain name detection system[J/OL]. Journal of Information Security and Applications，2021，61：102933. https://doi.org/10.1016/j.jisa.2021.102933.

[48] REN F，JIANG Z，WANG X，et al. A DGA domain names detection modeling method based on integrating an attention mechanism and deep neural network[J]. Cybersecurity，2020，3(1)：1-13.

[49] YANG L H，LIU G J，DAI Y W，et al. Detecting stealthy domain generation algorithms using heterogeneous deep neural network framework[J]. IEEE Access，2020，8：82876-82889.

[50] NAMGUNG J，SON S，MOON Y S. Efficient deep learning models for DGA domain detection[J]. Security and Communication Networks，2021，2021(2)：1-15.

[51] LIANG J，CHEN S，WEI Z，et al. HAGDetector：Heterogeneous DGA Domain Name Detection Model[J]. Computers & Security，2022：102803.

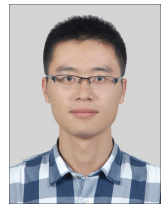
[52] SZEGEDY C，ZAREMBA W，SUTSKEVER I，et al. Intriguing properties of neural networks[J]. arXiv：1312.6199，2013.

[53] ANDERSON H S，WOODBRIDGE J，FILAR B. DeepDGA：adversarially-tuned domain generation and detection[C]// Proceedings of the 2016 ACM Workshop on Artificial Intelligence and Security. 2016：13-21.

[54] PECK J，NIE C，SIVAGURU R，et al. CharBot：A simple and effective method for evading DGA classifiers[J]. IEEE Access，2019，7：91759-91771.

[55] LIU X Y，LIU J M，LIU C，et al. Novel botnet DGA domain detection method based on character level sliding window and deep residual network[J]. Acta Electronica Sinica，2022，50(1)：250-256.

WANG Yu，born in 1996，postgraduate，is a member of China Computer Federation. Her main research interests include data mining and deep learning in DGA domain name detection.



PAN Rui，born in 1988，master，senior engineer. His main research interests include cyber security，data governance and data security.

（责任编辑：喻藜）