学士学位论文

基于异质信息网络的恶意 DNS 域名识别

学　　号：**20181000645**

姓　　名：**朱靖宇**

学 科 专 业：**信息安全**

指 导 教 师：**张锋 副教授**

培 养 单 位：**计算机学院**

二〇二二年五月

中国地质大学（武汉）学士学位论文原创性声明

本人郑重声明：本人所呈交的学士学位论文《基于异构信息网络的恶意 DNS 域名检测》，是本人在指导老师的指导下，在中国地质大学（武汉）攻读学士学位期间独立进行研究工作所取得的成果。论文中除已注明部分外不包含他人已发表或撰写过的研究成果，对论文的完成提供过帮助的有关人员已在文中说明并致以谢意。

本人所呈交的学士学位论文没有违反学术道德和学术规范，没有侵权行为，并愿意承担由此而产生的法律责任和法律后果。

学位论文作者签名：_____

日 期：　　年　　月　　日

摘要

域名系统（Domain Name System）是因特网的核心组成部分，用于在域名和 IP 地址之间建立映射关系。近年来，许多恶意域名被攻击者注册，用以进行网络攻击，损害了正常用户的利益，极大地威胁到网络空间安全。因此，如何高效准确地检测出恶意域名，成为了研究者关注的重点。

最初，研究者希望利用黑名单的方法来检测恶意域名，但是黑名单难以跟上网络环境发展的速度，无法及时覆盖到正在活跃的恶意域名。后来人们训练机器学习模型来检测恶意域名，但是这些方法仅考虑了域名自身的特征，特制的恶意域名就可以绕过检测。为了解决上述问题，研究人员着眼于域之间的联系。他们将 DNS 场景建模成图，从中挖掘各域间的联系，用于检测。然而，鉴于 DNS 场景的复杂性，这些检测方法仍不能很好地挖掘域之间的联系。

为了解决上述检测方法中出现的问题，本文提出一种基于异质信息网络的恶意域名检测方法 HANdom。该方法首先将 DNS 场景表示为含客户端、IP 地址、CNAME 记录和域名等实体的异质信息网络（Heterogeneous Information Network，HIN）；接着利用基于元路径的异质图卷积网络生成嵌入向量，并利用语义级别注意力机制来学习各元路径的重要性；最后利用全连接神经网络来对域名节点进行二分类，完成恶意域名检测任务。

本文利用 DataCon 开源数据集来对 HANdom 恶意域名检测技术进行测试，验证了提取的元路径的有效性。同时，本文将 HANdom 方法与传统的机器学习方法和其它三种典型图卷积模型进行比较。实验结果表明，本文提出的 HANdom 方法在指标上取得了最好表现。HANdom 在以 8:2 划分训练集和测试集的情况下，达到了 95.4%的检测精度。

关键词，恶意域名检测，异质信息网络，异质图卷积，元路径，注意力机制

Abstract

Domain Name System (DNS) is the core component of the Internet, used to build the mapping relationship between domain names and Internet Protocols (IPs). Unfortunately, in recent years, many malicious domain names have been registered by attackers for cyber-attacks, which damages the interests of regular users and significantly threatens network security. Therefore, how to efficiently and accurately detect malicious domain names has become the focus of researchers.

Initially, the researchers wanted to use the blacklist method to detect malicious domain names. However, the blacklist is difficult to keep up with the development of the network environment and cannot cover the active malicious domain names in time. Later, machine learning models were trained to detect malicious domain names, but these methods only consider the features of the domain names themselves and can be bypassed by well-crafted attacks. To tackle the above issues, the researchers focused on the connection between domains. They modeled the DNS scene as a graph and then mined associations among domains for detection. However, given the complexity of the DNS scene, some detect methods cannot fully excavate relations between domains.

To address the limitations in the above detection methods, this thesis proposed a new detection method HANdom, which is based on a heterogeneous information network. In this method, the DNS scene is represented as Heterogeneous Information Network (HIN) with diverse entities like clients, IP addresses, CNAME records and domains. Then HANdom uses meta-path-based Heterogeneous Graph Convolutional Networks (HGCN) to generate embedding vectors and uses the semantic-level attention mechanism to learn the importance of each meta-path. Finally, the fully connected neural network is used to double classify the domain nodes and complete the task of malicious domain name detection.

Finally, this thesis uses DataCon open-source dataset to test the Handom method and verify the validity of the extracted meta-path. At the same time, the Handom method is compared with traditional machine learning methods and three other typical graph convolution models. Experimental results show that the proposed HANdom

method achieves the best performance on metrics. HANdom achieved a detection accuracy of 95.4% when dividing the training and test set at 8:2.

Key Words, Malicious domain detection, Heterogeneous information network, Heterogeneous Graph Convolutional Network, Meta-path, Attention Mechanism

目 录

第一章 引言

1.1 研究工作背景和意义

域名系统（Domain Name System，DNS）是一种在全球范围内提供互联网服务的分布式协议[1]。该协议主要被用于在域名和 IP 地址之间建立联系，将便于记忆的域名转为数字化的 IP 地址。该协议是目前网络体系中的骨干部分，除了静态 IP 通信以外，任意主机和互联网之间的通信都必须使用该协议。考虑到域名在网络环境中的重要性，可以将其看作攻击者进行网络恶意行为所需的重要资源。比如，攻击者会利用域名来创建僵尸网络中的 C&C（command and control）通信。因此，如何有效地区分和屏蔽参与恶意行为的域名是网络安全研究中的热点课题之一。

早期，网络管理员们利用黑名单来检测恶意域名。当时，恶意域名的发现主要依赖于人工的分析。一些专业的域名黑名单网站，如 Malwaredomains.com[1]等，可以覆盖绝大部分活跃的恶意域名。网络管理员们通过将网络环境中的域名与恶意域名进行简单的对比，就可以检测出当前环境中的恶意域名，并及时阻断恶意域名与被感染主机之间的通信。

但是，域名生成算法（Domain Generation Algorithm，DGA）的提出，加快了恶意域名的生成速度，恶意域名的生存周期也大为缩短，意味着恶意域名和 IP 地址之间的映射关系持续时间更加短暂，也更加难以被及时发现，这样，攻击者就可以加以利用，规避黑名单检测。因此，域名黑名单网站不仅无法及时收集恶意域名，也无法对当前正在活跃的恶意域名高度覆盖。

因此，研究者将目光投向了基于特征的机器学习检测方法，通过从 DNS 的流量和日志中得到 DNS 的特征，然后利用机器学习方法进行检测。这些方法实现了对恶意域名的快速检测，取得极大成功。然而，有经验的攻击者可以通过伪造域名的相关特征来避开检测，如控制生成的恶意域名的字符分布，控制通信报文发送间隔，等等。

为了解决基于特征的检测方法中出现的问题，一些研究者将目光转向了域名

[1] Malwaredomains.com, 2019. DNS-BH - Malware Domain Blocklist by RiskAnalytics. http://www.malwaredomains.com. [Online].

之间的联系，他们发现攻击者难以伪造域名间联系。因此，他们将 DNS 场景建模成一幅图，并通过挖掘域名之间的联系来进行检测。如何准确、高效地对 DNS 场景建模，并有效利用域名之间的联系来检测出恶意域名，成为了当前恶意域名检测研究的重要方向。

1.2 国内外研究历史与现状

随着互联网的发展和网络攻击手段的不断更新，恶意域名检测方法也在不断进步，本文按照检测方式的不同将恶意域名的检测方法划分为基于黑名单的检测方法、基于特征的检测方法和基于图的检测方法三大类别。

1.2.1 基于黑名单的检测方法

在最初，网络规模较小，攻击手段落后单一的情况下，恶意域名的规模和变化都较小，用黑名单的方式就可以满足检测恶意域名的要求。这时候，恶意域名的发现主要依靠用户和安全分析人员向专业的网站提交恶意域名相关信息，然后由专业人员来管理和维护黑名单，如 360 安全服务恶意网址举报平台[2]等。企业网络安全管理人员可以从各大专业平台获取黑名单并设置拦截规则，以过滤企业内部人员对恶意域名的访问。虽然，目前基于黑名单的检测方法依然是最有效的检测方法之一，但是由于互联网规模的不断壮大和攻击者使用 DGA 和 Fast-Flux 等方法，该检测方法已经无法实时覆盖所有恶意域名。

1.2.2 基于特征的检测方法

研究者们提出该类方法用以应对数据量爆发式增长的恶意域名，该类方法的特点是从大量 DNS 流量和日志文件中提取出特征，然后将这些特征与机器学习算法或者深度学习算法相结合，构建恶意域名分类器。

在 2011 年，Bilge 等人[2]通过对 DNS 流量分析提出 EXPOSURE 模型。该模型提取了时间特征，DNS 响应特征等共 15 维特征，然后结合决策树分类算法得到分类模型。该方法使用了多角度信息特征区分恶意域名与正常域名，准确度较高。

考虑到主机 DNS 活动完整性，Manmeet Singh 等人[3]提出基于 DNS 指纹的检测方法。该方法着眼于指定时间窗口内的 DNS 活动特征，主要包括 DNS 请求特征，域名特征，DNS 响应特征以及 FQDN-IP 特征。最后，通过随机森林算法训练

[2] 360.360 安全服务恶意网址举报平台[EB/OL].https://fuwu.360.cn/jubao/wangzhi,May 10,2022

出 DNS 分类器。Yan 等人[4]提出一种 APT 攻击检测框架 AULD。该框架主要关注 DNS 的时域特征，通过自动无监督的机器学习对大量 DNS 日志信息进行分析，并得到一组可疑域名。

此后，为了提高 DNS 流量检测效果，研究者将目光转向多种检测方法相结合。Lu Huang 等人[5]通过结合多种检测手段来检测恶意域名。他们首先利用白名单来过滤正常流量，再从剩下的 DNS 流量中提取包括 C&C 特征、主机特征、IP 特征、请求序列特征以及请求时间特征，最后使用 BP 算法检测出恶意域名。Shaojie Chen [6]等人基于 FQDN 中的次级域和顶级域对 DNS 包进行分类。他们将 FQDN 的子域提取出，转化为词向量后依次喂入 LSTM 网络，实现端到端的检测，接着使用基于域分组和白名单过滤方式过滤掉非隐蔽的 DNS 通道流量和合法的 DNS 隐蔽通道流量，以降低假阳性率。

此后，随着 DGA 算法在恶意域名生成过程中的大量使用，研究者在提取特征时增加了对域名词特征的关注。刘浩杰等人[7]提出了利用集成学习的方法进行恶意域名检测。该方法提取了域名的域名长度、域名元辅音字符占比等统计特征以及 N-gram 衍生特征，并集成了 HMM、朴素贝叶斯算法和 LSTM 算法合成了恶意域名检测系统。Zhouyu Bao 等人[8]提出了一个 DGA 域名检测系统，该系统首先从被动 DNS 数据集和域名黑白名单中提取域名的字符特征和域名访问活动的特征，接着使用 word2vec 进行训练并生成域名特征向量，最后利用域名特征向量进行 DGA 分类。

1.2.3 基于图的检测方法

虽然基于特征和机器学习的方法在评估过程中表现出了高准确性，但是它们依旧可以被攻击者绕过。攻击者可以通过改变 FQDN 的字符分布[9]，控制查询的时间间隔等方式使 DNS 恶意域名的特征接近正常域名的特征。但是一些研究者发现，攻击者难以伪造 DNS 域之间的联系。于是他们采取基于图的方式来检测 DNS 中的恶意域，将问题转化为对图中节点进行二分类的问题。

最开始，研究者们使用二部图来进行恶意域名的检测。Mohamed Nabeel 等人[10]通过建立域和 IP 的二部图，利用 IP 向域发送查询请求这一行为建立域之间的联系。他们对私有 IP 和公共 IP 分别进行了考虑，并将被同一个私有 IP 查询和在一个时间窗口内被多个公共 IP 共同查询作为域之间的联系。Segugio[11]主要关注发出 DNS 请求的主体以及 DNS 请求内容，该研究者通过监控 DNS 流量，构建了主机-域名查询图，图中的边表示主机查询域名的关系。Segugio 在查询图中标记出已知的良性或恶意节点后，对图进行修剪以减少噪声，再从图中提取机器学习、域名活动、

IP 滥用三种类型特征放入分类器训练，最终得到恶意域名检测模型。Futai Zhou 等人[12]基于 IP 和域名的映射关系构建了 DNS 查询映射图以及基于域名 CNAME 关系的被动 DNS 图，并在图上结合先验知识、MRF、BP 算法推断图中恶意域名。Khalil 等人[13]构建了 IP 和域名的映射图，并通过统计两个域名之间共享 IP 地址的个数，将 IP 和域名的映射图转化为域名和域名的关联图，最后根据域名关联图中的权值关联等信息推断出恶意域名。

更进一步，一些研究者考虑将 DNS 场景建模成异构图。HInDom[14]利用 DNS 流量中的信息构建了异质信息网络，并从中提取了 6 条元路径，最后结合 Tranductive Classifier 从图中检测出恶意域名。在此基础上，DeepDom 系统[15]更进一步考虑了 DNS 场景中的更多联系，包括被相同 IP 查询，有相同 CNAME，在 WHOIS 数据库中有相同注册者等等。他们依据这些联系建立了异质信息网络，并利用元路径引导的短随机游走和图卷积网络为每个 DNS 域节点生成一个用于分类的嵌入向量。

根据上述分析，我们可以发现，基于黑名单的恶意域名检测方法依然是一种有效的检测方法，但是随着 DGA 和 Fast-Flux 攻击技术的出现，黑名单已经无法及时覆盖恶意域名。人工智能技术的发展为基于特征的恶意域名检测方法提供了基础，其主要是提取恶意域名相关特征后喂入深度学习或者机器学习模型进行训练。从目前的研究来看，这类方法具有良好检测效果。但是，该类方法仅关注单个域名的特征，未关注域名间联系。此外，该类方法还需要大量数据进行模型训练，当攻击者采用新的攻击算法或者刻意模仿正常恶意域名的特征时，该类模型就需要考虑新的特征，训练新的检测模型。这在现实检测中将耗费巨大算力和时间，并导致用户在新的模型生成前蒙受损失。而基于图的检测方法则是考虑到了域名之间的联系，这种联系是在通讯需求中提取的。当遭遇新的攻击算法时，新型攻击中域名间的联系特征依然不会发生改变。因此，基于图的检测方法在近几年成为了域名检测的热点之一。本文中所提出的基于异质信息网络的恶意域名检测方法正是基于图的检测技术，在本方法中提取了客户端查询行为，域名-IP 映射关系，域名 CNAME 关系等信息构建异质信息网络，并在构建的异质信息网络上利用异构图卷积技术进行恶意域名检测。

1.3 研究目的和贡献

这篇论文的目的如下所示：利用异质信息网络对 DNS 场景进行建模，以利用

域名之间的联系来进行恶意域名识别。

本文的主要贡献如下：

1）提出一种新的基于异质信息网络的恶意 DNS 域名检测系统 HANdom，该检测系统引入异质信息网络对 DNS 场景进行刻画，通过元路径建立 DNS 域之间的主要联系。

2）提出了一种新的带注意力机制的图卷积神经网络，利用语义级别注意力机制和 GAT 卷积层，完成异质信息网络上的聚类学习，并同时处理其节点特征和图结构特征。

3）从多角度考虑了 DNS 域名的特征，通过域名本身特征和 DNS 解析特征两方面入手，提高检测准确度。

1.4 文章结构

这篇文章的结构如下：在第二章中讨论了本文提出的方法中使用的相关理论基础。在第三章中呈现了对 DNS 流量的分析。在第四章中展示了基于异质信息网络和图卷积网络的检测模型。对该检测模型的实验和测试将在第五章呈现。本文在第六章中对全文进行了总结，并提出了一些不足和有待改进部分。

第二章 相关理论基础

2.1 DNS 相关知识

2.1.1 DNS 基础

DNS 协议是网络架构中的骨干部分，大多数网络会话的进行依赖于 DNS 协议。该协议被定义为一种用于在客户端和服务器之间交换名称和 IP 地址（Internet Protocol Address）的协议。DNS 协议可以看作一个有层次的分布式的数据库，其中存放了一组注册过的主机的信息。该协议中可以存放不同类型的数据，主要为 IP 地址（IPv4 和 IPv6）和域名。

为了向全球提供服务，DNS 协议中数据的存储带有一定结构，该结构被称作 DNS 命名空间。DNS 命名空间是 DNS 协议中的一种树型结构。在一个域名中的每一个点（.），表示了在树形结构中层次间的分隔。DNS 树型结构最高层代表以点开始的根级别。顶级域（TLD，Top-Level Domain）是根节点下的子节点，如 .com，.net 等。顶级域也有子域表示权威域名服务器或者二级域名(2LD，Second-Level Domain)。最后，完全限定域名 FQDN 确定了主机或者 DNS 层级结构中的子域[16]。例如，在查找域名 123.mydomain.com. 的过程中，通常是从在树型结构中表示根级别的点号开始。接着，根服务器将请求转发到下属顶级域中表示 com.的域名服务器。然后，顶级域将查询转发到二级域。权威服务器在收到查询后查找到 FQDN 对应的目标主机并返回其 IP 地址。图 1.1 表示 DNS 域名结构。

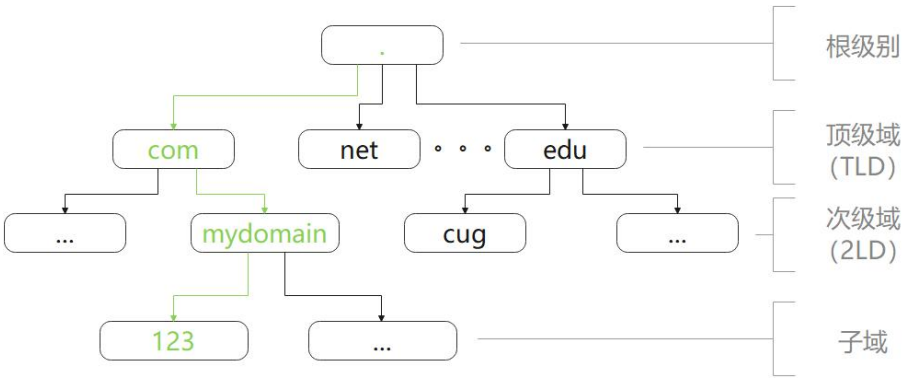


图 1.1 DNS 层次结构

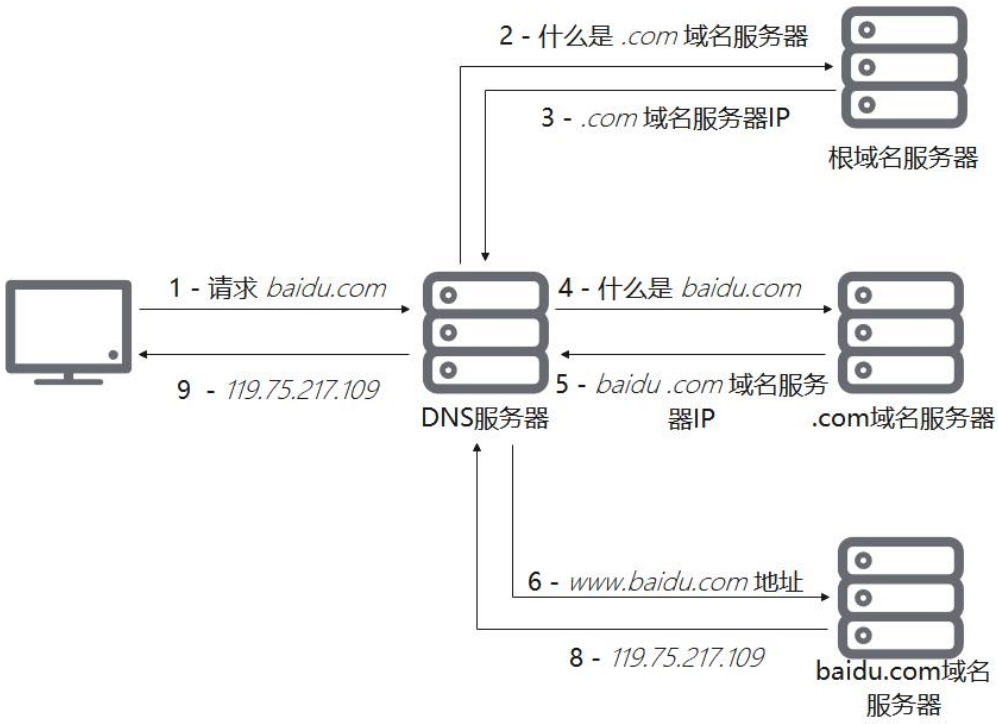


图 1.2 DNS 工作过程

2.1.2 域名解析

DNS 协议的目的是在便于记忆的域名和数字化的 IP 地址之间建立相互映射关系。图 1.2 表示了 DNS 的工作过程，以用户在浏览器中键入 *baidu.com* 为例。首先，浏览器在本地缓存中查找域名对应的 IP 地址。如果没有查找到 IP 地址，浏览器将把查询转发到本地 DNS 服务器或者网络服务提供商（ISP，Internet Service Provider）的 DNS 服务器。若依旧没有找到对应 IP 地址，则本地 DNS 服务器将请求转发到根域名服务器。根域名服务器分布在世界各地，并指向合适的下级域名服务器。然后，请求将会被转发到顶级域 *.com* 对应的域名服务器上，该服务器中存储了次级域 *baidu.com* 对应的 IP 地址。接着，DNS 服务器将向 *baidu.com* 对应的域名服务器查询 *www.baidu.com* 并得到其 IP 地址。最后，得到信息的 DNS 服务器将会响应浏览器的请求，返回 IP 地址。

2.2 异质信息网络

考虑到现实世界中的多样性，将一个系统简单的用同构图来表示显然是不合理的。在这种情况下，Sun 等人[17-18]提出异质信息网络（Heterogeneous Information

Network，HIN）用于融合更完善的特征，更全面的表现出现实系统中不同类型的组件和其间的联系。下面列出了异质信息网络的一些基本概念。

定义 1：异质信息网络[17]。给定一个图$G=<V,E,T,X>$，其中 V 代表节点集合，E 代表边集合，$\mathrm{X}=\{\mathrm{x_i}\mid \mathrm{v_i}\in \mathrm{V}\}$表示节点上的特征集合。在图中存在函数$\varphi:V\to T_V$和$\phi:E\to T_E$表示节点/边到其对应类型的映射，这里$T_V$和$T_E$分别表示节点/边的类型集合。如果$|T_E|+|T_V|>2$，则可以将图 G 称为异质信息网络，并且将$T=<T_E,T_V>$称为 G 的网络模式。

定义 2：元路径[18]。给定一个异质信息网络$G=<V,E,T,X>$，元路径 P 就是节点$\mathrm{V_1}$到节点$\mathrm{V_{L+1}}$之间的复合关系$\mathrm{R}=R_1\circ R_2\circ\cdots\circ R_L$。P 可以表示为$\mathrm{V_1}\xrightarrow{\mathrm{R_1}}\mathrm{V_2}\xrightarrow{\mathrm{R_2}}\cdots\xrightarrow{\mathrm{R_L}}\mathrm{V_{L+1}}$，在不产生歧义（如两个节点之间有多个联系）的情况下，也可以简写为$\mathrm{V_1V_2}\cdots\mathrm{V_{L+1}}$。如果一条元路径初始节点了类型和结尾节点的类型相同，即$\mathrm{T_{V_1}}=\mathrm{T_{V_{L+1}}}$，则将其称之为一条对称元路径。

简而言之，异质信息网络是一种半结构化模型，该模型在网络模式 T 的约束下表示不同类型的节点和联系。

定义 3：元路径子图提取。给定一个元路径p表示节点v_1到v_{l+1}之间的复合关系$r_1\circ r_2\circ\cdots\circ r_L$。邻接矩阵$A_p$可以表示为复数关系邻接矩阵相乘，如式 2.1 所示：

$$A_p=A_{r_1}A_{r_2}\cdots A_{r_l} \tag{2.1}$$

符合邻接矩阵A_p的图就被称为元路径子图，如果元路径p为对称元路径，则提取出的子图为同构图，否则为二部图。

2.3 图卷积网络

随着卷积神经网络在结构化的数据（图片，文本等）上取得重大成果，图卷积网络（Graph Convolution Network ，GCN）被提出并应用到了复杂的图上。GCN 的关键点在于参考节点本身的特征和图的结构为每个节点生成一个嵌入向量[19]，

用于后续的分类任务。现存的 GCN 基本可以分为两大模式：基于谱的方式和基于空间的方式。

基于谱的 GCN 模型[19-20]是从图信号处理的角度，引入滤波器来定义图卷积。该类方法基于层次传播规则和谱图理论对卷积运算重新定义，如式 2.2 所示：

$$H^{(l+1)}=\sigma(D^{-1/2}\widetilde{A}D^{-1/2}H^{(l)}W^{(l)}) \tag{2.2}$$

其中 $\widetilde{A}$ 是自连接的邻接矩阵，D 是一个对角矩阵，满足 $D_{ii}=\sum_i \widetilde{A}_{ij}$，而σ表示一个激活函数。$H^{(l)}$ 是第 i 个隐藏层的矩阵，且 $H^{(0)}$ 即为 X。

同时，基于空间的 GCN 模型来自于图片上的卷积神经网络(CNN)，如图 2.3 所示，将原本像素间直接且规则的连接推广到了节点和其邻居间的不规则联系上。基于空间的 GCN 将图卷积定义为了节点与其邻居之间的特征聚合，并通过不同的采样策略提高效率和灵活性。

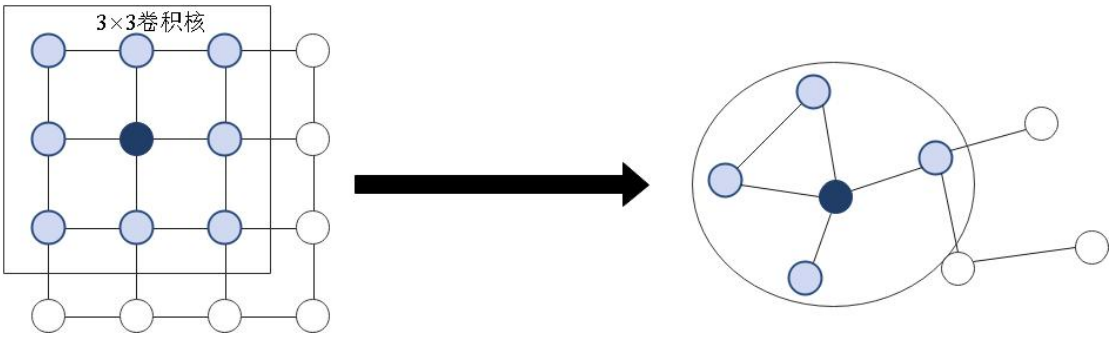


图 2.3 将像素间的连接推广到节点间的联系。每个节点的邻接点由采样函数决定

2.4 图注意力网络

图注意力网络（GAT，Graph attention network）是一种带自注意力机制的基于空间的图卷积模型，该模型由 Petar Veličković等人[21]在 2017 年提出。在该模型中，对每个节点 v_i 生成一个输出特征 $\widetilde{h}_i$，计算过程中只考虑其一阶邻居集合 N_i 中节点对 v_i 的作用，如式 2.3 所示：

$$\widetilde{h}_i=\sigma(\sum \alpha_{ij}W\widetilde{h}_j) \tag{2.3}$$

其中 $\widetilde{h}_j$ 是节点 v_i 邻居节点 v_j 的表示特征。$W\in\Re^{\widetilde{F}\times F}$ 是一个线性变换矩阵，其作用是将维度为 F 的节点特征映射到维度 $\widetilde{F}$ 上。α_{ij} 是在节点 v_i 和 v_j 间的标准化注意力系数，可以定义为：

$$\alpha_{ij}=\mathrm{softmax}_i(e_{ij})=\frac{\exp(e_{ij})}{\sum_{k\in N_i}\exp(e_{ik})} \tag{2.4}$$

节点v_i和它的邻居$j \in N_i$之间的相关系数可以由注意力系数e_{ij}来表示，其计算公式如式 2.5 所示：

$$e_{ij} = \mathrm{LeakyReLU}\left(a^T\left[\mathrm{Wh_i} \,\|\, \mathrm{Wh_j}\right]\right) \tag{2.5}$$

此外，采用多头注意力机制来扩散注意力对模型的稳定性有所提升，常见的 K 头注意力机制有拼接和均值两种计算方式。

第三章 数据分析

在本章节，本文对 DNS 合法流量和异常流量的各特征进行了分析，并确定了需要提取的特征。

3.1 域名字符串特征分析

本文分别对 DNS 域名在字符数和标签两个类别上的特征进行了比对和分析。这里本研究把 Cosico-top-1m-benion 数据集[3]用作合法域名数据。该数据集是由思科每天更新的列表，包括了前 1m 最常被查询的 DNS 域名。这里使用该数据集而不是常见的 Alex-top-1m 数据集[4]的理由在于，Alex 数据集中仅包括了顶级域 TLD，而思科中包含完整域名。而在恶意域名方面，本文使用的是 Jawad Ahmed 等人[22]在检测 DNS 渗漏时所使用的 FQDN 数据集。

3.1.1 字符数

FQDN 中的字符数是区分恶意域名的一个重要特征。因为在恶意域名往往是随机生成，这导致恶意域名中 FQDN 的字符数要大于合法的 FQDN。图 3.1 展示了不同的字符数特征在合法 FQDN 和非法 FQDN 上的分布。

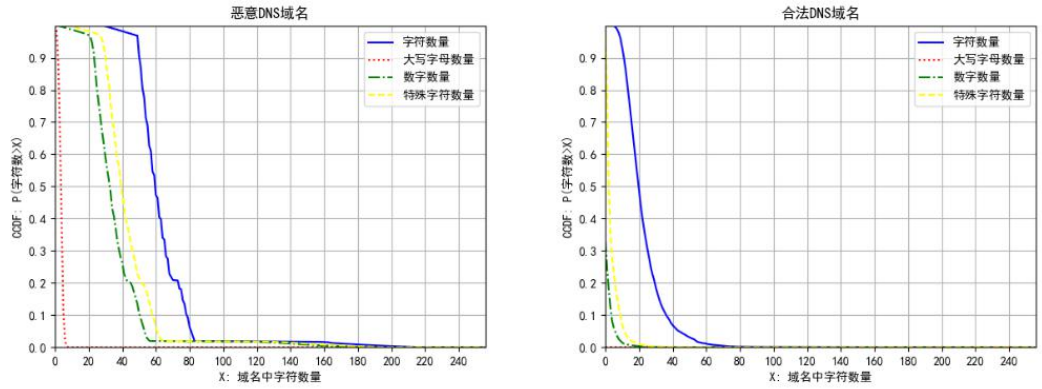


图 3.1 查询域名中字符数的 CCDF 图，(a) 恶意 DNS 域名，(b) 合法 DNS 域名

[3] Cosico-top-1m-benion. http://s3-us-west-1.amazonaws.com/umbrella-static/index.html
[4] Alexa Web Information Company. Topsites, 2020. https://www.alexa.com/topsites.

在图 3.1 中可以很容易的发现，在恶意 DNS 域名中，超过 90%的 FQDN 有超过 50 个字符，而在合法的 DNS 域名中，仅有不到 0.5%的 FQDN 总字符数超过 50。故而，本文选择将查询的域中的总字符数作为第一个特征。此外，本文还将大写字母数量和数字数量纳入特征中。这么做的理由是经过加密或编码后的数据中大写字母和数字的比例相比普通数据要更高[23]。而在图 3.1 中，大写字母和数字数量的分布也符合这个结论。除此以外，本文还将特殊字符数量纳入了考虑，这是基于在正常 DNS 域中极少存在特殊字符这一猜测。这里的特殊字符是指除了大小写字母和数字外所有出现的字符。

3.1.2 标签

这一类别由 FQDN 中标签的数量和长度两大属性构成。这是由于 DNS 恶意域名中，查询的域名倾向于符合一定格式，以添加一定辅助信息。这会导致其标签与正常 FQDN 的标签存在差异。如图 3.2 (a) 所示，对于正常的 FQDN，大部分标签数在 3 个以内，而恶意 FQDN 的标签数则在 3 个以上。但是，这并不代表标签数较多的 FQDN 必然是恶意的。有 10%左右的良性 FQDN 的标签数在 4 个及以上。其次，就长度而言，恶意 FQDN 的长度要普遍大于合法 FQDN，如图 3.2 (b) (c) 所示。这是因为攻击者需要将信息嵌入 FQDN 中，而嵌入的信息不仅包括了要传递的信息，还有如编号等附加信息。这会导致虽然在大多数情况下恶意 FQDN 的标签数要多于正常 FQDN，但是其每个标签的上的长度依旧长于正常 FQDN。

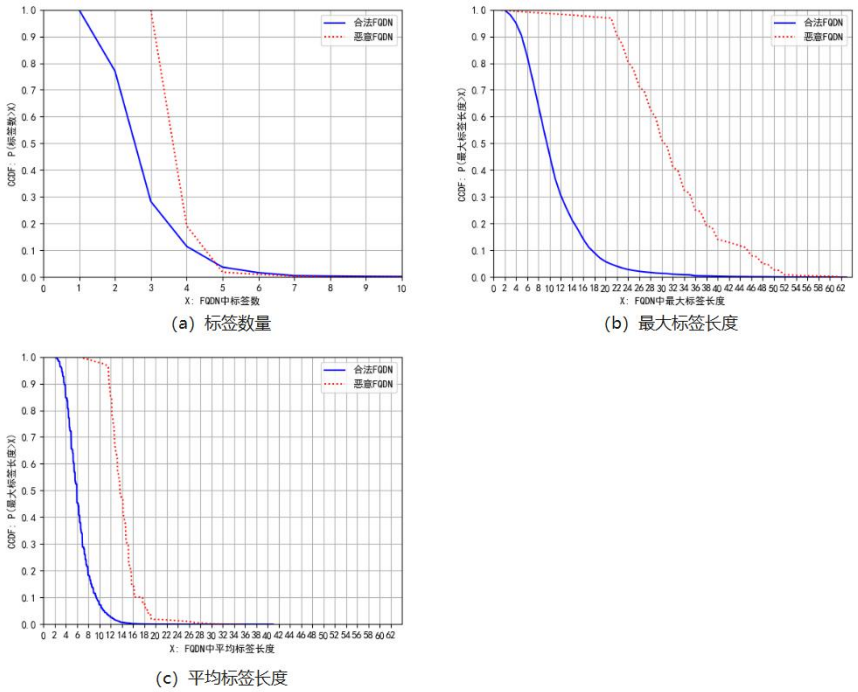


图 3.2 标签特征的 CCDF 图 (a) 标签数量 (b) 最大标签长度 (c) 平均标签长度

3.2 客户端访问行为分析

基于前人研究[24]，可以得出猜想：受感染的客户端会查询一组恶意域名，而相同攻击受害者的客户端查询的恶意域名集大概率是重叠的，因为这些被感染的客户端的查询行为由相同技术产生。此外，正常的客户端几乎不会取查询这些恶意域名集合，因为大部分恶意域名并不会提供正常服务。

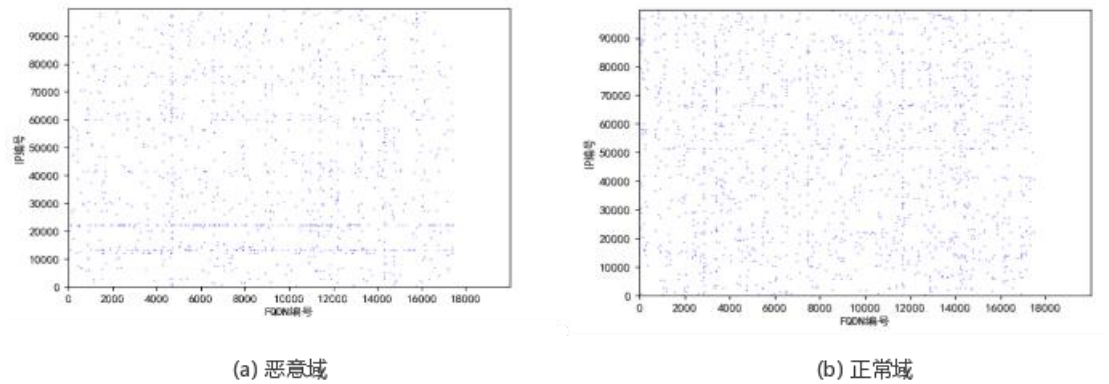


图 3.3 客户端查询行为

图 3.3 是客户端的查询行为，其数据来自于 DataCon 数据集中 2020 年 5 月 31 日的客户端查询记录，其中每一个蓝色点代表了一个客户端对一个域名的查询。在图 3.3(a)中，可以清晰地发现存在横向的线条，也就是说有一些客户端访问了大量的恶意域。然而在图 3.3(b)中点的分布更加随机。

第四章 HANdom 设计思想

图 4.1 展示了 HANdom 检测系统的结构，其中一共包括四个主要部分：特征提取、HIN 构建、元路径提取、图裁剪和图卷积分类器。特征提取从两个层面提取 DNS 域的相关特征；HIN 构建中利用包含不同关系和组件的异质信息网络对 DNS 场景建模；图裁剪对数据中噪音进行处理；元路径提取在该异质图上提取三种元路径，以展示域之间的关系；图卷积分类器首先基于元路径提取子图，在子图上使用 GATConv[21]卷积获取每个节点的表示特征，接着用语义级别注意力机制将特征聚合成嵌入向量，最后将用全连接网络对嵌入向量进行判断。

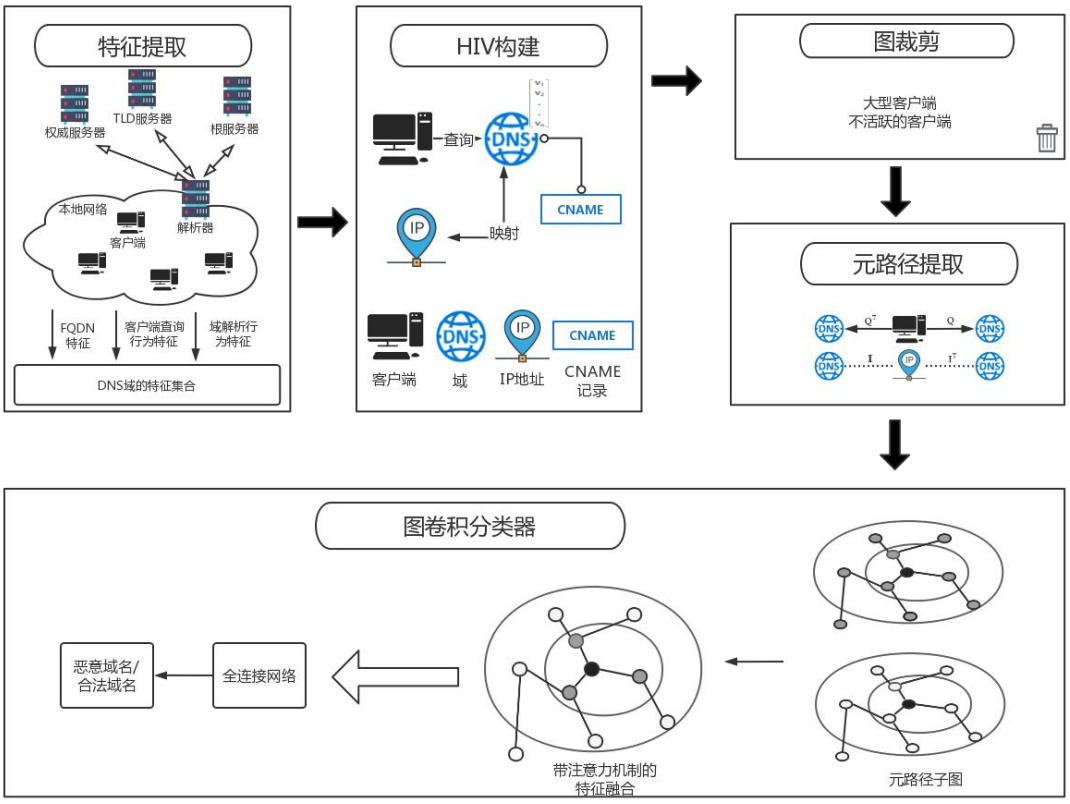


图 4.1 HANdom 系统架构

4.1 特征提取

在该阶段本文通过两个不同的方向提取 DNS 域的特征，来对 DNS 域进行描

述，分别是 FQDN 特征和 DNS 解析特征。

FQDN 特征：如第 3.1 节所示，本文在 FQDN 上提取 7 维特征，分别是 FQDN 中字符总数，大写字符数，数字字符数，特殊字符数以及标签方面的总标签数，最大标签长度以及平均标签长度。此外考虑到使用的数据集中对 FQDN 做了去隐私的处理，导致无法提取 FQDN 字符串中的熵作为特征。故而使用 FQDN 中单词相关的特征进行替代，分别为 FQDN 中单词数，平均单词长度和最大单词长度。

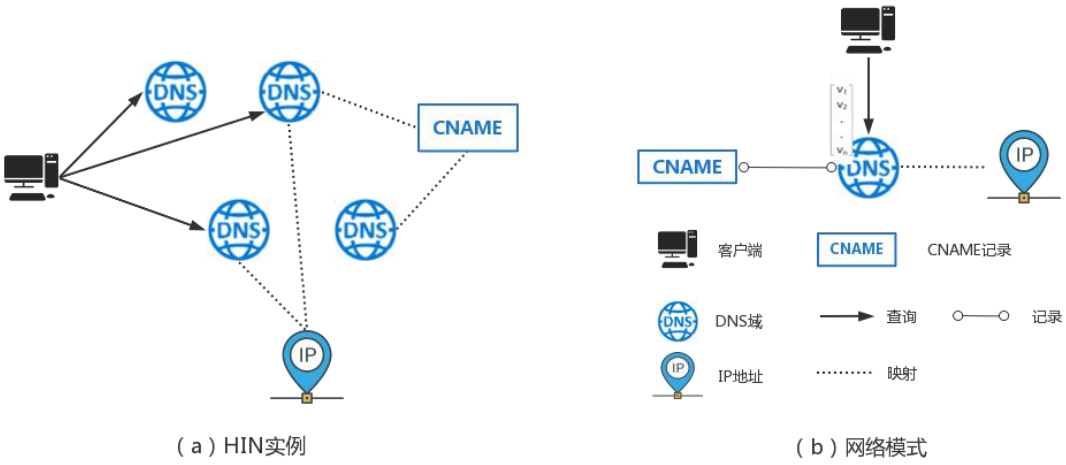
DNS 解析特征：在 DNS 解析方面，本文考虑的是被访问资源记录的类型和 DNS 解析到的 IP 地址数量。再资源记录的类型上，提取了包括每个域名被访问的资源记录类型数，A 和 AAAA 记录占该域名所有被访问记录的比例。在 DNS 解析到的 IP 地址数量方面，则是分别统计了 DNS 映射到的地址数量。

总而言之，本文一共从两个不同方向，提取了总共 13 维特征来对 DNS 域进行描述，具体如下表所示。

表 4.1 DNS 初始特征

特征名称	特征类型
FQDN 字符总数	int
大写字符比率	float
数字字符比率	float
特殊字符比率	float
总标签数	int
最大标签长度	int
平均标签长度	int
单词数	int
平均单词长度占比	float
最大单词长度占比	float
被访问的资源记录类型数	int
A 和 AAAA 记录访问占比	float
IP 地址数量	int

4.2 HIN 构建



（a）HIN实例　　（b）网络模式

图 4.2 HIN 实例（a）即其网络模式（b）

如图 4.2 所示，在本文提出的系统 HANdom 中，利用具有四种不同类型节点的 HIN 来对 DNS 场景建模。这三种节点分别是客户端(C)，DNS 域(D)，IP 地址(I)和 CNAME 记录(R)。本文通过建立以下邻接矩阵来表示这些节点之间的联系。

R1：为了表示域和 IP 地址之间关系，本文建立了矩阵 M 来表示这之间的映射关系，即在 M 中，当域 i 映射到 IP 地址 j 时，元素 $M_{(i,j)}=1$，否则元素 $M_{(i,j)}=0$。

R2：为了表示客户端和域之间的联系，本文建立矩阵 Q 来表示客户端查询域名，其中若客户端 i 查询到域 j 时，元素 $Q_{(i,j)}=1$，否则元素 $Q_{(i,j)}=0$。

R3：为了表示域和其 CName 记录之间的联系，本文建立矩阵 O 来表示域和 CNAME 记录之间关联，其中若域 i 的 CNAME 记录为 j，元素 $O_{(i,j)}=1$，否则 $O_{(i,j)}=0$。这么做的原因在于，如果两个域有相同 CNAME 记录，则这两个域在同一主机上，极有可能属于同一攻击者。

同时，将特征提取阶段得到的特征作为每个域节点上的特征向量 V。

4.3 图裁剪

在现实网络中，收集到的数据常常带有大量噪音，这会导致建立的异构图中

存在一些无用的节点，这些节点不仅会消耗大量计算资源，甚至会对分类效果产生消极影响。为了降低这些噪音的恶性后果，提高计算效率，HANdom 系统删除异构图中如下节点。

a) 大型客户端。在网络中存在一些大型的客户端，它们访问的域占整个网络环境中域集的相当比例，这些客户端大部分是转发或者代理，也就是说，它们极有可能同时访问恶意域名和正常域名。这些客户端的存在会扰乱图卷积中的特征聚合过程，从而降低检测效果。由于这些客户端在查询行为中产生的噪音，本文删除访问的域数量排在前 k_a%的客户端(这里 k_a 设置为 0.1)

b) 不活跃的客户端。本文把访问的域数量少于 k_c 的客户端视为不活跃的客户端。删除这些客户端的理由是它们并没有给挖掘 DNS 域间联系提供太大帮助，删除后不会对检测效果照成太大损害，但是能降低运算复杂度。（k_c 设置为 3）

4.4 元路径提取

如图 4.3 所展示，为表示图中 DNS 域节点之间的关系，本小节建立了以下几条元路径。由于该系统的目的是识别恶意域，故而每条元路径 P 都是对称元路径。其中 P1 表示，如果不同域被同一批客户端请求，则这些域的性质可能相同。这是基于正常客户端几乎不会查询恶意域这一假设。而 P2 展示了攻击者资源的限制。如果不同的域在一段时间内解析到同一 IP 地址，则这几个域极有可能是被同一攻击者所掌握。P3 则是表示了有相同 CNAME 记录的域名，这类域名常属于同一主机。

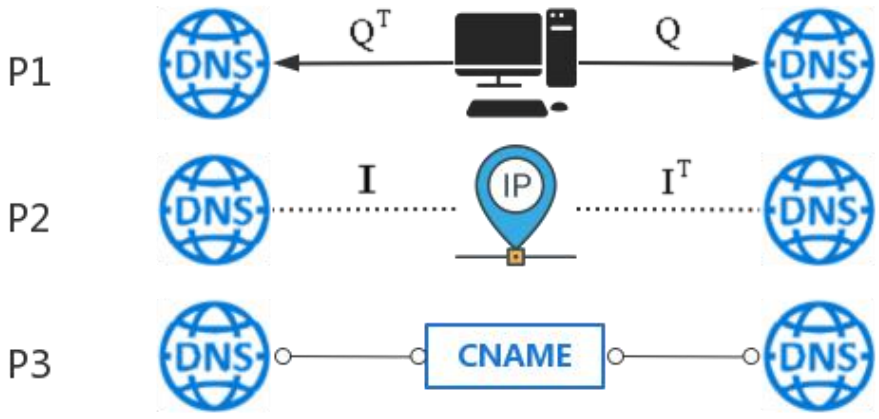


图 4.3 元路径

4.5 图卷积分类器

HANdom 系统中的分类器将异构图和 DNS 域节点的特征作为输入，将未标记的域的检测结果作为输出。该分类器的目标在于，首先利用异构图中各节点间的联系和 DNS 域节点本身的特征，为 DNS 域生成用于识别恶意域的嵌入向量，然后在利用机器学习模型对嵌入向量进行判断，得到 DNS 域的标签。因此，本文中可以将该分类器如式 4.1 所示。

$$\mathrm{F} = \gamma_{w_1}\left(p_{w_2}(G,h)\right) \tag{4.1}$$

其中 γ_{w_1} 表示类型判断函数，p_{w_2} 表示嵌入向量生成函数，G 表示异构图，h 表示节点上的特征。首先 p_{w_2} 将整张图和节点上的特征集作为输入，为每个 DNS 域节点 v_i 生成嵌入向量 z_i，然后 γ_{w_1} 根据每个节点的 z_i 为其进行分类。接下来，本文将介绍这两个函数的实现细节。

4.5.1 嵌入向量生成函数

参考 HAN 模型[25]，本小节使用基于元路径的图卷积方法。该嵌入向量生成函数可以表示为

$$Z = p_{w_2}(G,h) = \varphi(\psi(G,h)) \tag{4.2}$$

其中 ψ 表示元路径上提取特征向量的函数，φ 表示对各元路径上提取到的向量进行聚合，Z 为嵌入向量集。下面详细介绍这两个函数。

首先是对 ψ 函数进行介绍。该函数基于元路径 p_i 从异质信息图 G 中提取子图 $\mathrm{g}_{\mathrm{p_i}}$，其中 $p_i \in P$。然后在每个子图 $\mathrm{g}_{\mathrm{p_i}}$ 上利用 GATConv 生成嵌入向量集 Z_{p_i}，具体如下所示。

$$\psi(G,h) = \{GATConv(g_{p_i},h) \mid p_i \in P\} = \{Z_{p_i} \mid p_i \in P\} \tag{4.3}$$

本文中对 GATConv 采用 K 头注意力机制，将每一头得到的特征向量取均值得到在每个 DNS 域节点 v_i 在子图 g_{p_i} 上的表示向量 $z_i^{p_i}$。

然后使用带语义级别注意力机制的特征融合函数 φ 来对节点 v_i 在各子图上的表示向量 $z_i^{p_i}$ 进行聚合。元路径在异构图中表示不同的意义的语义关系，不同元路径的语义不同，对任务的贡献度也不同。用于学习元路径重要性的注意力机制可

以形式化描述如下：

$$\left(\beta_{p_i} \mid p_i \in P\right)= SemATT\left(Z_{p_i} \mid p_i \in P\right) \tag{4.4}$$

其中 $\beta_{\mathrm{p_i}}$ 是各个元路径上的注意力权重，在其中，本文利用单层神经网络和语义级别注意力向量来学习各个元路径的重要性，并利用 softmax 函数来进行归一化，具体如下所示：

$$w_{p_i} = \frac{1}{|V|}\sum_{i \in V} q^T \cdot \tanh\left(\mathrm{W} \cdot \mathrm{z}_i^{p_i} + b\right) \tag{4.5}$$

$$\beta_{p_i} = \mathrm{softmax}(w_{p_i}) = \frac{\exp(w_{p_i})}{\sum_{i=1}^{|P|}\exp(w_{p_i})} \tag{4.6}$$

其中 w_{p_i} 代表元路径 p_i 的重要性。在对每个元路径上的表示向量 $\mathrm{z}_i^{p_i}$ 进行非线性化后，使用语义级别注意力向量 q 将之转化为相似度，接着将每个节点 $i \in V$ 上的相似度取均值得到元路径 p_i 的重要性 w_{p_i}，再将 w_{p_i} 进行归一化后得到注意力权重 $\beta_{\mathrm{p_i}}$。

最后，通过对多个元路径上的表示向量 $z_i^{p_i}$ 进行加权融合，得到每个节点上的嵌入向量 $z_i = \sum^{|P|} \beta_{p_i} \cdot z_i^{p_i}$。

4.5.2 分类函数

在得到嵌入向量后，对域进行分类的任务就变成了简单的二分类任务。本文采用由三层隐藏层构成的全连接网络作为分类器，其具体结构如图 4.4 所示。

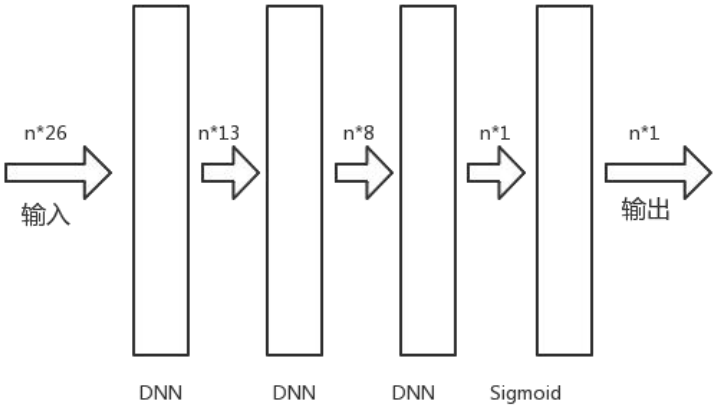


图 4.4 分类函数结构图

在得到分类器预测结果后，与阈值 t 进行比较，得到最后的标签。

本方法中使用 Adam 优化器对图卷积网络中的所有参数进行训练，并采用交叉熵损失（cross-entropy loss）作为损失函数来估测预测值和真实标签之间的距离。此外，考虑到得到的数据中，存在大量无标签的 DNS 域，所以在训练过程中使用半监督训练方法，仅考虑有标签的 DNS 域的嵌入向量。

第五章 实验

5.1 数据集

在实验中，使用的数据集是恶意域名检测方向的开放数据集——DataCon2020-DNS，该数据集由奇安信技术研究院提供[26]，面向科研人员、教学人员进行持续开放。该开放数据集中一共有三个子数据集，本文使用其中的第三个：猜猜我是谁之域名大分类（域名聚类）。

该数据集中包括了来自约 10000 个黑白域名的 DNS 请求数据集，囊括了 DNS 僵尸网络、DNS 木马、APT 攻击、gamble，以及正常域名。数据集中具体包括了 fqdn、ip/ipv6、access、flint 以及 label 数据。

FQDN 内包括了所有待分类的域名以及 flint 中出现的域名。这里给出的 encoded_fqdn 是经过清洗后的数据，仅保留了字符串特征。其格式为：编码后的域名.tld。编码方式为：a 表示字母、0 表示数据、[aaa]表示为一个词语、特殊符号无变化。例如， abchello-12.com 将会被编码为 aaa[aaaaa]-00.com。此外还有每个 fqdn 的域名编号 fqdn_no。具体见表 5.1。

IP/IPv6 中包括了所有客户端的 IP 和 Flint 解析结果中出现的 IP 信息，如表 5.2 所示。其中不仅包括了加密后的 IP，还有 IP 的地理信息和网络供应商信息。

Access 中包括了客户端访问域名记录，该部分是按小时聚合，分别包括了域名标号，加密后的 IP，请求量，该时间段全网请求量，日期，时刻，具体见表 5.3。

Flint 中包括了域名解析记录，包括被解析到的域名（cname，ns）的 A 和 AAAA 记录，这部分具体包括了域名编号，解析类型，解析结果，访问量和 TTL，如表 5.4 所示。flintType 的 1 表示 A 记录，28 代表 AAAA 记录。

最后，在 label 中包括了编码后的域名，域名编号和其对应的标签，如表 5.5 所示，标签包括了“botnet”，“trojan”，“apt”，“gamble”和“white”五类标签。

表 5.1 FQDN 中数据

encoded_fqdn	fqdn_no
[aaaaaa][aaaaaaaaa].com	fqdn_0
aaa[aaaa]aaa.ir	fqdn_1
[aaaaa]aa[aaaaa].com	fqdn_2
a[aaaaaaa]0.[aaaaaaaaa]-[aaaaaaa].com	fqdn_3
aa.aaa[aaaa]a[aaaaaaa].com	fqdn_4

表 5.2 IP/IPv6 中数据

encoded_ip	country	subdivision	city	latitude	longitude	isp
79e14e0a6aa30c8d7d98650bc2413baf.16	中国	浙江	丽水	28.45191	119.908722	电信
73c667fc9226f4299c05aa92c713c097.168	中国	广东	中山	22.516701	113.366699	电信

表 5.3 客户端访问域名记录

fqdn_no	encoded_ip	request_cnt	total_request	date	hour
fqdn_13606	4d02b59739e27c5da16c448939b700f2.86	3	19868706108	20200531	23
fqdn_5700	5e272d088400512d6d01f36e6c1e6d71.181	1	19868706108	20200531	23
fqdn_10493	82bc26fa5d8fa221f8afdd8e20a3acb1.10	2	19868706108	20200531	23
fqdn_13960	7ee88083cc1c685f0e07a7884f0f4c91.227	12	19868706108	20200531	23

表 5.4 域名解析记录

fqdn_no	flintType	encoded_value	requestCnt	ttl	date
fqdn_1204	1	034555d1f94800816818a412359c45ee.93	89	600	20200531
fqdn_2664	5	fqdn_8144	29	3600	20200531
fqdn_9156	1	0acb50bd3eb661e55dfb3db7357fd1f3.3	1	3600	20200531
fqdn_150	28	33939c566589867a698d2dd17f17c1d8:0:bd73:4880:93a1	5	60	20200531
fqdn_9293	28	f19c1b72ce1f76d32e1616a85660dd2b:8:d3fb:39c0:93a1	1	60	20200531

表 5.5 FQDN 及其对应标签

encoded_fqdn	fqdn_no	label
[aaaaaaa][aaaa].com	fqdn_6709	botnet
aaa.000aaa.com	fqdn_1465	trojan
Aaa[aaaaaaa].aa.kr	fqdn_1025	apt
a000.com	fqdn_2365	gamble
aa[aaaa]aaaa.com	fqdn_15995	white

5.2 实验设置与指标

本文中，将 DataCon2020-DNS 恶意域名数据集中标签为“white”的域名看作正常域名，其余看作恶意域名，并以 8:2 的方式划分训练集和测试集，利用训练集将模型训练 1000 轮后得到最终模型。同时，4.3 小节中提到的 k_a 和 k_c 分别设为 0.1 和 3，分类函数中阈值 t 设置为 0.5。

本文中的异构图卷积是在 Python3.6.5 上基于 DGL 和 pytorch 进行实现，DGL 是亚马逊提供的一款用于简化图运算和图神经网络的软件包。本次实验基于 Windows 10 家庭中文版系统，64 位操作系统，处理器为 Intel(R) Core(TM) i7-10870H CPU，内存 16.0GB，使用显卡为 GeForce RTX 2060。

在本次实验中，所有结果都采用 5 折交叉检验取平均值的方法得到，采取以下指标来评价提出的检测模型：

1）真阳率 TP：恶意域名被标记为恶意域名。

2）真阴率 TN：正常域名被标记为正常域名

3）假阳率 FP：正常域名被标记为恶意域名

4）假阴率 FN：恶意域名被标记为正常域名

5）精度 Accuracy：(TP+TN)/(TP+TN+FP+FN)

6）准确度 Precision：TP/(TP + FP)

7）召回率 Recall：TP/(TP + FN)

8）F1：2×(precision × recall)/(precision + recall)

5.3 性能评价实验

为了检验 HANdom 方案的性能，本文决定将其与 SVM、决策树、随机森林，以及全连接神经网络（DNN）进行比较。在其他四种检测方式中，使用的特征为在 4.1 节中提取初始特征。其检验结果如表 5.6 所示。从表 5.6 中可以看出，本文提出的 HANdom 性能远高于直接使用原始特征进行检测的模型，这是因为本文提出的检测模型不仅包括 DNS 域名自身特征，还利用了各 DNS 域名之间的联系来辅助检测。

表 5.6 与基于原始特征的检测模型比较

检测模型	Accuracy	Precision	Recall	F1
SVM	72.9%	61.1%	39.6%	49.0%
决策树	73.1%	63.8%	39.4%	48.7%
随机森林	74.6%	70.1%	41.6%	52.2%
DNN	74.7%	89.5%	30.9%	45.9%
HANdom	**95.4%**	**94.2%**	**91.5%**	**92.8%**

同时，为了检验本文中提出的卷积模型与其他图卷积方案的性能差异，本文还将该方案与三种具有代表性的图卷积方案进行比较。这三种方案分别是 DeepWalk[27]、Metapath2vec[28]和 GraphSAGE[29]。比较结果如表 5.7 所示。

表 5.7 与其它图卷积模型比较

检测模型	Accuracy	Precision	Recall	F1
DeepWalk	76.2%	70.3%	69.8%	70.1%
Metapath2vec	91.0%	89.1%	84.0%	86.4%
GraphSAGE	88.6%	83.9%	80.4%	82.1%
HANdom	**95.4%**	**94.2%**	**91.5%**	**92.8%**

这里，GraphSage 的检测结果优于 DeepWalk，原因是它不仅考虑了节点之间的联系，还考虑了节点自身的特征。Metapath2vec 比前两者表现更为优秀，其原因是该模型更加注重于节点之间的联系，而域名检测任务中有大量有用信息隐藏在节点的联系上。同样，这个猜测也可以解释为何选择的初始特征检测效果不尽如人意，但本文最终的模型有出色的检测效果。此外，本文提出的算法能比另外三种图卷积模型表现更加优异，原因在于该算法不仅可以充分利用域名间的联系信息，还考虑到了不同联系（元路径）的重要程度。

5.4 元路径和特征融合效果实验

本小节进一步研究各元路径以及本文中采用的特征融合机制对恶意域名检测的贡献。表 5.8 包括了在语义级别注意力机制中，各元路径的注意力权重。在这里，我们可以发现权重最大的是元路径 P1，其次是 P2，最后是 P3。这个发现与本文最初的猜想有一定差异，最初本文猜想 P3 所占权重应当最大。本文认为造成差异的原因在于连接的数量问题。虽然 P_3 中拥有相同 CNAME 的域名本质上属于同一主机，但是由于在整个网络环境中拥有 CNAME 的域名数量较少，比如本文使用的数据集中就只包含 888 条 CNAME 记录，但是相比之下客户端访问记录有约 3639848 条，而域名到 IP 地址的映射记录也有 102413 条。

表 5.8 元路径注意力权重

路径编号	元路径名称	注意力权重
P_1	D-C-D	0.6160
P_2	D-I-D	0.2694
P_3	D-R-D	0.1146

本小节对仅考虑一种元路径的图卷积模型（分别以表 5.8 中元路径名称表示）进行测试，同时，为了与特征融合后的方案进行比较，还对不使用注意力机制，仅对各原路径中取得的表示向量取均值的模型进行了比较。这些模型都运行 1000 轮，结果如表 5.9 和图 5.1 所示。

表 5.9 特征融合机制作用评估

检测模型	Accuracy	Precision	Recall	F1
D-C-D	90.8%	89.0%	82.0%	85.4%
D-I-D	72.3%	66.9%	30.7%	42.0%
D-R-D	69.8%	78.1%	11.1%	19.5%
均值方法	93.7%	91.8%	88.7%	90.2%
HANdom	**95.4%**	**94.2%**	**91.5%**	**92.8%**

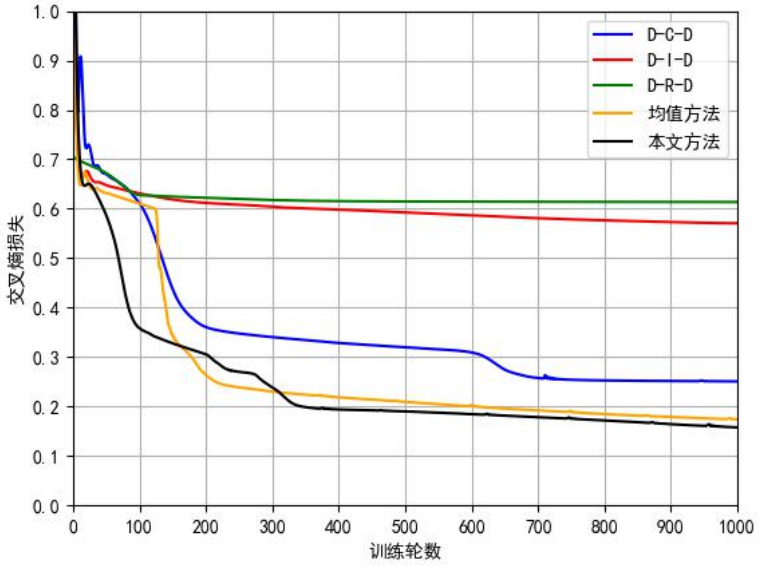


图 5.1 模型损失变化图

在表 5.9 中,可以发现三种单独元路径的检测效果都不如本文提出的 HANdom 方法，但是其中元路径 D-C-D 的检测结果较为理想，以元路径 D-R-D 表现最不如人意，这个结果也与表 5.8 中的注意力权重相吻合。我们可以从该现象中得出以下结论：客户端的访问行为给恶意域名检测提供了最多的信息，而其它路径上的信息是其有益补充。值得一提的是，在使用均值方法的模型表现也较为可观，各项指标都接近本文提出的方法。原因可能是：在子图上使用的 GATConv 卷积中包括了自注意力机制，在一定程度上调整了各元路径对检测结果的贡献度。

第六章 总结与展望

6.1 总结

本文建立了一种基于异质信息网络的恶意域名检测系统 HANdom。HANdom 将 DNS 场景建模为一个包括客户端、IP 地址、CNAME 记录以及域名的异构图，并从该图上自动提取复杂语义，这使得攻击者难以逃避检测系统。考虑到 DNS 场景的不断变化，HANdom 采用类似于 HAN 的基于空间的异构图卷积方法来处于域名节点分类问题。通过基于元路径子图提取方案和基于语义级别注意力机制的特征融合方案，HANdom 可以将域名特征和域名间联系相结合。本文基于 DataCon2020-DNS 恶意域名开放数据集，对 HANdom 系统进行了实验检验，发现 HANdom 系统的检测精度达到了 95.4%，HANdom 的检测效果远优于基于特征的机器学习方法和其它三类图卷积方法。此外，本文还对 HANdom 中的基于语义级别注意力机制的特征融合方法进行评估，发现该方法有效提升了检测效果。

6.2 不足与展望

在本文中存在以下问题有待解决。

1）冷启动问题。对于新加入网络的恶意域名，由于其还没有与其它恶意域名发生联系，本文提出的方案对其检测效果将退化到利用原始特征的检测效果。要解决该问题，应当想办法提升原始特征的质量。

2）没有充分利用资源记录。在本文中，仅利用到了 A、AAAA 和 CNAME 资源记录，但是在 DNS 资源记录中有多达 16 种不同类型的资源记录。在后续的工作中，可以考虑充分利用其它类型资源记录数据。

3）加密 DNS。目前，几乎所有的恶意域名检测系统都是基于一个前提：DNS 流量是透明的，可被直接观察和分析。但是，出于保护用户隐私的目的，一些加密 DNS 流量的协议已经被提出，并被一些知名企业采纳，如 Google、Firefox 等[30]，这将给恶意域名检测带来极大挑战。

致谢

岁月如流水，匆匆而过，转眼已是四年光阴，而我的本科学业也即将告一段落。点击运行，模型基本达到预期效果，虽然觉得有众多不完善的地方，比如提取的初始特征不够完善，模型还有继续训练的空间等等。但转眼又会安慰自己，最后结果看得过去，就可以了。呵，这就是所谓的自欺欺人吧。

毕业设计，应该是我大学所提交的最后一份作业了。在此感谢大学期间所有给我帮助的老师，同学，与你们的记忆是我人生的财富，也是我生命中不可或缺的一部分；同时也要感谢我的父母亲人，感谢他们的养育之恩和对我的支持；最后，我要特别感谢我的指导老师张锋，感谢他不厌其烦的指导和帮助。

大学期间，欢乐也有，悔恨也有，曾经年少轻狂，也曾故作老成。最终，一切也就归结为一句话：我来到，我看过，我走了，我记得。接下来，我将走上硕士生的学业，在这里祝福和我一路上风雨同舟的朋友们，一路走好，未来可期。

参考文献

[1] Paul V. Mockapetris. Domain names - implementation and specification[J]. Proceedings of the ACM on Programming Languages, 1983, 883: 1-74.

[2] Leyla Bilge, Engin Kirda, Christopher Kruegel, Marco Balduzzi. EXPOSURE: Finding Malicious Domains Using Passive DNS Analysis[C]. 18th Annual Network and Distributed System Security Symposium. :Internet Society, 2011:195-211.

[3] Manmeet Singh, Maninder Singh, Sanmeet Kaur. Detecting bot-infected machines using DNS fingerprinting[J]. Digital Investigation, 2019, 28: 14-33.

[4] Guanghua Yan, Qiang Li, Dong Guo, Bing Li. AULD: Large Scale Suspicious DNS Activities Detection via Unsupervised Learning in Advanced Persistent Threats[J]. Sensor, 2019, 19(14): 3180-3198.

[5] Lu Huang, Jingfeng Xue, Weijie Han, Zixiao Kong, Zequn Niu. Detection of malcious domains in apt via mining massive dns logs[C]. International Conference on Machine Learning for Cyber Security. :Springer, Cham, 2020: 140-152.

[6] Shaojie Chen, Bo Lang, Hongyu Liu, Duokun Li. DNS covert channel detection method using the LSTM model[J]. Computers & Security, 2021, 104:102095-102110.

[7] 刘浩杰，皇甫道一，李岩，王涛. 一种通用的恶意域名检测集成学习方法[J]. 网络空间安全, 2019, 10(09): 26-32.

[8] Zhouyu Bao, Wenbo Wang, Yuqing Lan. Using Passive DNS to Detect Malicious Domain Name[C]. Proceedings of the 3rd International Conference on Vision, Image and Signal Processing. 2019:1-8.

[9] Lior Sidi, Asaf Nadler, Asaf Shabtai. MaskDGA: A Black-box Evasion Technique Against DGA Classifiers and Adversarial Defenses[J]. CoRR, 2019: abs/1902.08909.

[10] Mohamed Nabeel,Issa M. Khalil, Bei Guan, Ting Yu. Following Passive DNS Traces to Detect Stealthy Malicious Domains Via Graph Inference[J]. ACM Transactions on Privacy and Security, 2020, 23(4): 1-36.

[11] Babak Rahbarinia, Roberto Perdisci, Manos Antonakakis. Segugio: Efficient behavior-based tracking of malware-control domains in large isp networks[C]. 2015 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks.

2015: 403-414.

[12] Futai Zou, Siyu Zhang, Weixiong Rao, Ping Yi. Detecting Malware Based on DNS Graph Mining[J]. International Journal of Distributed Sensor Networks, 2015, 11(10): 102687.

[13] Issa Khalil, Ting Yu, Bei Guan. Discovering Malicious Domains through Passive DNS Data Graph Analysis[C]. Asia Conference on Computer and Communications Security. 2016: 663-674.

[14] Xiaoqing Sun, Mingkai Tong, Jiahai Yang. HinDom: A Robust Malicious Domain Detection System based on Heterogeneous Information Network with Transductive Classification[C]. 22nd International Symposium on Research in Attacks, Intrusions and Defenses. 2019: 399-412.

[15] Sun Xiaoqing, Wang Zhiliang, Yang Jiahai, Liu Xinran. Deepdom: Malicious domain detection with scalable and heterogeneous graph convolutional networks[J]. Computers & Security, 2020, 99(prepublish): 102057-102073.

[16] Michael Dooley, Timothy Rooney. DNS Security Management[M]. John Wiley & Sons, 2017: 17-29.

[17] Yizhou Sun, Yintao Yu, Jiawei Han. Ranking-based clustering of heterogeneous information networks with star network schema[C]. Acm Sigkdd International Conference on Knowledge Discovery & Data Mining. 2009: 797-806.

[18] Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S. Yu, Tianyi Wu. PathSim: Meta Path-Based Top-K Similarity Search in Heterogeneous Information Networks[J]. Proceedings of the Vldb Endowment, 2011, 4(11): 992-1003.

[19] Thomas N. Kipf, Max Welling. Semi-Supervised Classification with Graph Convolutional Networks[C]. International Conference on Learning Representations. 2017: 1-14.

[20] Ruoyu Li, Sheng Wang, Feiyun Zhu, Junzhou Huang. Adaptive Graph Convolutional Neural Networks[C]. Proceedings of the AAAI Conference on Artificial Intelligence. 2018: 3546–3553.

[21] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, Y. Bengio. Graph Attention Networks[J]. CoRR, 2017: abs/1710.10903.

[22] Jawad Ahmed, Hassan Habibi Gharakheili, Qasim Raza, Craig Russell, Vijay Sivaraman. Monitoring Enterprise DNS Queries for Detecting Data Exfiltration From Internal Hosts[J]. IEEE Transactions on Network and Service Management, 2020, 17(1):

265-279.

[23] Anirban Das, Min-Yi Shen, Madhu Shashanka, Jisheng Wang. Detection of Exfiltration and Tunneling over DNS[C]. 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA). 2017: 737-742.

[24] Rahbarinia B, RobertoPerdisci, Antonakakis M. Segugio: Efficient Behavior-Based Tracking of Malware-Control Domains in Large ISP Networks[C]. 2015 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks. NW Washington, DC. United States. 2015: 403-414.

[25] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Peng Cui, Philip S. Yu, Yanfang Ye. Heterogeneous Graph Attention Network[C]. WWW '19: The World Wide Web Conference. 2019: 2022-2032.

[26] DataCon社区. DataCon开放数据集-DataCon2020-DNS恶意域名数据集方向开放数据集. 2021-11-11. https://datacon.qianxin.com/opendata/openpage?resourcesId=1.

[27] Bryan Perozzi, Rami Al-Rfou', Steven Skiena. DeepWalk: Online Learning of Social Representations[C]. KDD '14: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. 2014: 701-710.

[28] Yuxiao Dong, Nitesh V. Chawla, Ananthram Swami. Metapath2vec: Scalable Representation Learning for Heterogeneous Networks[C]. KDD '17: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017: 135-144.

[29] William L. Hamilton, Rex Ying, Jure Leskovec. Inductive Representation Learning on Large Graphs[C]. NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017: 1025-1035.

[30] 孟德超, 邹福泰. DNS 隐私保护安全性分析[J]. 通信技术, 2020, 53(02): 445-449.