doi :10.3969/j.issn.1671-1122.2021.03.011

一种基于被动 DNS 数据分析的 DNS 重绑定攻击检测技术

郭烜臻 [1,2]，潘祖烈 [1,2]，沈毅 [1,2]，陈远超 [1,2]

（1. 国防科技大学电子对抗学院，合肥 230037；2. 网络空间安全态势感知与评估安徽省重点实验室，合肥 230037）

摘　要：基于域名系统（DNS）的 DNS 重绑定攻击能够有效绕过同源策略、防火墙，窃取敏感信息，控制内网设备，危害巨大。DNS 重绑定需要通过设置恶意域名才能实现。针对 DNS 重绑定相关恶意域名的检测问题，文章提出一种基于被动 DNS 数据分析的 DNS 重绑定攻击检测模型（DNS Rebinding Classifier，DRC）。通过引入被动 DNS 数据，从域名名称、时间、异常通信及恶意行为等 4 个测度集刻画 DNS 重绑定相关域名；基于 C4.5 决策树、KNN、SVM 及朴素贝叶斯等分类方法对数据进行混合分类、组合训练及加权求值。交叉验证实验表明，DRC 模型对相关恶意域名的识别能够达到 95% 以上的精确率；与恶意域名检测工具 FluxBuster 进行对比，DRC 模型能够更准确地识别相关恶意域名。

关键词：DNS 重绑定；被动 DNS；恶意域名检测；混合分类

中图分类号：TP309　**文献标志码：**A　**文章编号：**1671-1122（2021）03-0087-09

中文引用格式：郭烜臻，潘祖烈，沈毅，等 . 一种基于被动 DNS 数据分析的 DNS 重绑定攻击检测技术 [J]. 信息网络安全，2021，21（3）：87-95.

英文引用格式：GUO Xuanzhen, PAN Zulie, SHEN Yi, et al. DNS Rebinding Detection Technology Based on Passive DNS Data Analysis[J]. Netinfo Security, 2021, 21(3): 87-95.

DNS Rebinding Detection Technology Based on Passive DNS Data Analysis

GUO Xuanzhen[1,2], PAN Zulie[1,2], SHEN Yi[1,2], CHEN Yuanchao[1,2]

(1. *College of Electronic Engineering, National University of Defense Technology, Hefei* 230037, *China*; 2. *Anhui Province Key Laboratory of Cyberspace Security Situation Awareness and Evaluation, Hefei* 230037, *China*)

Abstract: DNS rebinding attack based on the domain name system (DNS) can effectively bypass the homologous strategy and firewall, steal sensitive information, and control intranet devices, causing great harm to the Internet community. DNS rebinding can only be realized by setting malicious domain name. Aiming at the detection of malicious domain names related to DNS rebinding, this paper proposes a DNS rebinding classifier (DRC) based on passive DNS data analysis. By introducing passive DNS data, the domain names related to DNS rebinding

收稿日期：2020-06-18

基金项目：国家重点研发计划 [2017YFB0802900]

作者简介：郭烜臻（1996—），男，江西，硕士研究生，主要研究方向为网络空间安全；潘祖烈（1976—），男，安徽，副教授，博士，主要研究方向为网络空间安全；沈毅（1986—），男，重庆，讲师，硕士，主要研究方向为网络空间安全；陈远超（1996—），男，福建，硕士研究生，主要研究方向为网络空间安全。

通信作者：郭烜臻 guoxuanzhen@nudt.edu.cn

are characterized from the four measure sets of domain name, time, abnormal communication and malicious behavior. Based on C4.5 decision tree, KNN, SVM and naive Bayes classification methods, the data are classified, trained and weighted. Cross validation experiments show that the accuracy of DRC model for identifying related malicious domain names can reach more than 95%. Compared with the malicious domain name detection tool FluxBuster, DRC model can identify related malicious domain names more accurately.

Key words: DNS rebinding; passive DNS; malware domain name detection; mixed classification

0 引言

域名系统（DNS）是现代互联网的核心组成部分，提供了方便记忆的域名与复杂的IP地址的相互映射，推动了互联网的发展。但随着网络技术的发展，DNS因其协议及系统的脆弱性，成为不法分子发动网络袭击的跳板，使互联网在可用性、安全性和完整性等方面受到损害。如何保障DNS安全成为当下亟待研究与解决的重点问题[1]。

本文所研究的DNS重绑定攻击属于基于DNS协议的攻击。该类攻击基于DNS协议漏洞，重绑定域名解析IP地址，以达到绕过同源策略渗透进内网，从而控制设备及窃取敏感信息的目的[2]，危害巨大。DNS重绑定攻击早在20世纪90年代就已出现。随着时间的推移，DNS重绑定的种类日益增多，攻击手段也更加丰富。在文献[3]中，DNS重绑定可以通过浏览器对内部网络进行攻击。DAI[4]等人将DNS重绑定与DNS 洪水攻击相结合，能够有效绕过包括DNS pinning在内的防御策略。随着物联网技术的普及，DNS重绑定技术也随之发展。在第27届DEFCON黑客大会上，研究人员展示了利用DNS 重绑定技术，对Google home、Roku TV、SONY音响设备、SonosWiFi Speakers、radio thermostat等设备进行攻击，并获取设备控制权限[5]。据物联网安全公司Armis披露，全球有近5亿台设备受到这种攻击方式的威胁[6]。CVE-2018-1002103[7]通过DNS重绑定，在Minikube上远程执行代码，并实现了虚拟机逃逸的效果。可见，亟需提出针对DNS重绑定攻击的检测方法。

DNS的攻击检测往往与相关恶意域名的检测相关联。传统的DNS攻击检测主要分为两种：1）通过DNS实时流量检测DNS攻击；2）通过分析DNS日志检测DNS攻击。近年来，随着机器学习的兴起，研究者通过将DNS数据与机器学习相结合以准确高效地完成对DNS攻击的检测。

文献[8]提出对DNS日志中特定类型的报文进行检测，如AAAA和PTR类型的报文，确定被攻击的IP地址，并进行防护。文献[9]收集获取DNS流量，分析DNS流量特性，利用分类的思想，从数据采集、特征提取、分类器选择等方面建立分类模型，识别DNS恶意域名。该方法存在的问题是DNS流量难以获取，并且涉及隐私问题。

值得注意的是，近年来，作为基于流量和日志的检测技术的补充，被动DNS数据因其数据的及时性、合理性及全面性，替代了数据流分析，被广泛应用于DNS攻击检测及恶意域名识别领域。

ANTONAKAKIS[10]等人于2010年提出一种名为Notos的基于被动DNS数据的域名信誉评价系统。该系统采用网络特征、域特征和黑名单证据特征等3大类特征，并对数据进行标记，利用随机森林算法识别恶意域名。但是该系统仅关注目标地理位置及逻辑归属，忽略了访问特征。2014年，BILGE[11]等人提出Exposure方法对恶意域名进行识别。该方法提出了15个特征集，利用决策树的方法实现恶意域名的识别。但该方法局限于DNS的时间特征，无法对DNS重绑定的相关域名进行有效识别。

Segugio也是一款利用DNS数据进行僵尸网络恶意域名识别的系统[12]。Segugio可作为攻击检测系统部署在两个大型ISP网络之间，然而其局限性在于只能依据待测域名与已知域名的关系进行恶意域名的判定。PERDISCI[13]等人提出FluxBuster方法，利用DNS数据，

通过包括IP变迁在内的13个特征，采用聚类算法进行速变域名识别。然而该方法难以应对种类日益繁多的DNS攻击。张维维[14]等人提出DOAS系统，通过分析DNS实时流量，从依赖性和使用位置等角度，利用多分类器实现对恶意DNS流量的检测。然而在缺乏足够样本的情况下，该方法的结果会有较大波动。

目前针对DNS攻击的检测主要是围绕钓鱼网站、拒绝服务攻击及僵尸网络的检测，针对DNS重绑定攻击的检测研究较少，对相关恶意域名的检测存在不足。为了提高DNS重绑定攻击检测效果，本文深入研究DNS重绑定攻击特性，提出一种基于被动DNS数据分析的DNS重绑定攻击检测模型（DNS Rebinding Classifier，DRC）。本文基于被动DNS数据的域名属性、存活时间等特性提取部分静态特征，基于网络通信行为、域名解析等特性提取动态特征，最终共提取出域名名称、时间、恶意行为、异常通信等4个测度集的16种特征，结合机器学习与混合分类算法对模型进行组合训练并加权求优，从而对相关恶意域名进行识别。实验结果表明，本文方法能够有效检测DNS重绑定攻击。与相关研究进行对比实验表明，本文方法能够实现更好的DNS重绑定攻击检测效果。

1 相关知识

1.1 DNS 重绑定的实现原理

为了更好地检测DNS重绑定攻击，需要了解其基本的实现原理。在发起DNS重绑定攻击之前，需要设置相关恶意域名及恶意DNS服务器。图1展示了基于时间变化的DNS重绑定攻击实现步骤。

DNS重绑定攻击步骤具体如下：

1）攻击者设置恶意网站和恶意域名服务器，受害者申请访问恶意网站，向恶意域名服务器发出解析请求；

2）恶意域名服务器接收解析请求，返回Web服务器对应IP地址；

3）受害者访问Web服务器，下载包含恶意Javascript脚本的页面代码；

4）因攻击者设置的*TTL*（DNS解析记录存活时间）

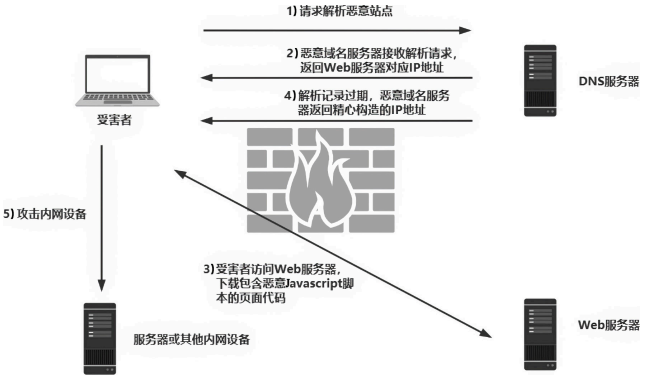

图 1 DNS 重绑定攻击步骤

值较短，步骤2）所返回的DNS解析地址过期，受害者再次申请解析恶意网站域名，此时恶意域名服务器返回精心构造的IP地址，如内网IP地址；

5）攻击者配合页面中的攻击脚本，绕过浏览器同源策略的限制，达到渗透攻击的目的。

DNS重绑定攻击并不只局限于内网设备。事实上，DNS重绑定攻击所针对的目标类型多样。按照威胁对象划分，如果DNS重绑定攻击返回的是公网IP地址，则为基于公网IP的DNS重绑定攻击，主要威胁目标为公网服务器；如果DNS重绑定攻击返回的是私有IP地址，则为基于内网IP的DNS重绑定攻击，攻击者能够获取内部网络中的敏感信息及设备控制权，危害较大。

1.2 被动 DNS 数据

本文在DNS重绑定攻击检测研究中引入了被动DNS（Passive DNS）数据。被动DNS数据是指从真实互联网中采集DNS流量，经过解析、去重、过滤等步骤后存储到相应数据库中的信息。被动DNS数据最早由WEIMER[15]于2004年提出，WEIMER将被动DNS数据的作用总结为恶意域名识别、流量分析、黑名单分析以及DNS错误配置检测。因对域名描述的全面性，可以通过被动DNS数据来展现互联网的具体情况，为安全研究者的研究提供相应依据。被动DNS数据结构如图2所示。

2 DNS 重绑定攻击检测模型

2.1 系统模型

本文提出的DNS重绑定攻击检测模型（DRC）如图3

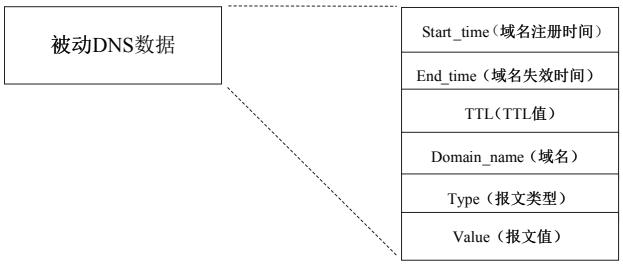

图 2 被动 DNS 数据结构

所示。首先，通过开源数据库获取被动DNS数据，并进行数据清洗与存储，将被动DNS数据存储到MySQL数据库中。其次，特征统计模块对数量众多的特征进行筛选，根据DNS重绑定的具体实现方法，删除掉一些无关或者效果不明显的特征，以减小机器学习的计算复杂度和系统开销。本文选取了4个测度集的16个特征对数据进行处理。对于已经确定为正常和恶意的域名，打上相应的标签，将数据传输到分类器引擎。该引擎的作用是通过标签集训练恶意域名的识别模型，通过多个分类器的训练学习，检测待测数据集中与DNS重绑定有关的恶意域名；在此基础上，将结果传入加权求优模块，以加权优化的方式，对结果进行混合分类，提高最终预测的准确率。最后，将结果生成报告。

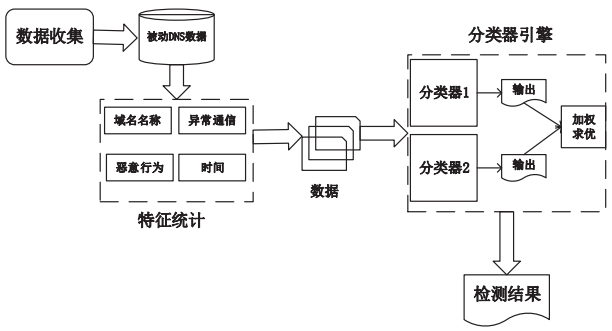

图 3 DRC 模型

2.2 DNS 重绑定攻击特征分析与恶意域名特征选择

2.2.1 DNS 重绑定攻击特征分析

本文所选取的特征均是从DNS重绑定的实现方式所得。DNS重绑定按照实现方式的不同可以分为3类：基于时间变化的DNS重绑定、基于多个A记录的DNS重绑定和基于插件的DNS重绑定。具体如表1所示。

表 1 DNS 重绑定分类

种类	描述
基于时间变化	同一域名在较短时间内有多个不同 IP 的 DNS 解析报文，利用此特点进行 DNS 重绑定攻击，往往与速变域名相联系
基于 A 记录	利用从接收到的 A 记录中解析出的多个 IP 进行 DNS 重绑定攻击
基于插件	利用浏览器插件漏洞进行 DNS 重绑定攻击

事实上，DNS重绑定相关恶意域名与正常域名之间存在着较大区别，如表2所示。

表 2 正常域名与 DNS 重绑定相关恶意域名的区别

正常域名	DNS 重绑定相关恶意域名
容易记忆	记忆复杂
域名解析行为正常	域名解析行为异常
域名存活时间长	域名存活时间短
IP 解析结果正常	IP 解析返回私有 IP
网页源码正常	网页源码返回恶意攻击脚本
应答报文 *TTL* 值正常	应答报文 *TTL* 值偏短

DNS重绑定中，DNS解析报文的*TTL*值较短，大多数DNS重绑定攻击的*TTL*值均在10s以下。综上所述，DNS重绑定具有如下基本特征：

1）出现私有IP地址。

私有IP地址包括A类：10.0.0.0～10.255.255.255；B类：172.16.0.0~172.31.255.255；C类：192.168.0.0～192.168. 255.255。域名解析中一般不出现私有 IP 地址，但 DNS 重绑定相关恶意域名的 IP 解析结果中会出现私有 IP 地址。私有 IP 地址用集合和正则表达式可表示为 Private_ip= {Domain()|Domain()∈r'^1((((0|27)(.(([1-9]?|1[0-9])[0-9]|2([0-4][0-9]|5[0-5])))|(72.(1[6-9]|2[0-9]|3[01])|92.168))(.(([1-9]?|1[0-9])[0-9]|2([0-4][0-9]|5[0-5]))){2})$'}。其中包括了 A 类、B 类、C 类及回环 IP 地址。

2）*TTL*值过短。

DNS重绑定应答报文中的*TTL*值过短，小于特定阈值。

3）行为异常。

短时间内同一域名接收到多个A记录，或者接收到的一个A记录中包含多个IP。用集合可表示为Abnormal_behavior={$m|m \in$ time_varying_record$\cup$ multiple_A_record }。

4）出现异常流量。

浏览器与Web服务器的HTTP应答报文中存在恶

意脚本，可表示为Abnormal_script。

因此，满足DNS重绑定攻击的特征匹配规则为DNS_rebinding_set()={ $n \mid \forall n \in$ Private_ip()∪TTL_time()∪Abnormal_behavior()∪Abnormal_script()}。其中，Private_ip表示出现私有IP，TTL_time()表示*TTL*值过短，Abnormal_behavior()表示行为异常，Abnormal_script()表示出现异常流量。

2.2.2 恶意域名特征选择

分析出DNS重绑定攻击的基本特征后，从实现方式出发，进一步对DNS重绑定相关恶意域名特征进行选择。

1）基于域名名称的特征

通常恶意域名是数字、字母和特殊符号的随机组合，与正常域名差异较大。字母数量指在域名中出现的字母数量，数字数量指在域名中出现的数字数量，域名中出现的特殊符号数量也纳入考虑范围，三者作为域名的基本特征描绘了域名的基本面貌。是否含有著名域名是指该域名是否是著名域名或其子站，著名域名可以参考Alexa Topsites排行。同理，将是否与恶意域名相关也纳入考虑范围。

2）异常通信特征

异常通信特征包括域名地理位置、域名解析服务器及是否含有私有IP。某些特定区域的恶意域名活动较频繁，故需判定域名是否属于这些敏感区域。与DNS重绑定相关的域名解析服务器往往是伪造的服务器或者是受攻击者控制的服务器，故当NS（域名服务器）记录对应一个未知可疑服务器时，有存在DNS重绑定攻击的可能。当前DNS重绑定倾向于控制内部网络中的IoT设备，或者窃取敏感文件，域名解析结果中出现私有IP也不是正常行为，故可将此作为DNS重绑定的一个判断依据。

3）基于时间的特征

基于时间的特征包括*TTL*和域名存活时间等多种。*TTL*值过短是DNS重绑定的一项重要特征。每次的DNS解析报文往往对应不同的*TTL*值，反映了域名解析的变化情况。故将最大*TTL*值、最小*TTL*值及*TTL*值的变化次数纳入本测度集。域名存活时间是域名的一个重要特征。一般来说，与DNS重绑定相关的域名存活时间较短。域名存活时间计算方法如公式（1）所示。

$$Survivaltime(n)=\frac{starttime(n)-endtime(n)}{3600} \quad (1)$$

式中，*starttime*(*n*)为域名存活开始时间，*endtime*(*n*)为域名过期时间。

4）恶意行为特征

DNS重绑定的一大特征是域名对应的IP重绑定前后解析结果不同。进行DNS重绑定攻击时，域名解析请求数量暴增，域名服务器会对相同域名在短时间内进行多次解析；也可采用如文献[4]所展示的firedrill方法，对目标浏览器进行DNS洪泛攻击，通过发送大量的DNS解析报文，填充浏览器DNS缓存以达到篡改域名解析地址的目的。

本文所使用的4个测度集的16个特征如表3所示。

表 3 特征信息

测度集	编号	特征名	描述
域名名称	1	Domain_special	特殊字符
	2	Famous_domain	与著名域名相关
	3	Malware_domain	与恶意域名相关
	4	Num_word	域名字母数量
	5	Num_digit	域名数字数量
	6	Domain_length	域名长度
异常通信	7	NS	域名解析服务器是否正常
	8	Geo_feature	地理位置信息
	9	Num_IP	解析是否对应多个 IP
	10	Private_IP	是否为私有 IP
时间	11	Survival_time	域名存活时间
	12	Min_TTL	最小 *TTL* 值
	13	Max_TTL	最大 *TTL* 值
	14	Change_ttl	*TTL* 值变化次数
恶意行为	15	Abnormal_behavior	是否有异常解析行为
	16	Flood_behavior	是否有洪泛 DNS 解析

2.3 分类器引擎

分类器引擎由两部分组成：分类算法的选择及分类结果的加权求优。

2.3.1 分类算法

分类的目的是确定一个对象的类别。分类算法是指给定一个对象*X*，将其划分到预定义好的某个类别*Y*中的算法。本文选取较常用的几种分类算法进行对比实验，从中选出最适合本文数据集的两种算法。本文选取的分类算法有C4.5决策树、支持向量机（SVM）、

贝叶斯模型和最近邻算法（KNN）。

决策树方法是通过训练数据构建决策树，对未知数据进行分类，找出最高信息增益的属性。决策树的特点是简单、易于理解，能够在短时间内处理大量数据，但容易出现过拟合的现象。支持向量机把分类问题转化为寻找分类平面的问题，通过最大化分类边界点距离分类平面的距离来实现分类。支持向量机能够解决小样本下的机器学习问题，提高泛化能力，但是内存消耗过大。贝叶斯模型是基于贝叶斯定理的统计学分类方法，通过预测一个给定元组属于一个特定类的概率来进行分类。最近邻算法是一种监督学习方法，其核心思想是未标记样本的类别由距离最近的K个邻居来决定，但该方法存在样本分类不均衡的问题，容易产生误判。

2.3.2 混合分类

传统的学习模型采用单一算法，单一算法在某些领域存在局限性，准确率不高。本文采用混合分类算法，混合分类算法的目的是使效果更优的算法所占权值更大，综合各算法之长，求得最优值。基于现有条件，使问题简单化，本文采用两种分类算法对DNS重绑定相关域名样本进行分类学习、组合训练及加权求值，以获得更精确的判断。

从混合分类的基本思想出发，数据需要先经过两种机器学习算法进行预处理，再传入混合分类算法。即在进入加权求优阶段之前，两种分类算法已经对同一数据集进行了处理，并分别进行了多次分类实验。将每次实验得到的结果与真实情况进行对比[16]，计算准确率 $\alpha_{i,j}$（即正确检测的域名数与域名总数的比值）。其中，i表示所采用的分类算法，i=1,2；j表示实验的次数，$1\leqslant j\leqslant 10$ 且$j\in \mathbf{Z}$。例如，$\alpha_{1,3}$表示使用第1种分类算法第3次实验的准确率。

设$v_{i,j}$为算法i在第j次实验的结果，即正确检测的域名数。$\bar{v}$为n次实验结果的平均值，计算$v_{i,j}$与$\bar{v}$的逆平方差值$d_{i,j}$，即

$$d_{i,j}=\frac{1}{\sqrt{|\bar{v}^2-v_{i,j}^2|}} \tag{2}$$

$d_{i,j}$越大，表示该算法所得结果越稳定。综合逆平方差值与准确率，判断算法i在第j次实验中所占权值β，即

$$\beta_{i,j}=\frac{d_{i,j}\alpha_{i,j}}{\sum_{i=1,2}^{j=n}d_{i,j}\,\alpha_{ij}} \tag{3}$$

$\beta_{i,j}\in(0,1)$。综合每次权值，计算

$$W=\frac{\sum_{i=1,2}^{j\leqslant 10}f_{i,j}\,\beta_{i,j}}{n} \tag{4}$$

其中，$f_{i,j}$表示算法i在第j次实验中的某个评价指标，W为经过混合分类算法最终得到的该评价指标。

3 实验与分析

3.1 实验环境与评价指标

3.1.1 实验环境

实验部署在一台个人计算机上，CPU型号为Intel（R）Core（TM）i7-4790，主频为3.6GHz，内存为10 GB，硬盘为500 GB。使用的机器学习库为Scikit-Learn。

本文采用360公司的被动DNS数据库，利用360公司提供的flint工具，对被动DNS数据进行查询。flint工具工作在python 2.6或2.7环境下，能够直接查询特定域名的被动DNS数据。另外以VIRUSTOTAL、Passive Total[17]、CIRCL[18]等数据库中的被动DNS数据作为补充。

实验中涉及的黑名单主要来自于phishtank、kaggle及github上公开的恶意域名集[19]。为了增加恶意样本的多样性，通过实际抓包，分析DNS和HTTP应答报文，采用手工标记的方式补充与DNS重绑定相关的恶意域名。

本文的合法域名来自Alexa列表，在前期准备工作中，通过scrapy爬虫框架，获取了Alexa Top[20]100万个的域名作为参考。本文所使用的数据集包括合法域名2545个、DNS重绑定相关恶意域名412个。

3.1.2 评价指标

实验共涉及两种情况：与DNS重绑定无关的正常

域名和与DNS重绑定相关的恶意域名。评价标准包括精确率（Precision）、准确率（Accuracy）和召回率（Recall）3项指标。这3项指标计算涉及4个概念：真阳性（True Positive，TP）指判定为正常域名实际也为正常域名的样本；假阳性（False Positive，FP）指DNS重绑定相关恶意域名被误判成正常域名的样本；真阴性（True Negative，TN）指判定为恶意域名实际也的确是DNS重绑定相关恶意域名的样本；假阴性（False Negative，FN）指实际是正常域名却被判定为DNS重绑定相关恶意域名的样本。3项指标定义如下。

精确率指模型判断出的正例中，*TP*样本所占的比例。即

$$Precision=\frac{TP}{TP+FP} \quad (5)$$

准确率指模型判断出的正例占所有样本数量的比例。即

$$Accuracy=\frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

召回率指模型判断出的正例中，*TP*样本占所有正例的比例。即

$$Recall=\frac{TP}{TP+TN} \quad (7)$$

3.2 实验结果及分析

1）单独分类算法实验

实验采用10次交叉验证的方式以优化结果。把数据集分成10个大小一致的子集，每个子集中的恶意域名均匀分布。把其中一个子集作为测试集，剩下9个子集作为训练集，进行10次测试训练，最终返回10次结果的平均值。

目前，针对分类算法的研究相对成熟。本文不研究具体算法的优化和改进，只采用4类较常见的分类算法进行实验：C4.5决策树、KNN、SVM及朴素贝叶斯，并从准确率、精确率、召回率和完成分类所用时间4个维度进行评价。图4给出了各分类算法10次实验的精确率、准确率和召回率的平均值比较。图5给出了各分类算法10次实验的分类所用时间比较。

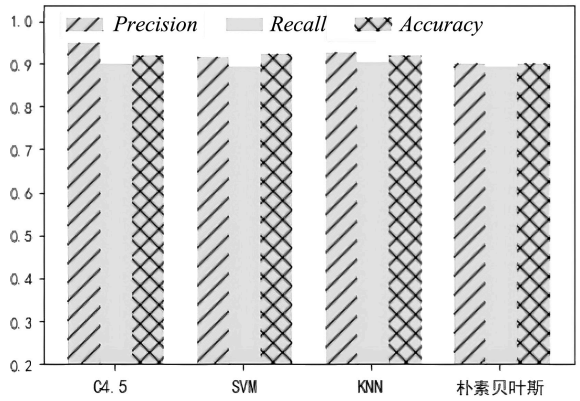

图 4 精确率、准确率和召回率对比

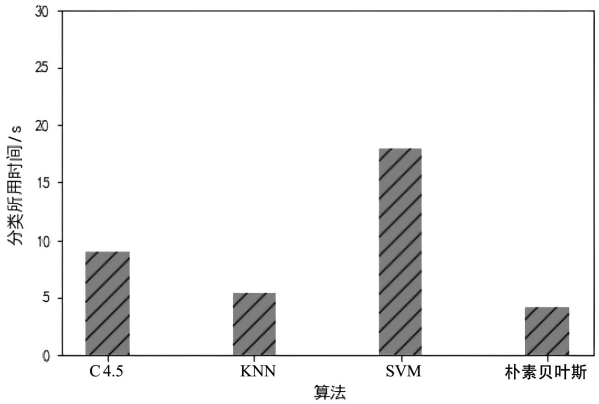

图 5 分类所用时间对比

由图4和图5可知，C4.5决策树在各项指标中均占优。其他3类算法中KNN占略微优势，且KNN在时间效率上占优。故本文选择C4.5决策树和KNN算法作为混合分类算法的输入，这两种算法的准确率和召回率均维持在90%以上。

2）混合分类算法实验

将由C4.5决策树和KNN得出的数据输入DRC模型，经过10次实验，运用公式（2）~公式（4），得到如图6所示的结果。从图6可以看出，相关恶意域名的识别精确率在95%以上，准确率在 90%以上，召回率在93%以上。可见DRC模型要优于单独的C4.5决策树、KNN、SVM和朴素贝叶斯算法。

3）同类分类算法比较

将DRC模型与同类型基于被动DNS数据的分类算法进行比较。每种算法的侧重点各不相同，如表4所示。Exposure算法选择了18个特征，主要针对僵尸网络相

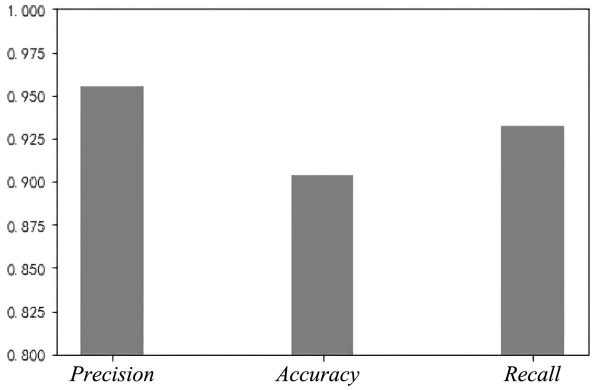

图 6 DRC 模型分类结果

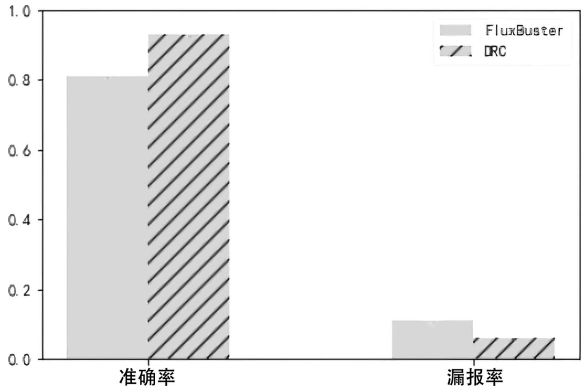

图 7 与 Fluxbuster 的比较结果

关域名进行检测；Kopis算法选择了23个特征，对域名的信誉进行检测，判断是否是恶意域名；Fluxbuster算法选择了13个特征，侧重于速变域名的检测。DRC模型选择了16个特征，该模型相较于其他3种分类算法的优势在于其在DNS重绑定相关恶意域名的检测上，增加了恶意域名检测的完备性。

表 4 同类算法的比较

算法	类别	作用	提取特征
Exposure	决策树	识别僵尸网络相关恶意域名	域名字面特征、资源记录、时间特征等 18 个特征
Kopis	随机森林	识别恶意域名	请求密度、解析 IP 声誉等 23 个特征
Fluxbuster	决策树	识别速变恶意域名	IP 地址数量、解析行为等 13 个特征
DRC	混合分类算法	识别 DNS 重绑定相关恶意域名	域名名称、时间、异常通信和恶意行为等 4 个测度集的 16 个特征

为进一步验证DRC模型对DNS重绑定相关恶意域名的识别能力，与Fluxbuster算法进行对比实验，评价指标为准确率和漏报率，如图7所示。漏报率反映了被误判成正常域名的样本占总的恶意域名样本的比例，漏报率越小，分类性能越好。实验结果表明，在同一数据集下，在DNS重绑定相关恶意域名的识别上，DRC模型在准确率和漏报率上均优于Fluxbuster。可见，DRC模型能够改进和补充现有分类算法的识别测度，提高DNS重绑定相关恶意域名的检测精度。

4 结束语

本文提出了一种基于被动DNS数据分析的DNS重绑定攻击检测模型DRC，解决了传统算法无法有效识别DNS重绑定相关恶意域名的问题。该方法目前检测的准确率和召回率均在90%以上，且有一定的提升空间。然而当前检测仅依据历史数据开展，实时性不够，无法及时进行响应。为了提升对DNS重绑定攻击的检测效果，后期拟采用集成学习的方法，并通过实时跟踪DNS流量，提升DNS重绑定攻击检测的时效性与准确性。

参考文献：

[1] HONG Bo, GENG Guanggang, WANG Liming, et al. System to Discover Phishing Attacks Actively Based on DNS[J]. Application Research of Computers, 2013, 30(12): 3771–3774.
洪博, 耿光刚, 王利明, 等. 一种基于DNS主动检测钓鱼攻击的系统[J]. 计算机应用研究, 2013, 30(12): 3771–3774.
[2] TATANG D, SUURLAND T, HOLZ T. Study of DNS Rebinding Attacks on Smart Home Devices[EB/OL]. https://www.researchgate.net/publication/335243263_Study_of_DNS_Re–bin–ding_At–tacks_on_Smart_Home_De–vices, 2020–01–20.
[3] JACKSON C, BORTZ A, BONEH D, et al. Protecting Browser State from Web Privacy Attacks[C]//ACM. The 15th International Conference on World Wide Web, May 23–26, 2006, Edinburgh, Scotland, UK. New York: ACM, 2006: 737–744.
[4] DAI Yunxing, RESIG R. FireDrill: Interactive DNS Rebinding[C]// USENIX. The 7th USENIX Conference on Offensive Technologies, August 13, 2013, Washington, DC, USA. Berkeley: USENIX Association, 2013: 2.
[5] DORSEY B. Attacking Private Networks from the Internet with DNS Rebinding[EB/OL]. https://medium.com/@brannondorsey/attacking–private–networks–from–the–internet–with–dns–rebinding–ea7098a2d325, 2018–06–19.
[6] Armis. DNS Rebinding Exposes Half a Billion Devices in the Enterprise[EB/OL]. https://www.armis.com/resources/iot–security–blog/dns–rebinding–exposes–half–a–billion–iot–devices–in–the–enterprise/, 2020–01–20.
[7] F–Secure. Minikube RCE & VM Escape[EB/OL]. https://labs.

f-secure.com/advisories/minikube-rce/, 2020-01-20.
[8] ZHA Chengji. Analysis of Mobile Internet Based on DNS Log[D]. Beijing: Beijing University of Posts and Telecommunications, 2014.
查诚吉 . 基于 DNS 日志的移动互联网分析 [D]. 北京：北京邮电大学，2014.
[9] TIAN Shiqi. Implementation and Traffic Analysis of DNS Traffic Collection System[D]. Beijing: Beijing University of Posts and Telecommunications, 2018.
田世奇 . DNS 流量采集系统的实现与流量分析 [D]. 北京：北京邮电大学，2018.
[10] ANTONAKAKIS M, PERDISCI R, DAGON D, et al. Building a Dynamic Reputation System for DNS[C]// USENIX. The 19th USENIX Conference on Security, August 11-13, 2010, Washington, DC, USA. Berkeley: USENIX Association, 2010: 18.
[11] BILGE L, SEN S, BALZAROTTI D, et al. Exposure: A Passive DNS Analysis Service to Detect and Report Malicious Domains[J]. ACM Transactions on Information and System Security, 2014, 16(4): 1-28.
[12] RAHBARINIA B, PERDISCI R, et al. Efficient and Accurate Behavior-based Tracking of Malware-control Domains in Large ISP Networks[J]. ACM Transactions on Privacy and Security, 2016, 2016(19): 1-31.
[13] PERDISCI R, CORONA I, GIACINTO G. Early Detection of Malicious Flux Networks via Large-scale Passive DNS Traffic Analysis[J]. IEEE Transactions on Dependable and Secure Computing, 2012, 9(5): 714-726.
[14] ZHANG Weiwei, GONG Jian, LIU Shangdong, et al. DNS Surveillance on Backbone[J]. Journal of Software, 2017, 28(9): 2370-2387.
张维维，龚俭，刘尚东，等 . 面向主干网的 DNS 流量监测 [J]. 软件学报，2017，28（9）：2370-2387.
[15] WEIMER F. Passive DNS Replication[EB/OL]. https://www.researchgate.net/publication/265577184_Passive_DNS_Replication, 2020-01-20.
[16] FARID D M, ZHANG Li, RAHMAN C M, et al. Hybrid Decision Tree and Naive Bayes Classifiers for Multi-class Classification Tasks[J]. Expert Systems with Applications, 2014, 41(4): 1937-1946.
[17] RISKIQ. RiskIQ PassiveTotal[EB/OL]. https://www.riskiq.com/products/passivetotal/, 2020-01-20.
[18] circl. lu. CIRCL Passive DNS[EB/OL]. https://www.circl.lu/services/passive-dns/, 2020-01-20.
[19] ZHOU Changling, CHEN Kai, GONG Xuxiao, et al. Detection of Fast-flux Domains Based on Passive DNS Analysis[J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2016, 52(3): 396-402.
周昌令，陈恺，公绪晓，等 . 基于 Passive DNS 的速变域名检测 [J]. 北京大学学报（自然科学版），2016，52（3）：396-402.
[20] Alexa. The Top 500 Sites on the Web[EB/OL]. https://www.alexa.com/topsites, 2020-01-20.