

# 基于J48决策树的CDN域名检测方法

杜 飞<sup>1</sup>, 李国静<sup>1</sup>, 王杉杉<sup>2</sup>

<sup>1</sup>北京锐驰信安技术有限公司, 北京

<sup>2</sup>中国联合网络通信有限公司河南省分公司, 河南 郑州

Email: move170@163.com

收稿日期: 2021年6月23日; 录用日期: 2021年7月20日; 发布日期: 2021年7月27日

## 摘 要

内容分发网络(Content Delivery Network)是新兴的网络加速技术, 可以避开互联网上影响数据传输速度和稳定性的瓶颈和环节, 使内容传输更快、更稳定。CDN有效的利用DNS, 将用户重定向到附近的CDN网络的边缘节点服务器来提高用户获取内容信息的效率和质量。DNS重定向技术的运用为CDN网络带来了显著的优势和灵活性。本文介绍了CDN的基础原理、工作流程及CDN域名监测的必要性, 采用基于J48决策树算法对CDN域名进行检测, 构建CDN基础资源库。实验结果表明, 该算法能有效提高CDN域名检测的效率, 准确率达到98.8%, 召回率达到98.1%。

## 关键词

域名系统, 内容分发网络, CDN

# CDN Domain Name Detection Method Based on J48 Decision Tree

Fei Du<sup>1</sup>, Guojing Li<sup>1</sup>, Shanshan Wang<sup>2</sup>

<sup>1</sup>Beijing Ruichixin'an Technology Co., Ltd., Beijing

<sup>2</sup>Henan Branch, China United Network Communications Co., Ltd., Zhengzhou Henan

Email: move170@163.com

Received: Jun. 23<sup>rd</sup>, 2021; accepted: Jul. 20<sup>th</sup>, 2021; published: Jul. 27<sup>th</sup>, 2021

## Abstract

Content Delivery Network is an emerging network acceleration technology. CDN can avoid bottlenecks and links on the Internet that affect the speed and stability of data transmission, that makes

the content transfer faster and more stable. CDN effectively utilizes DNS that redirects user to the edge node server of the nearby CDN network to improve the efficiency and quality of the user's access to content information. The application of DNS redirection technology brings significant advantages and flexibility to CDN. This paper introduces the basic principles, workflow, and the necessity of CDN domain name monitoring. It uses the J48 decision tree algorithm to detect CDN domain names and builds a basic CDN resource library. Experimental results show that the algorithm can effectively improve CDN domain name detection efficiency, with an accuracy rate of 98.8% and a recall rate of 98.1%.

## Keywords

Domain Name System, Content Delivery Network, CDN

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

随着互联网技术和多媒体技术的不断发展, 图像、视频、音频等服务在互联网中的占比越来越大, 并且随着互联网用户持续增加, 网络访问距离过长, 导致网络负载迅速增加, 从而使用户的访问质量受到严重影响[1] [2] [3] [4] [5]。

内容分发网络技术(Content Delivery Network)是新兴的网络加速技术, 它采用分布式缓存、负载均衡、流量工程和客户端重定向等技术, 通过建立若干加速节点, 尽量将用户互联网请求的内容存储到距离用户“最后一公里”的边缘节点上, 在互联网上构筑一个地理上分布的内容分送网络, 将信息资源向网络边缘推进, 用户可以在“最近”的位置快速访问到所需的内容, 很大程度上提高了终端用户的访问速度和服务质量[6] [7] [8]。

通过对 CDN 域名的监测, 可以减少网络上不良行为、较少对国家网络安全有危害的风险。本文通过对 CDN 域名检测技术的研究, 采用基于 J48 决策树算法对 CDN 域名的分类检测, 构建 CDN 基础资源库, 对基于域名的网络安全分析将起到积极的促进作用。

## 2. CDN 原理及组成

### 2.1. 概念

#### 1) 加速域名

加速域名即需要使用 CDN 加速的域名。域名是一组服务器的地址, 可以是网站、电子邮件、FTP 等。在 CDN 服务中加速域名通常指域名。

#### 2) 服务域名

域名接入 CDN 服务后, CDN 将分配一个 CNAME 域名, 该 CNAME 域名的形式例如为\*.kunlun.com, 自己的加速域名指向\*.kunlun.com 的域名。记录生效后, 域名解析的工作就正式转向 CDN 服务, 该域名所有的请求都将转向 CDN 节点, 达到加速效果。

#### 3) 边缘节点

在 CDN 服务中, 边缘节点、CDN 节点、Cache 节点、缓存节点、加速节点、节点等都指 CDN 边缘

节点。边缘节点是相对于网络的复杂结构而提出的一个概念，指与最终接入的用户之间具有较少中间环节的网络节点，对最终接入用户有相对于源站而言更好的响应能力和连接速度。其作用是将访问量较大的内容缓存在边缘节点的服务器上，以此来提高网端用户访问网站内容的速度和质量。

#### 4) CNAME 记录

CNAME 记录也是一种和 A 记录功能相似的域名记录方式，也叫别名记录，用来把域名解析到别的域名上。常用于 mail 邮箱解析和 CDN 加速解析，特别是 CDN 加速解析是现在 CNAME 记录最大的用途了。

下表演示 CDN 域名解析过程，<https://www.jd.com> 经过 CDN 加速，通过两次 CNAME 解析，最终通过 A 记录解析到固定 IP。表 1 为 CDN 域名解析过程。

**Table 1.** CDN domain name resolution process

**表 1.** CDN 域名解析过程

域名	解析形式	解析值
<a href="https://www.jd.com/">https://www.jd.com/</a>	CNAME	<a href="https://www.jd.com.gslb.qianxun.com">https://www.jd.com.gslb.qianxun.com</a>
<a href="https://www.jd.com.gslb.qianxun.com/">https://www.jd.com.gslb.qianxun.com/</a>	CNAME	<a href="https://www.jdcn.com">https://www.jdcn.com</a>
<a href="https://www.jdcn.com">https://www.jdcn.com</a>	A	58.218.206.1

#### 5) DNS

DNS (Domain Name System)，即域名解析服务。DNS 的作用：把域名转换成为网络可以识别的 IP 地址。人们习惯记忆域名，但机器间互相只识别 IP 地址。域名与 IP 地址之间是一一对应的，它们之间的转换工作称为域名解析，域名解析需要由专门的域名解析服务器来完成，整个过程自动进行。比如：上网时输入的 <https://www.baidu.com> 会自动转换成为 220.181.112.143。

## 2.2. CDN 技术原理

CDN 网络主要通过 DNS 机制进行资源定位。

CDN 网络节点由一系列分布在 Internet 上的节点组成，这些 CDN 节点位于网络的边缘，分布在不同的地理位置。在正常情况下(无 CDN)，用户的 DNS 请求通过 LDNS (本地 DNS)及 ADNS (权威 DNS)解析出域名对应的 IP 地址，实现对资源的访问。

引入 CDN 后，将修改用户目标域名对应的 IP，将其引导到其新部署的网络边缘节点上。一般实现是通过增加 CDNS 网元，即 CDN 运营商 DNS 解析服务器。修改用户的 LDNS，将用户未命中请求首先引导到 CDNS，由 CDNS 分配最优 CDN 节点并将地址返回用户。

采用各种 Cache 服务器(缓存设备)，并将它们分布到用户访问相对集中的地区或网络中，利用全局负载均衡技术(Global Server Load Balance)，将用户访问指向离用户最近、工作正常 Cache 服务器上，由 Cache 服务器直接响应用户的请求。当用户访问已加入 CDN 服务的网站时，首先通过 DNS 重定向技术确定最接近用户的最佳 CDN 节点，同时将用户的请求指向该节点。当用户请求到达指定节点时，CDN 的服务器(节点上的高速缓存)负责将用户请求的内容提供给给用户。如果 Cache 服务器中没有用户要访问的内容，它便根据配置自动到原服务器中抓取相应的页面提供给用户。

CDN 是一种组合技术，包括源站、缓存服务器、智能 DNS、用户客户端等几个部分。图 1 所示为 CDN 的全局结构图。

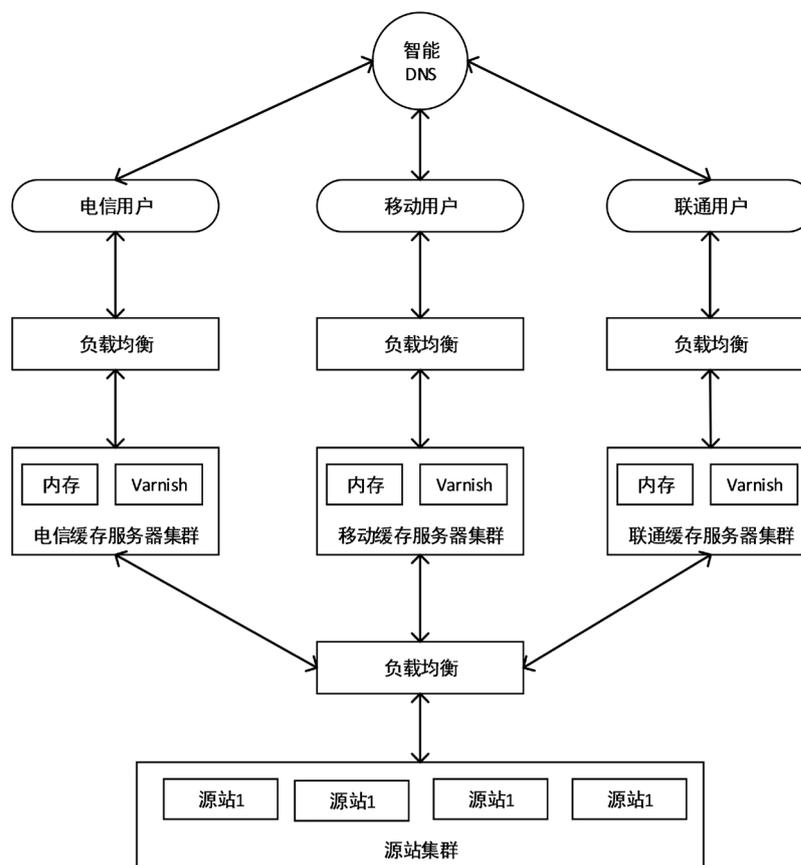


Figure 1. CDN global structure diagram

图 1. CDN 全局结构图

源站为发布内容的原始站点，可以新增、删除和更改网站的文件。缓存服务器抓取的对象也全部来自于源站。采用 CDN 技术后源站的压力会变得很小，为了提高可用性，将源站部署成负载均衡集群。智能 CDN 实现用户访问的转向功能，是整个 CDN 的核心，它根据用户访问的来源，将其访问请求转向到离用户较近或较合适的缓存服务器。

实现智能 DNS 的一种技术是 BindView，在本文中，使用电信、联通、移动网 3 个 IP 地址列表，每个列表关联独立的 Bind 视图，共有 3 个视图。

Varnish 作为缓存服务器，当用户发起访问时，它的访问请求被智能 DNS 定位到离它最近的缓存服务器，如果访问所需的内容没有被缓存，则缓存服务器向邻近的缓存服务器或直接向源站抓取内容，然后再返还给用户。如果用户所请求的内容刚好在缓存里面，则直接把内容返还给用户。为保证缓存服务本身的高可用性，每个点的服务器都以负载均衡集群的方式存在。

## 2.2. CDN 工作流程

当用户访问已经加入 CDN 服务的网站时，首先通过 DNS 重定向技术确定最接近用户的最佳 CDN 节点，同时将用户的请求指向该节点。当用户的请求到达指定节点时，CDN 的服务器负责将用户请求的内容提供给用户。

用户访问的基本流程如下：

- 1) 用户在自己的浏览器中输入要访问的网站域名；

- 2) 浏览器向本地域名解析系统(DNS)请求对该域名的解析;
- 3) 本地 DNS 将请求发到网站的主 DNS, 主 DNS 再将域名解析请求转发到其他 DNS;
- 4) DNS 根据路由策略确定当时最适当的 CDN 节点, 并将解析的结果(IP 地址)发给用户;
- 5) 用户向给定的 CDN 节点请求相应网站的内容;
- 6) CDN 节点中的服务器负责响应用户的请求, 提供所需的内容。

### 3. 解决方案

CDN 服务通过大量 CDN 加速节点缓存图片、静态资源等信息。CDN 节点由负载均衡设备和高速缓存服务器两部分组成。负载均衡设备负责每个节点中各个 Cache 的负载均衡, 保证节点的工作效率及收集节点与周围环境的信息, 保持与全局负载 DNS 的通信, 实现整个系统的负载均衡。高速缓存服务器就是个靠近用户的网站服务器, 响应本地用户的访问请求, 负责存储客户网站的大量信息[9] [10] [11] [12] [13]。

网站若使用 CDN 服务, 必定会将网站静态资源缓存到各地 CDN 加速节点中, 在不同地点访问网站会解析到不同的 IP 地址。

整体方案架构图如图 2 所示。

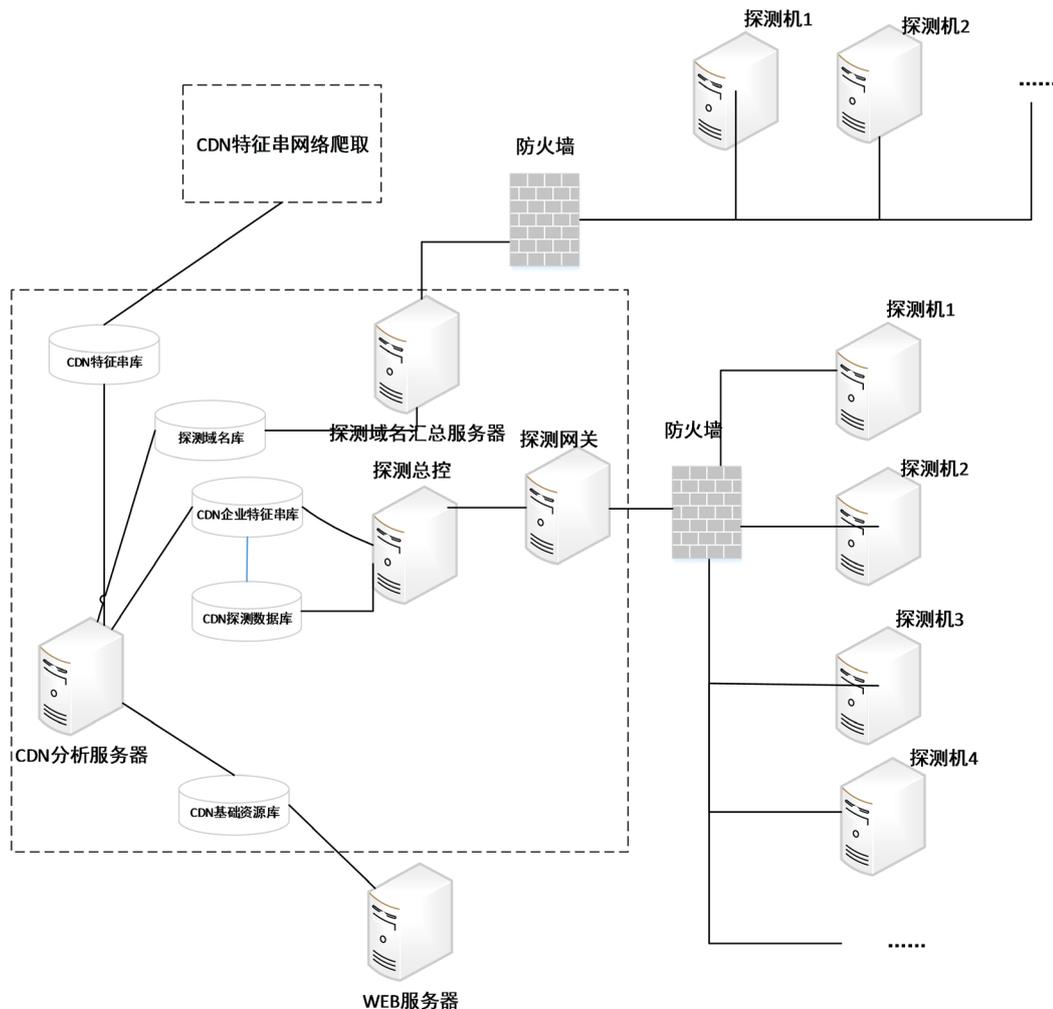


Figure 2. CDN domain detection deployment diagram  
图 2. CDN 域名检测部署图

使用域名主动探测的方式，在全国各省部署不同运营商的探测节点，有针对性的探测域名，获取域名解析记录，根据解析记录(cname)进行统计分析，分析出 CDN 企业特征串，录入 CDN 特征串库。

探测域名汇总服务器对探测机的探测域名解析记录进行记录，将过滤后的域名记录存入探测域名数据库。

CDN 分析服务器，网络爬取的 CDN 特征串与探测分析的 CDN 特征串，组成 CDN 特征串库。并定期针对探测域名库的 CDN 域名进行分析，通过 DNS 解析的 CNAME 记录是否包含特征串进行标记，以此判断此域名是否为 CDN 域名。若为 CDN 域名，将 CDN 域名、CDN 加速节点存入 CDN 基础资源库中。

CDN 特征串依赖于网络爬取的 CDN 特征串和通过主动探测分析的 CDN 特征串，是一个循环迭代的过程。其中 CDN 特征串的提取，采用基于决策树算法的技术实现。具体的介绍如下文所述。

### 3.1. 域名探测

域名探测子系统由探测机、探测汇聚服务器、CDN 探测数据库组成。

域名探测子系统在全国各省部署探测节点，并且根据运营商的不同进行针对性的域名探测。对探测节点返回的探测结果进行返回汇聚到汇聚机，通过汇聚级对其整合，返回准确、完整的探测结果。将结果存入 CDN 域名库中。

图 3 所示为域名探测系统的部署图。

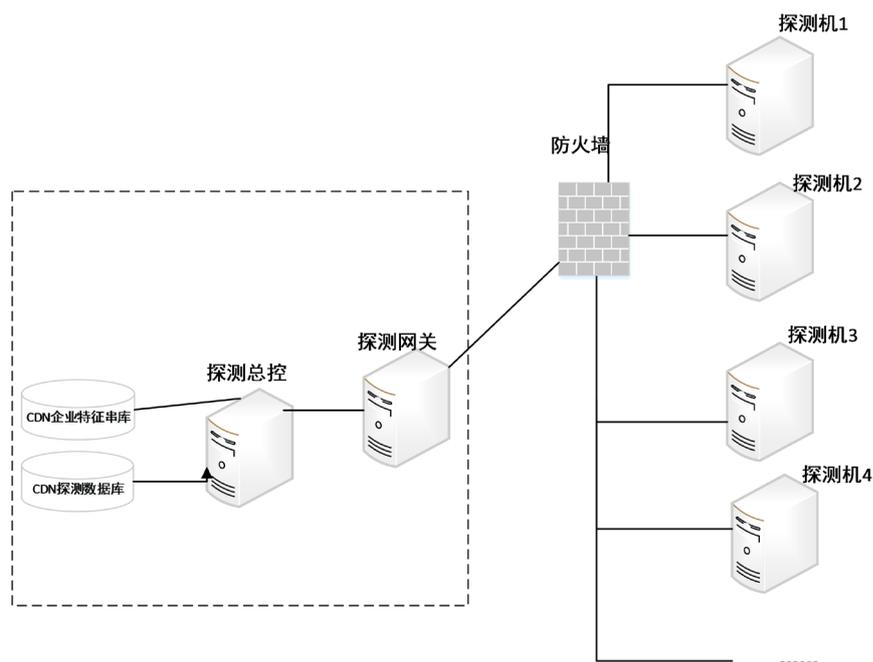


Figure 3. Deployment diagram of domain name detection system

图 3. 域名探测系统的部署图

域名探测子系统获取对域名探测解析的记录值，需要实时、长期的对被探测域名的解析。从而获取大量的原始数据，作为提取特征串的输出。

### 3.2. CDN 企业特征串分析

对 CDN 探测数据库中的数据进行数据分析，针对域名的 CNAME 记录进行分析。提取 CDN 特征串

特征, 使用基于决策树的方法判定是否属于 CDN 特征串。

1) CDN 域名特征

针对域名的解析记录与统计的思想, 把 CDN 域名的特征总结分为 6 个属性, 特征域名对应的加速域名频次(SpeedNum)、特征域名的三级域名频次(ThreeNum)、特征域名的边缘节点频次(IPNum)、特征域名是否包含 cdn (CdnString)、是否使用 CDN 服务(CdnService)、顶级域是否为通用顶级域名(TopName)。采用这 6 个属性对 CDN 域名的特征进行刻画。

2) 决策树训练过程

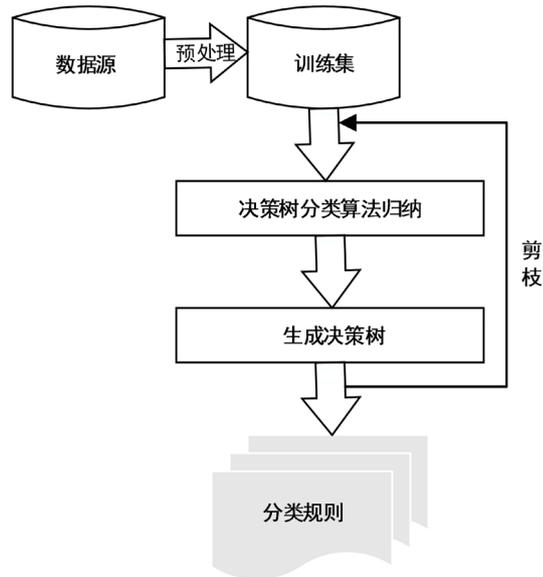


Figure 4. Decision tree overall training process diagram  
图 4. 决策树总体训练过程图

决策树总体训练过程如图 4 所示:

a) 数据预处理, 对数据源进行数据预处理, 因 CDN 必定使用 CNAME, 所以只提取 CNAME 记录。并且对 CNAME 记录进行去重。

b) 设  $S$  是  $s$  个数据样本的集合。假定类标号属性具有  $m$  个不同值, 定义  $m$  个不同类  $C_i (i=1, 2, \dots, m)$ 。设  $s_i$  是类  $C_i$  中的样本数。对一个给定的样本分类、所需要的期望信息如下:

$$I(s_1, s_2, \dots, s_m) = -\sum_{i=1}^m p_i \log_2(p_i) \tag{1}$$

其中  $p_i$  是任意样本属于  $C_i$  的概率。

c) 设属性  $A$  具有  $v$  个不同值  $(a_1, a_2, \dots, a_v)$ 。用属性  $A$  将  $S$  划分为  $v$  个子集  $(S_1, S_2, \dots, S_v)$ , 设  $S_{ij}$  是子集  $S_j$  中类  $C_i$  的样本数。由  $A$  划分成子集的熵表示如下

$$E = \sum_{j=1}^v \frac{S_{1j} + S_{2j} + \dots + S_{vj}}{S} (S_{1j}, S_{2j}, \dots, S_{vj}) \tag{2}$$

d) 在  $A$  分枝将获得的信息增益标识为:

$$\text{Gain}(S, A) = I(s_1, s_2, \dots, s_m) \tag{3}$$

e) 用信息增益率进行属性选择, 信息增益率定义为:

$$\text{GainRatio}(S, A) = \frac{\text{Gain}(S, A)}{\text{SplitInfo}(S, A)} \quad (4)$$

分裂信息  $\text{SplitInfo}(S, A)$  代表了按照属性  $A$  分裂样本集  $S$  的广度和均匀性。

### 3) 决策树算法的种类选择

目前决策树的典型算法有 ID3 [14]、C4.5 [15]、J48 [16]、CART [17] 等，不同的决策树会影响系统判断的准确度。首先对建立好的训练集进行预处理，优化属性。之后对训练集进行分类回归。采用留余法，将数据集分成 10 份，轮流将其中的 9 份做为训练数据、1 份做为测试数据进行试验。每次试验都会得到相应的正确率，将 10 次结果正确率的平均值做为对算法精度的估计。运用不同的决策树算法进行训练，根据设置的实例情况，共选择了 5 种决策树进行对比分析，试验结果如图 5 所示，结果表明，J48 决策树算法的正确率最高，所选属性集为最优的属性组合，其正确率为 93.38%。

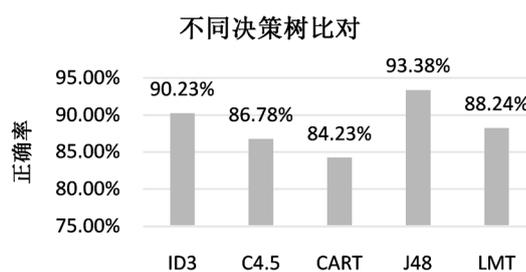


Figure 5. Comparison chart of correct rate of various decision trees  
图 5. 各类决策树正确率对比图

### 3.3. CDN 域名分析

探测系统实时对域名进行探测，数据库中更新域名解析结果，结合 CDN 企业特征串、CNAME 记录进行 CDN 域名分析。若 CNAME 记录符合 CDN 特征串，则此域名经过了 CDN 加速，并记录此 CDN 加速节点。

CDN 域名分析流程图如图 6 所示。

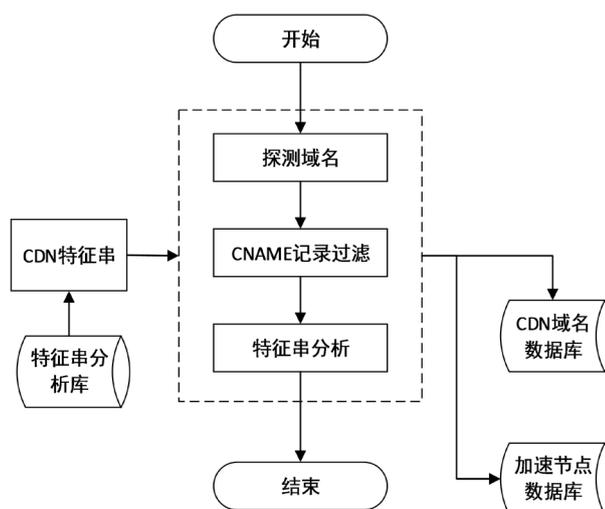


Figure 6. CDN domain name analysis flowchart  
图 6. CDN 域名分析流程图

通过特征串分析库提供的 CDN 特征串,对探测发现的域名进行过滤、分析,将最终的分析结果,CDN 域名、CDN 加速节点存入 CDN 基础资源库中(CDN 域名数据库、CDN 加速节点数据库)。

## 4. 实验结果与分析

### 4.1. 实验分析

本文为了训练和评估决策树模型,测试集选取网络收集域名 1000W,并通过域名探测系统针域名进行全国多节点探测,获得域名解析记录。该记录包含 A 记录、CNAME 记录、MX 记录等。通过已知 CDN 特征串确认 CDN 域名。加速域名较少、边缘节点 IP 较少等确认为非 CDN,从而进行模型建立。表 2 是采集的部分 CDN 特征串。

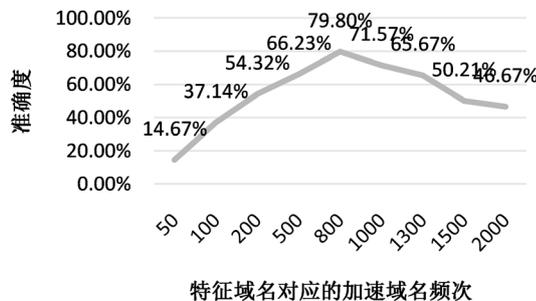
**Table 2.** Known CDN enterprise characteristic string  
**表 2.** 已知 CDN 企业特征串

CDN 企业	省份	特征串
腾讯云计算有限责任公司	北京	dnsv1.com spcdntip.com cdntip.com tcdnvod.com ovscdns.com tcdnlive.com
北京金山云网络技术有限公司	北京	download.ks-cdn.com
大连致远信息科技有限公司	辽宁	hirpo.com
北京爱奇艺科技有限公司	北京	iqiyiedge.com
上海帝联网络科技有限公司	上海	bbbcdns.com
上海逸云科技发展有限公司	上海	scsdns.com
上海云熵网络科技有限公司	上海	miitsp.com
上海哎娘娘网络科技有限公司	上海	ulute.net
上海聆云信息技术有限公司	上海	lingyuncloud.com

构建数据集分别针对特征域名对应的加速域名频次(SpeedNum)和特征域名的边缘节点频次(IPNum)的阈值对准准确度的影响。

#### 1) 加速域名频次对准准确率的影响

分析特征域名对应的加速域名频次阈值依次在 50 至 2000 对 CDN 域名准确度的影响,如图 7 所示。



**Figure 7.** The impact of accelerated domain name frequency on accuracy  
**图 7.** 加速域名频次对准准确率的影响

从表中可以看出准确率随着特征域名对应的加速域名频次阈值增加而增加,当阈值为 800 左右时,准确率最高,达到了 79.80%。而在阈值为 100 以下时,准确率非常的低。

说明如果此特征串为 CDN 特征串,则有很多加速域名会指向此特征域名。当指向的特征域名比较少的时候,大概率不是 CDN 特征域名。

#### 2) 边缘节点频次对准确率的影响

分析特征域名的边缘节点频次阈值依次在 10 至 200 对 CDN 域名准确度的影响,如图 8 所示。

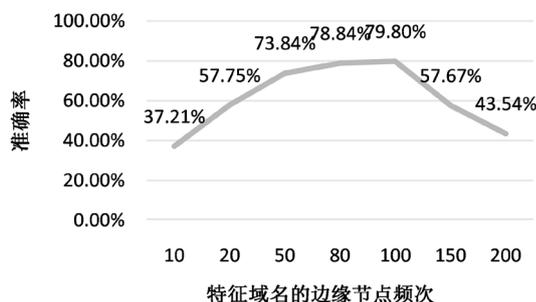


Figure 8. The influence of edge node frequency on accuracy

图 8. 边缘节点频次对准确率的影响

从表中可以看出准确率随着特征域名对应的边缘节点频次阈值增加而增加,当阈值为 90 左右时,准确率最高,在 10 以下与 200 以上的准确率比较低,达到 40% 以下。

说明如果此特征串为 CDN 特征串,则此特征域名的解析结果(边缘节点)大概率存在多个,全国多个 CDN 加速节点为其提供 CDN 加速服务。

## 4.2. 评估指标

为量化评估决策树模型针对域名测是否为 CDN 的检测能力。本文采用如下 2 个评估指标:

1) 准确率(Precision): 是分类结果中的某类别判断正确的文档中有多少是真正的正样本的比例,是针对预测结果而言的,衡量的是分类系统的查准率。

2) 召回率(Recall): 是原来某个类别的文本的分类结果中有多少被预测为正确的比例,是针对原来样本而言的,衡量的是分类系统的查全率。

## 4.3 结果分析

取 CDN 域名主要影响特征的特征域名对应的加速域名频次 N、边缘节点频次 K 做为变量分别根据准确率、召回率做比较。根据不同的 N、K 值的检测结果如表 3 所示。

Table 3. Accuracy and recall rates of different combinations of N and K values

表 3. 不同 N、K 值组合的准确率、召回率

参数 N	参数 K	准确率(%)	召回率(%)
800	50	98.3	95.3
800	80	98.7	97.6
800	100	98.6	97.8
1000	50	97.7	95.1
1000	80	98.8	98.1
1000	100	99	97.9

从表 3 中可以分析得出, 加速域名频次  $N$  和边缘节点频次  $K$  值的不同组合, 对 CDN 准确率、召回率的影响较大, 当  $N = 1000$ ,  $K = 80$  时, 准确率达到 98.8%, 召回率达到 98.1%。J48 的决策树分类模型在该参数下达到最优解。

针对探测获取的域名进行分析, 通过本研究方法经过近两个月的分析。目前收录 CDN 企业 195 家, CDN 特征串 367 个。部分特征串如表 4 所示。

**Table 4.** Part of enterprise characteristic string  
**表 4.** 部分企业特征串

企业	特征串
腾讯云	dnsv1.com, spcdntip.com, cdntip.com, tcndvod.com, ovscdns.com, tcndlive.com
北京金山云	download.ks-cdn.com
北京动力在线	powercdn.cn, iproot.com, y163.net, y163.org, fcdn.cn
北京快网科技	cloudglb.com
北京蓝汛通信	ccgslb.net, chinacache.net, ccgslb.com.cn, lxsvc.cn
北京魔方互联	morefuntech.com
北京晟元亿讯	h2comm.com.cn
广州云硕	cloudcdscdn.cn
上海纪耀网络	ji-yao.com
北京友普	urcloudcdn.com
浪潮软件	inspurcloud.cn
广州酷狗	kugou.com

表 3 中的特征串, 可以作为 CDN 特征串, 构建 CDN 基础资源库的组成部分。

## 5. 结论

域名系统(Domain Name System)将便于用户识别的网络资源域名转换为机器可识别的 IP 地址, 是十分重要的网络基础设施, 是互联网的基石。内容分送网络技术是新兴的网络加速技术, 可以避开互联网上影响数据传输速度和稳定性的瓶颈和环节, 使内容传输更快、更稳定[18] [19] [20] [21]。识别 CDN 域名在域名分析领域有着十分重要的作用。针对域名管制发挥着重要的作用, 针对目前国内域名混乱使用 CDN 服务的情况有针对性的辅助作用, 并提高对 CDN 厂商的监管力度。本文通过采用基于 J48 决策树的 CDN 检测技术, 构建 CDN 基础资源库, 从而高效的识别和检测 CDN 域名。为基于 CDN 技术的网络安全应用和管理提供了重要的依据, 具有广泛的应用前景。

## 参考文献

- [1] 李聪颖, 王瑞刚, 梁小江. CDN 技术的研究与设计[J]. 物联网技术, 2015, 5(12): 28-30.
- [2] 熊明. CDN 技术研究及在宽带中的应用[D]: [硕士学位论文]. 天津: 天津大学, 2015.
- [3] 蒋杰. CDN 系统的关键技术[J]. 数字通信世界, 2018(8): 12-13.
- [4] 田光辉. 移动内容分发网络节点位置部署建模及部署方案研究[D]: [硕士学位论文]. 北京: 北京邮电大学, 2013.
- [5] 张戈. 基于 CDN 的视频网络架构研究[J]. 电脑编程技巧与维护, 2018(12): 161-163.
- [6] 唐宏, 陈戈, 陈步华, 余媛. 内容分发网络原理与实践[J]. 电信科学, 2018, 34(11): 181.

- 
- [7] 郎丰凯. CDN 技术及发展趋势分析[J]. 电子世界, 2019(14): 106.
- [8] 王海洋, 赵建福, 韩增辉, 王晟. 分布式 CDN 带来的挑战与应对方案[J]. 山东通信技术, 2018, 38(1): 22-25.
- [9] 乔爱锋. CDN 体系架构及运营部署方案[J]. 电信快报, 2018(10): 17-21.
- [10] 罗明. 基于移动边缘计算的网路信息监测与缓存业务优化[D]: [硕士学位论文]. 北京: 北京邮电大学, 2019.
- [11] 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.
- [12] 薛景安. 大规模内容分发网络客户端映射性能测量与优化研究[D]: [博士学位论文]. 北京: 清华大学, 2018.
- [13] 徐翔, 任昌燕, 管黎晨, 等. 一种高效的全球 DNS 智能调度系统研究[J]. 新一代信息技术, 2019, 2(22): 33-38.
- [14] Jin, C., De-Lin, L. and Mu, F.-X. (2009) An Improved ID3 Decision Tree Algorithm. 2009 4th International Conference on Computer Science & Education. Nanning, 25-28 July 2009, 127-130.
- [15] 韩存鸽, 叶球孙. 决策树分类算法中 C4.5 算法的研究与改进[J]. 计算机系统应用, 2019, 28(6): 198-202.
- [16] Mathuria, M. (2013) Decision Tree Analysis on J48 Algorithm for Data Mining. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3, No. 6.
- [17] Lewis, R.J. (2000) An Introduction to Classification and Regression Tree (CART) Analysis. *Annual Meeting of the Society for Academic Emergency Medicine in San Francisco, California*, 14.
- [18] Al-Abbasi, A., Aggarwal, V., Lan, T., et al. (2019) FastTrack: Minimizing Stalls for CDN-Based Over-the-Top Video Streaming Systems. *IEEE Transactions on Cloud Computing*, 1, 1. <https://doi.org/10.1109/TCC.2019.2920979>
- [19] Helt, J., Feng, G., Seshan, S., et al. (2019) Sandpaper: Mitigating Performance Interference in CDN Edge Proxies. *Proceedings of the 4th ACM/IEEE Symposium on Edge Computing*, Arlington, 7-9 November 2019, 30-46. <https://doi.org/10.1145/3318216.3363313>
- [20] Flores, M. and Bedi, H. (2019) Caching the Internet: A View from a Global Multi-Tenant CDN. In: *International Conference on Passive and Active Network Measurement*, Springer, Cham, 68-81. [https://doi.org/10.1007/978-3-030-15986-3\\_5](https://doi.org/10.1007/978-3-030-15986-3_5)
- [21] Alabbasi, A.O. (2020) A Quantitative Framework for CDN-Based Over-the-Top Video Streaming Systems. Purdue University Graduate School, West Lafayette.