加密流量下网络公害行为主体关联技术

崔天宇 [1,2,3], 侯承尚 [2,3], 刘 畅 [2,3], 石俊峥 [2,3], 苟高鹏 [2,3], 熊 刚 [2,3]

[1] 中关村实验室 北京 中国 100094
[2] 中国科学信息工程研究所 北京 中国 100093
[3] 中国科学院大学网络空间安全学院 北京 中国 100049

摘要 网络公害治理一直是信息内容安全从业者面临的重要课题。海量的网络公害信息的传播不仅污染了互联网环境, 也阻碍了社会的健康发展。因此, 打击网络公害行为成为国家网络安全事业的重点防线。然而, 流量的加密化和客户端地址更迭问题给网络公害行为的分析带来了巨大的挑战, 公害用户主体的定位和公害行为的追踪溯源在流量场景下都难以实现。针对上述问题, 在本文中, 我们提出了一种加密流量下网络公害行为主体关联技术。方法提取每个客户端地址一段时间的流量特征作为地址的网络行为知识图, 并基于图神经网络和孪生网络构建地址关联模型 PolluTracker, 实现网络公害用户的地址关联和长期溯源分析工作。我们在 5 个月的真实用户流量数据集上进行了广泛的实验, 结果表明, 方法能够以 99%的准确率实现公害主体的地址关联工作, 相比现有的四种关联方法最多提升了 0.90 倍。消融实验、对抗实验、实际案例分析等多项测试表明, 我们的方法能够有效实现目标公害用户的长期行为关联分析工作, 并且关联效果兼具鲁棒性和逃逸对抗能力。

关键词 网络公害; 网络行为分析; 加密流量; 图表示学习; 度量学习
中图法分类号 TP309.2 **DOI 号** 10.19363/J.cnki.cn10-1380/tn.2025.09.04

Correlation Technology between Cyber Pollution Behaviors and Subjects under Encrypted Traffic

CUI Tianyu [1,2,3], HOU Chengshang [2,3], LIU Chang [2,3], SHI Junzheng [2,3], GOU Gaopeng [2,3], XIONG Gang [2,3]

[1] Zhongguancun Laboratory, Beijing 100094, China
[2] Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China
[3] School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract The governance of cyber pollution remains a pivotal challenge for those working within the realm of information content security. The pervasive spread of vast quantities of harmful information across the internet not only contaminates the digital landscape but also poses a significant barrier to societal progress. Hence, tackling cyber harmful activities has emerged as a forefront concern within national cybersecurity efforts. However, the encryption of traffic and the constant shifting of client addresses introduce formidable obstacles to analyzing and mitigating cyber pollution. Pinpointing the perpetrators and tracing the origins of these activities become increasingly complex in the context of encrypted traffic. In response to these complications, in this paper, we propose an innovative method for linking actors behind cyber pollution in the context of encrypted internet traffic. Our approach involves harvesting the traffic characteristics associated with each client address over a certain period to create a network behavior knowledge graph for that address. Utilizing this graph, we develop an association model named PolluTracker, which leverages the capabilities of Graph Neural Networks and the framework of Siamese Networks. This model aims to facilitate the association of addresses linked to cyber nuisances and supports the ongoing analysis of their source. Our extensive experimental analysis, conducted over five months using a dataset of real user traffic, indicates that our method can correlate harmful entities with a remarkable 99% accuracy rate. This performance significantly exceeds that of four existing correlation techniques by up to 0.90 times. Moreover, through a series of experiments including ablation studies, adversarial tests, and real-world scenario analysis, our method has demonstrated its effectiveness in conducting long-term behavioral correlation analysis of targeted harmful users. Notably, our approach stands out for its robustness and its adeptness at evading adversarial efforts, marking a significant advancement in the field of cybersecurity and digital environment protection.

Key words cyber pollution; network behavior analysis; encrypted traffic; graph representation learning; metric learning

通讯作者: 侯承尚, 博士, 工程师, Email: houchengshang@iie.ac.cn。
本课题得到国家重点研发计划项目“加密流量中网络公害检测与行为识别、处置研究”(No. 2021YFB3101400)资助。
收稿日期: 2024-02-05; 修改日期: 2024-04-02; 定稿日期: 2025-08-19

1 引言

网络公害是影响国家互联网环境和社会健康发展的严重阻碍。随着近些年互联网技术的蓬勃发展, 越来越多的网络公害服务或内容被散播在网络空间中, 给网络内容监管的从业人员带来了巨大的挑战。色情、赌博等黑灰产业的蔓延、漫无止境的有害信息传播无疑触及了国家网络安全事业的底线, 研究网络公害治理技术是打击网络犯罪、监管有害内容输出、保护互联网产业健康发展环境的首要任务。

然而, 网络流量的加密化给网络公害行为的分析工作带来了难题。加密协议的普及不仅保护了普通互联网用户的隐私, 也同样使这些有害国家和社会安全的内容传播难以甄别。同时, 由于固网的地址复用协议、移动网的动态地址分配问题, 客户端侧地址的频繁变更也使公害主体难以被定位, 公害行为也难以被追踪和溯源。在流量加密与地址更迭场景下, 实现网络公害行为的识别与追踪仍然面临诸多困难。

面对这些问题, 现有研究通常利用网络行为分析的相关技术以期望寻求突破。例如, 网站和移动应用指纹识别技术[1-9]通过学习加密流量中不同网站和应用的指纹特征, 从而帮助区分正常会话与访问赌博、色情等公害网站或应用的通信会话流量。用户行为细粒度分类技术[10-11]可以精细化地区分社交软件中用户的不同网络行为, 从而实现有害内容输出行为的识别和监管。然而, 这些方法只能做到实时流量中可能有害的网络行为识别检测, 对于网络公害行为的产生的追溯分析、公害团伙的网络画像构建与定位往往无法完成。因此, 网络公害行为产生的源头仍然无法被根除, 如何实现恶意团体的公害行为追踪是当前公害行为治理面临的最为困难的挑战之一。

在本文中, 我们提出了一种加密流量下公害行为主体关联技术。方法提取一段时间的 TLS 加密流量的协议特征和统计特征以对流量中的每一个客户端地址构建网络行为画像, 画像中包含的指纹信息和通信行为信息将帮助分析和刻画客户端地址背后的真实用户。随后, 方法利用图神经网络和孪生网络以判断不同画像间的关联关系, 深度学习不同画像中指纹特征和用户活动的相似性, 从而在地址变更的加密流量场景下实现目标主体的网络公害行为长期追踪。5 个月的真实用户流量数据集的评估表明, 方法能够以 99%的准确率实现网络公害行为的用户关联。

本文的主要贡献如下:

(1) 我们提出了网络公害行为主体关联技术, 面对流量加密和客户端地址动态更迭的双重挑战, 区别于传统的网络行为分析方法, 该技术能够跨越多个会话实现用户的网络行为追踪, 从而帮助网络公害行为的目标用户定位与追溯;

(2) 我们构建了客户端地址的网络行为画像, 并设计了基于图表示学习和度量学习的地址关联模型 PolluTracker。方法利用流量中的多维信息建模用户难以被精准描述的网络行为, 并基于图神经网络和孪生网络深度分析不同客户端地址的画像相似性;

(3) 我们收集了真实用户的流量数据以进行广泛的实验评估工作。方法能够以 99%的准确率实现网络公害行为的主体关联工作, 相比现有工作准确率提升了 14%。实验表明, 方法能够以较好的准确性和鲁棒性应用于网络公害治理工作中。

2 相关工作

为了帮助读者更好地理解本文的所有内容, 本节提供了与本文相关的基本知识和研究工作的介绍, 主要包括了加密行为建模和网络行为关联的相关背景。

2.1 加密行为建模

在如今的网络环境下, 加密流量已经成为网络服务或网络应用的主流流量表现形式, 网络服务商与互联网用户的隐私意识提升使安全传输层(Transfer Layer Security, TLS)等安全协议得到了广泛的应用和部署。为了逃避监管审查, 大部分公害网站或服务都使用了 TLS 协议加密通信过程, 恶意用户在社交网络上传播的有害信息也被正常的 TLS 加密渠道所保护。因此, 针对 TLS 加密流量的网络行为分析工作对于网络公害治理工作至关重要。虽然 TLS 协议加密了通信过程中的流量载荷内容, 然而在加密通信建立前的 TLS 握手信息包含了多种可以被建模网络行为的元信息。例如, ClientHello 报文中包含的客户端支持的 TLS 版本、加密套件等信息指示客户端可能使用的浏览器类型, ServerHello 报文与 Certificate 报文中包含的服务端选择的 TLS 版本、加密套件、证书的颁发者、主体等信息间接揭示了被访问的服务类型。目前, 海量的网络研究人员将工作重心转移到加密流量分析领域[1-9]以进行更加有效的用户行为知识提取与用户画像构建工作。然而, 流量下用户的复杂行为会形成多维元信息, 如何进行有效的知识编排实现加密行为的建模将成为网络行为分析工作重大的难点之一。本文考虑利用客户端与服务端的

行为知识图来刻画加密流量下的公害用户的网络行为，从而帮助实现流量特征提取，支撑下游的网络行为关联能力。

2.2 网络行为关联

网络行为关联旨在构建流量中用户的可区分性指纹，从而实现不同时间窗口下用户网络活动的长期关联与追踪。现有工作主要包括了用户画像方法、TLS 指纹方法、流序列方法。用户画像方法的主要思想是利用基于行为的统计特征构建用户画像。Kumpost 等人[12]采用了用户的客户端地址所连接的所有目的 IP 地址来构建用户画像，从而在未来的流量中重新识别这些用户。Herrmann 等人[13]收集了用户的客户端地址发送的所有 DNS 请求域以生成用户画像，并使用贝叶斯分类器在某大学网络内实现了用户追踪。Gonzalez 等人[14]提取了 TLS 流量的扩展字段中的服务器名称指示(Server Name Identifier, SNI)信息实现了用户画像的构建工作。其次，TLS 指纹方法的主要思想是提取 TLS Client Hello 消息域值构建可追溯指纹。在该领域下此前的工业界早已形成众多的开源工具如 JA3[15]、FingerprinTLS[16]、p0f[17]等。Husak 等人[18]利用 TLS 指纹在流量监听中实现了 HTTPS 的客户端识别。Anderson 等人[19]通过 TLS 指纹构建了被动操作系统指纹并实现了操作系统主要和次要版本识别。对于用户的客户端、操作系统的识别和检测也将在个人身份推断与行为关联场景起到巨大的帮助。最后，流序列方法的主要思想是利用数据包时延、长度序列等序列特征实现用户活动标识。例如，Nasr 等人[20]使用流序列来关联 Tor 网络用户的入口和出口流量。Bahramali 等人[21]利用流序列实现了即时通信应用中的频道用户识别。

然而，这些工作只能在封闭数据集上关联和追踪训练集中已知的用户，对于训练集中未出现的用户无法做到行为溯源和追踪。同时，现有工作没有在网络公害检测场景下的先验知识和专用场景，网络公害检测场景下方法迁移的效果仍然未知。对此，本文设计了先进的地址关联模型 PolluTracker 以更好地实现未知用户关联和网络公害检测能力。

3 方法实现

面对流量加密与客户端地址更迭的双重压力，本节详细介绍加密流量下网络公害行为主体关联技术的所有技术细节，包括网络公害行为溯源的问题定义、网络行为画像构建过程以及地址关联的模型实现。

3.1 问题定义

图 1 展示了网络公害行为主体关联的问题定义。设在 t 时间段内，有 N 个网络用户使用 M 个客户端地址访问了在线服务，其中具有公害行为的恶意用户数量为 N_1，正常用户的数量为 N_2。但是由于 TLS 协议加密的原因，仅通过报文内容无法检测到用户和地址之间的关系。例如，m_1 和 m_2 是在时间 t 内流量中观察到的用户的两个地址，但是我们无法判断他们是属于同一用户的两个地址还是来自两个不同的用户。因此，通常情况下，我们无法获知恶意用户的所有公害行为以实现目标用户的网络公害行为的分析追溯，也无法区分流量中存在的海量正常用户行为和网络公害行为。地址关联模型的目标是判断两个任意客户端地址的关联关系，以识别两个客户端地址所承载的网络行为是否属于同一用户。凭借监听时间 t 内所有地址的加密流量作为背景知识 K_t，地址关联模型 F 可以构建关联函数 f 来判断一对地址 $<m_1, m_2>$的关系 R:

$$R = f(<m_1, m_2>|K_t)$$

关联函数 f 可以通过关联模型进行学习，模型提供任意一对地址的距离度量，并通过阈值 η 来确定它们是否属于同一用户。

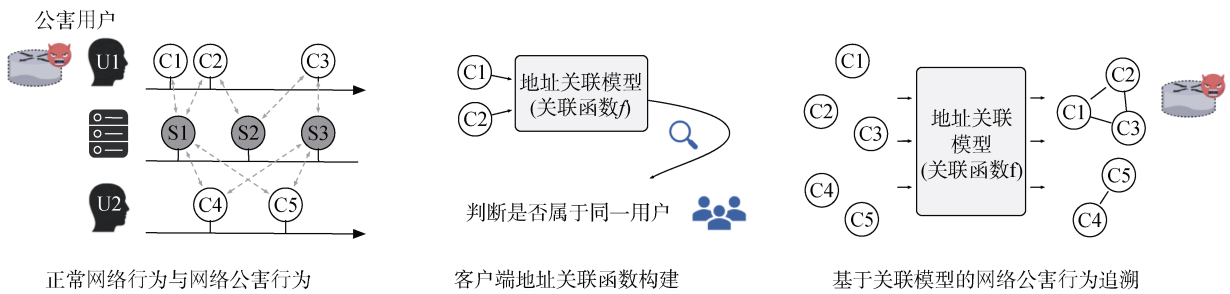


图 1 网络公害行为主体关联场景

Figure 1 Scenarios of the correlation between cyber pollution behaviors and subjects

在获得了地址关联模型后，已知流量中的所有客户端地址形成的地址集为 S，目标恶意用户的所有客户端地址形成的地址集为 $Y=\{y_0, y_1, \cdots, y_n\}$。地址关联模型 F 可以通过其中一个已知地址 y_0 的活动来

追溯完整的地址集 Y:

$$Y = F\left(\left(S, y_0\right) \mid K_t\right)$$

获得地址集 Y 后, 研究人员可以掌握这些地址所关联的所有网络公害活动来分析该目标恶意用户。在本文中, 我们称这些形成网络公害行为的恶意用户为公害主体, 并期望利用加密流量下的网络公害行为主体关联技术实现公害行为的识别与追溯。

3.2 网络行为画像构建

当长期监听路由器或服务器上的网络流量时, 研究人员可以收集有关客户端地址通信的大量元信息, 这些元信息可以被用户画像工作重构以帮助开展用户识别工作。为了在网络公害检测实现这一目标, 如图 2 所示, 方法为每个客户端地址构建了一个基于 TLS 加密通信的行为知识图, 并以此作为观测视角的背景知识。由于用户复杂的网络行为会在监听时间内生成多维的元信息, 我们使用多种类型的节点和邻接关系, 以此来更加准确地描述地址背后的用户活动。

3.2.1 节点和节点属性

基于背景知识, 每个客户端地址的行为知识图包含三种类型的节点, 即客户端节点 C、服务端节点 S 和指纹节点 F, 表 1 展示这些节点的基本信息。每个图节点都保留一个属性来表示节点的含义。

客户端节点 C。客户端节点是指监听时间内建立其行为知识图的该唯一客户端地址, 其属性是报文 IP 头部的客户端地址。每个客户端的行为知识图中只包含一个客户端节点来表示与其相关的元信息。

表 1 行为知识图中节点的详细描述

Table 1 The detail of nodes in the behavior knowledge graph

节点来源	来源	符号	节点属性	节点属性描述
客户端节点	IP 头部	C	客户端地址	构建知识图的唯一客户端地址
服务端节点	IP 头部	S	服务端地址	与客户端地址通信的服务端地址
客户端指纹节点	ClientHello 报文	Fc1	记录层版本(Record Version)	记录层中使用的 TLS 协议版本
		Fc2	客户端版本 (Client Version)	客户端选择的 TLS 通信版本
		Fc3	加密套件 (Cipher Suites)	支持的加密选项列表
		Fc4	压缩方法 (Compression)	支持的压缩方法列表
	ClientHello 报文文	Fs1	服务器名称指示 (SNI)	客户端指定的访问域名
服务端指纹节点	ServerHello 报文	Fs2	记录层版本 (Record Version)	记录层中使用的 TLS 协议版本
		Fs3	服务端版本 (Server Version)	服务端选择的 TLS 通信版本
		Fs4	加密套件 (Cipher Suites)	服务端选择的加密选项
	Certificate 报文	Fs5	算法 ID (Algorithm ID)	签署证书的加密算法
		Fs6	颁发者 (Issuer)	签署并颁发证书的实体
		Fs7	主体 (Subject)	公钥关联的实体

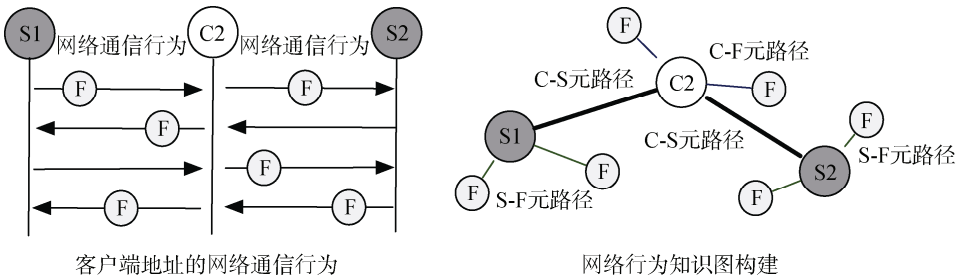


图 2 网络行为知识图构建过程

Figure 2 Network behavior graph contribution

服务端节点 S。服务端节点是与客户端地址建立了 TLS 通信的服务端地址, 其属性为报文 IP 头部的服务端地址。

指纹节点 F。指纹节点包括客户端指纹和服务端指纹, 其属性为 ClientHello、ServerHello、Certificate 消息的协议字段值以及与客户端地址建立 TLS 连接中的统计特征。我们选择常用的、可区分的 TLS 字段进行模型学习。此外, 我们还提取了一些统计特征

来提供更详细的用户行为描述。

3.2.2 邻接关系

在每个客户端地址的行为知识图中, 节点间可以通过不同的语义路径连接, 这些路径被定义为元路径。为了精确刻画复杂用户行为所产生的行为知识, 我们使用了三种元路径来连接行为知识图中的三种类型的节点: 客户端-服务端元路径(C-S)、客户端-指纹元路径(C-F)、服务端-指纹元路径(S-F)。

客户端-服务端元路径(C-S)。客户端-服务端元路径连接客户端节点 C 和多个服务端节点 S, 代表客户端和多个服务端之间的 TLS 通信活动。

客户端-指纹元路径(C-F)。客户端-指纹元路径连接客户端节点 C 和多个客户端指纹节点 F, 代表客户端背后可能使用的浏览器参数。

服务端-指纹元路径(S-F)。服务端-指纹元路径连接每个服务端节点 S 和与服务端相关的多个服务端指纹节点 F, 表示每个服务端背后的服务特征。

使用不同元路径来连接行为知识图中的节点的目的是学习能够帮助网络行为关联的语义信息, 从而能够区分网络公害行为和正常网络行为, 并对于可能产生于同一恶意用户的公害行为进行关联。例如, 客户端-指纹元路径和服务端-指纹元路径可以帮助模型有效地学习唯一的客户端和服务端指纹信息。客户端-服务端元路径则揭示了用户和每个服务端之间的通信活动背后的用户网络行为意图, 使模型掌握客户端地址背后用户的网上习惯, 从而更有效地实现用户身份定位和公害行为发现。

3.3 地址关联模型实现

地址关联模型利用图神经网络[22-25]和孪生网络[26]构架实现对不同地址的网络活动的关联判断, 从而实现恶意用户的网络公害行为的长期跟踪。基于图表示学习和度量学习原理, 方法将获取每个客户端地址的网络行为表示并判断表示之间的关联性。

3.3.1 图表示学习

在为每个客户端地址构建知识图谱后, PolluTracker 可以输入成对的网络行为知识图来建模它们的关联性, 从而推断它们是否来源于同一用户。每个知识图可以提取一个邻接矩阵 A 和特征矩阵 X 以供图神经网络处理, 其中邻接矩阵 A 包括每个节点的邻接关系, 特征矩阵 X 是所有节点的属性值。图神经网络方法通过聚合节点的邻接节点特征来迭代地更新图中节点的特征。地址关联模型 PolluTracker 使用多层次的自注意力机制[27]并根据邻接矩阵 A_i 和 A_j 更新特征矩阵 X_i 和 X_j , 并最终获得它们的图表示来度量两个地址的距离以进行关联推断。PolluTracker 的整体架构如图 3 所示。

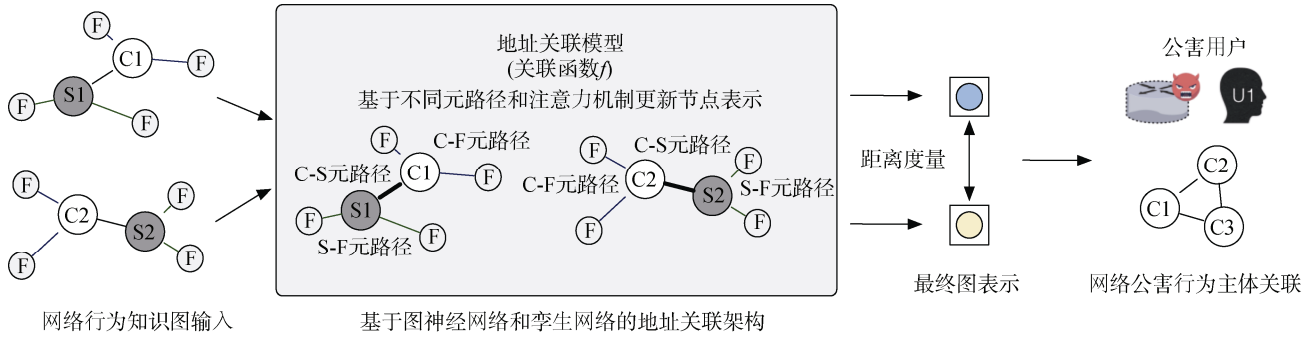


图 3 PolluTracker 的算法流程和模型架构

Figure 3 The algorithm flow and model architecture of PolluTracker

基于图注意力网络更新的原理, 注意力机制首先学习基于元路径的邻接节点的权重并聚合它们以获得节点表示, 从而学习图中的指纹信息和网络行为。给定 N_u^{Φ} 表示节点 u 基于元路径 Φ 的相邻节点, 节点 $v \in N_u^{\Phi}$。基于元路径的节点对<u, v>的重要性可以表述如下:

$$e_{uv}^{\Phi} = \sigma\left(\overset{T}{\Phi} \cdot \left[h_u \| h_v\right]\right)$$

$$a_{uv}^{\Phi} = \mathrm{softmax}\left(e_{uv}^{\Phi}\right) = \frac{\exp\left(e_{uv}^{\Phi}\right)}{\sum_{v \in N_u^{\Phi}} \exp\left(e_{uv}^{\Phi}\right)}$$

其中, h_u 和 h_v 是节点 u 和 v 的特征, a_{Φ} 是元路径 Φ 的节点级注意力参数化矩阵, σ 表示激活函数, $\|$ 表示连接操作。方法基于图注意力网络[24]原理连接不同节点特征并计算其注意力值。随后, 节点 u 的基于元路径的节点表示可以通过聚合所有具有相应权重系数的邻居属性来获得, 如下所示:

$$z_u^{\Phi} = \bigcap_{k=1}^{K} \sigma\left(a_{uv}^{\Phi} \cdot h_v\right)$$

其中, z_u^{Φ} 是元路径 Φ 下的节点 u 学习得到的节点表示, K 是使用多头注意力机制[27]的头数。

由于地址关联模型要求输入成对的用户元信息,

较大的权重系数 a_ϕ 表示在两个行为知识图中基于某一元路径匹配到了相似的邻居属性, 这将有助于关联任务实现同一用户的相似活动发现。例如, 在成对图中的两个客户端节点基于客户端-服务端元路径(C-S)发现其均链接到属性相同的服务端节点, 该现象即说明两个客户端地址访问了相同的服务内容, 注意力机制可以帮助发现这些基于元路径链接中的共同属性。在网络行为知识图的三种元路径中, 客户端-指纹元路径(C-F)和服务端-指纹元路径(S-F)帮助模型基于客户端和服务端指纹信息来学习唯一的客户端和服务端画像, 客户端-服务端元路径(C-S)则使用通信关系挖掘用户的网络行为活动。最终, 通过聚合图中更新后的每个节点表示, PolluTracker 可以获得两张输入的网络行为知识图的图表示 Z 来作为其行为的唯一指纹。

3.3.2 度量学习

地址关联模型 PolluTracker 的孪生网络架构的目标是度量任意两个客户端地址的网络行为知识图之间的距离 D, 并通过阈值 η 来判断两个地址的关联关系 R:

$$D = \|Z_1 - Z_2\|$$

$$R = \begin{cases} 1 & D < \eta \\ 1 & D \geqslant \eta \end{cases}$$

其中, Z_1 和 Z_2 是两个输入的网络行为知识图的最终图表示。R=1 表示两个客户端地址来自同一个用户, 否则 R=0。为了训练关联模型, PolluTracker 需要一组正样本和负样本来学习关联函数。正样本是属于同一用户的客户端地址的成对网络行为知识图。负样本是来自两个不同用户的任意一对地址的网络行为知识图。有了负样本和正样本, PolluTracker 可以通过最小化对比损失 L 来优化关联模型:

$$L = Y \cdot D^2 + (1-Y)\{\max(0, m-D)\}^2$$

其中, Y 是输入样本的标签, m 是一个用于控制可以考虑更新网络的最大距离的超参数。此外, 网络参数在孪生网络架构中共享, 从而帮助 PolluTracker 专注于输入差异并学习相似性。

4 实验设置

本节简要介绍了用于评估我们的加密流量下的网络公害行为主体关联技术的数据集、对比方法、评估指标和模型参数。

4.1 数据集

为了构建大规模的正常行为和网络公害行为组成的实验数据集, 我们在中国科技网 (China Science and Technology Network, CSTNET) 上被动监测了 5 个月的用户流量, 以此收集了 1.7 亿条网络流用于关联实验。在该网络中 80%的用户每月至少变更一次他们的客户端地址。实验利用 HTTP 明文中的持久化 Cookie(Persistent Cookie) 来标记地址的 TLS 流量, 图 4 显示了数据集标注的基本过程。

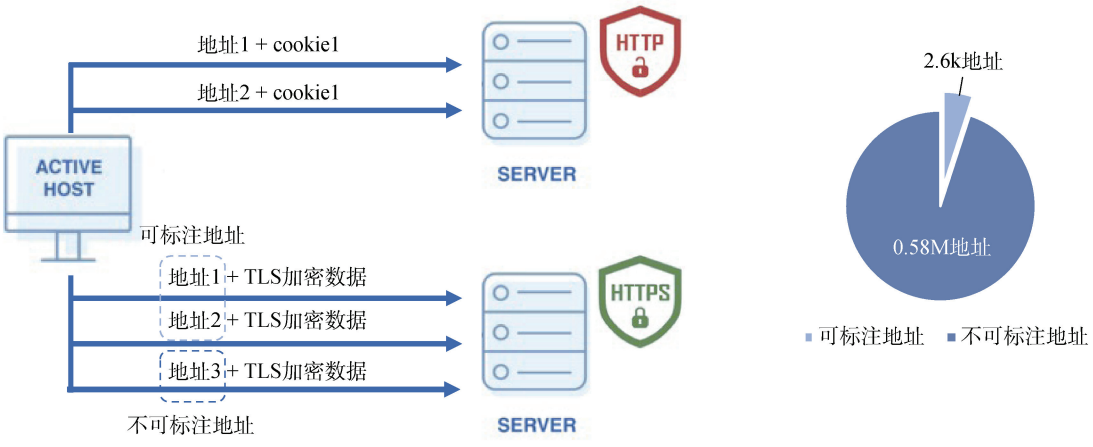


图 4 基于 HTTP 明文的持久化 Cookie 的 TLS 流量标注方法

Figure 4 TLS traffic annotation method based on HTTP plaintext persistent Cookies

背景流量数据收集。当用户在流量监测过程中访问部署了 HTTPS 的网站的同时, 使用相同的地址访问了其他一些没有部署 HTTPS 的网站, 这些用户的 HTTP 明文中的 Cookie 与客户端地址的对应关系将可以被用来构建 TLS 流量下的地址与用户的标签信息。实验通过 HTTP 明文中的 Cookie 字段查找监测期间持续使用的持久化 Cookie, 对频繁变换的客户端地址进行标注, 从而得到大量唯一 Cookie 对应的地址列表。实验只记录了第一天产生并在观察过程中持续使用的持久化 Cookie。标记过程不考虑在

监测期间出现的新 Cookie, 以防止用户标签准确性产生偏差。其次, 由于一个用户通常会生成多个 Cookie, 数据收集过程将具有相同地址的不同 Cookie 列表进行聚合, 以获取单个用户在流量监测期间产生的所有客户端地址。最后, 数据收集工作在 TLS 流量中搜索这些地址的通信记录, 从而获得 1.7k 用户和它们生成的 2.6k 地址的加密流量数据。由于关联模型需要成对的客户端地址的网络行为知识图作为输入, 实验在训练集、验证集或测试集中将任意两个地址组合成对。根据用户标签最终生成 150 万样本对的关联标签。当两个地址属于同一用户时, 关联标签为 1。否则, 关联标签为 0。

网络公害流量标注。为了获取网络公害行为的流量数据进行方法的实验评估, 我们提取 TLS ClientHello 报文中的 SNI 信息, 并采取主动验证的方式对数据集中客户端地址所访问的服务进行验证访问。当被访问的网站被我们判定为网络公害服务时, 这些客户端地址将被我们标记为网络公害用户。这些网站或应用主要包括赌博、色情等受国家监管和封禁的网络内容。最终, 我们收集到的公害网络行为流量占比约为 1%。公害主体关联技术的任务是在客户端地址频繁变更的海量背景流量下, 实现对网络公害用户变化的客户端地址关联, 从而实现目标恶意用户的长期追踪与溯源分析工作。

表 2 展示了数据集的基本组成。数据集中存在多种用户的不同网络行为, 因此具有较为良好的用户数据分布。例如, 实验统计了用户访问的 Top SNI 信息, 虽然排在前三的 SNI 域(*.google.com, *.adobe.com 和 *.microsoft.com)的访问量占比超过总访问量的三分之一, 但其余每个域的访问率都不高, 这表明用户在流量监测的数据集中保持着广泛的上网习惯和复杂的行为。为了更好地理解行为知识图中的元信息组成, 我们分析了为每个具有 1 个月背景知识的客户端地址构建的图数据统计。由于大多数客户端地址都是短期存活的临时地址, 每个客户端地址在 1 个月的观察期间平均访问 5～6 个在线服务。此外, 考虑到丢包等问题造成的观察偏差, 少数 TLS 连接不包含 ClientHello、ServerHello 或 Certificate 消息, 导致每个知识图中平均有 3.8 个客户端指纹节点和 41.3 个服务端指纹节点。最后, 实验使用前 3 个月的数据进行训练, 第 4 个月的数据用于验证, 第 5 个月的数据用于测试, 并保持网络公害标记用户在数据集中的占比为 1%。所有用于测试的用户被排除在训练集中。关联技术的现实目的是在历史流量的标签数据集上训练地址关联模型, 并利用学习到的关联函数对未来收集的数据进行地址关联, 从而实现公害行为溯源。

表 2　行为知识图中的元信息组成和数据集组成

Table 2　The composition of meta-information in a behavior knowledge graph and the composition of the dataset

元路径 Φ	关系(A-B)	A 数量	B 数量	实体	训练集	验证集	测试集
C-S	客户端节点-服务端节点	1.0	5.4	用户	1.0k (0.01k)	0.2k(0.002k)	0.5k(0.005k)
C-F	客户端节点-指纹节点	1.0	3.8	样本对	1.2M(4.9k)	0.1M(0.2k)	0.2M(1.2k)
S-F	服务端节点-指纹节点	5.4	41.3	流量数据	3 个月	1 个月	1 个月

注: 括号中数据表示公害用户的数据量, 例如 1.0k(0.01k)表示 1.0k 用户中公害用户数量为 0.01k

4.2　对比方法

先前的工作主要依靠用户画像、TLS 指纹和流序列等方法实现流量下的用户关联。其中, 本章实验部分实现了四种具有代表性的方法来与 PolluTracker 进行比较。

User IP Profiling[14]。User IP Profiling 方法通过客户端地址的所有目的 IP 来构建用户画像, 并使用贝叶斯分类器[28]来识别封闭数据集中的已知用户。为了在开放场景中应用地址关联来识别未知用户, 实验使用成对用户画像作为分类器的输入来评估地址关联的性能。

User SNI Profiling[12]。User SNI Profiling 是使用来自客户端的所有 TLS ClientHello 消息中的 SNI 作为用户兴趣标识。类似于 Herrmann 等人[13]的工作, 实验同样利用贝叶斯分类器输入成对的基于 SNI 的用户画像来关联用户活动。

Client Fingerprinting[19]。Client Fingerprinting 是提取 TLS ClientHello 消息的特定字段作为用户的客户端指纹, 并利用随机森林[29]来成对输入任意两个客户端的指纹信息来学习期间的相关性。

Deepcorr[20]。Deepcorr 利用流序列特征实现多场景下的关联任务[20-21,30]。为了与 Deepcorr 保持相同的设置, 实验还提取了每个客户端地址 300 个数据包的流序列以显示 Deepcorr 的性能。

4.3　评估指标

为了对比地址关联模型 PolluTracker 和其他对比

方法的关联效果, 实验评估的指标包括真阳性率、假阳性率、准确性和 ROC 曲线下面积。

真阳性率(TPR)。真阳性率指检测出来的真正例样本数除以所有真实正例样本数。该指标代表了所有真实正例样本中模型正确预测为正例样本的比例。

$$\mathrm{TPR}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$$

其中, 真正例(True Positive, TP)表示被模型预测为正类的正样本, 假反例 (False Negative, FN)表示被模型预测为负类的正样本。

假阳性率(FPR)。假阳性率指检测出来的假正例样本数除以所有真实反例样本数。该指标代表了所有真实反例样本中模型错误预测为正例样本的比例。

$$\mathrm{FPR}=\frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}}$$

其中, 假正例(False Positive, FP)表示被模型预测为正类的负样本, 真反例 (True Negative, TN)表示被模型预测为负类的负样本。

准确率(Accuracy)。准确率用于测量正确识别地址与目标用户样本相关或不相关的比例。

$$\mathrm{Accuracy}=\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}$$

ROC 曲线下面积(AUC)。AUC 度量是计算由多个阈值的 TPR 和 FPR 形成的 ROC 曲线下的面积, 在二元分类任务中经常使用。

$$\mathrm{AUC}=\frac{\sum I\left(P_+, P_-\right)}{M*N}\quad I\left(P_+, P_-\right)=\begin{cases}1 & P_+>P_-\\ 0.5 & P_+=P_-\\ 0 & P_+<P_-\end{cases}$$

其中, $\left(P_+, P_-\right)$ 是从数据集中的正样本和负样本中取出的一个正负样本对, P_+ 代表正样本的预测得分, P_- 代表负样本的预测得分。

4.4 模型设置

在数据预处理过程中, 实验设置将每个知识图谱中的最大节点数限制为 50, 最大节点属性长度限制为 50。因此, 邻接矩阵 A 和特征矩阵 X 的维数是 50×50。由于每个图的节点数或每个节点的属性长度之间的差异, 矩阵包含填充或截断操作。节点属性中的字符被编码为数字 token 值, 并基于 Word2Vec 完成节点特征的向量化处理过程, 特征矩阵 X 最终需要行归一化作为模型的输入。在训练 PolluTracker 时, 设置随机初始化参数并使用 Adam[31]优化模型, 实验将学习率设置为 0.005, 正则化参数设置为 0.001, 分层注意力参数化矩阵 a 的维度设置为 100。此外, 模型还将多头注意力的个数 K 设置为 4。激活函数 σ 为 LeakyReLU。根据孪生网络架构的模型输出, 方法将计算两个客户端地址的欧氏距离, 判定相似性的超参数阈值 m 为 20, 地址关联的阈值 η 为 10。训练过程使用 Early Stopping 并调整参数 Patient 为 100 来训练模型。模型参数总量为 124k, 训练时长为 3 小时。

5 方法评估

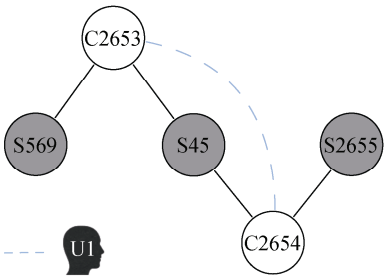
为了验证加密流量中网络公害行为主体关联技术的关联效果, 本节进行了广泛的实验并讨论方法的有效性。我们从注意力可视化分析、关联能力分析、消融实验、对抗实验、案例分析来验证实验效果。

5.1 注意力可视化分析

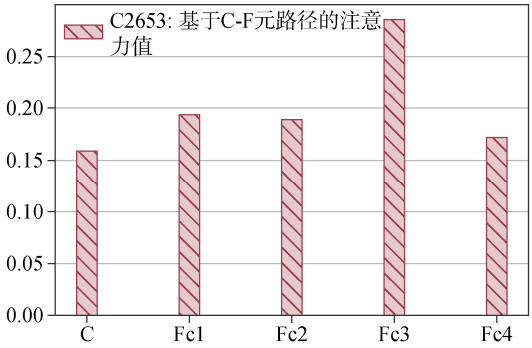
为了实施不同客户端地址行为之间的关联分析, PolluTracker 的一个显著特性是结合了注意力机制来帮助实现距离度量。图 5a 显示了数据集中的两个客户端地址的网络行为知识图例, 包括两个客户端节点(c2653 和 c2654)、三个服务节点(s569, s45, s2655)。在该实例中, c2653 和 c2654 都属于恶意用户 u1。

指纹级注意力分析。指纹级注意力分别利用客户端-指纹元路径(C-F)和服务端-指纹元路径(S-F)来学习客户端和服务端的唯一表示。图 5(b)和图 5(c)分别显示了地址节点 c2653 和服务节点 s569 的指纹级注意力。每个标签对应的指纹如表 1 所示。在 c2653 的客户端指纹中, Fc3 对任务贡献最大, 因为它是可用于识别恶意用户使用的客户端的浏览器参数。由于用户的客户端地址频繁变更, Fc1 显然对识别用户的有效性最低, 因此获得的指纹级注意力值是所有指纹中最低的。在 s569 的服务指纹中, Fs1 成为能够表示用户访问的服务属性的最具区分性的服务指纹。由于服务的域名可能部署在多个服务器地址上带来的复杂性。Fs25 的有效性次于 Fs1。Fs7 和 Fs8 也获得了相当高的注意力关注, 因为它们是来自服务器证书的有效特征。指纹级注意力最终通过学习语义信息获得唯一的客户端和服务表示, 从而帮助后续的服务级注意力进行地址节点表示的服务访问行为学习。

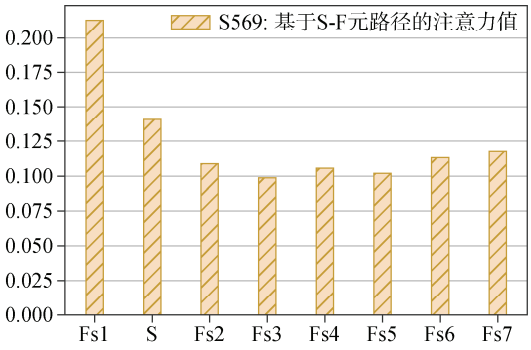
服务级注意力分析。在通过指纹学习获得了客户端和服务端表示后, 服务级注意力的目标是利用客户端-服务端元路径(C-S)学习客户端对不同服务端的访问行为。c2653 和 c2654 的服务级注意力如图 5(d)所示。可以看出, s45 的注意力值很高。该结果表示服务级注意力可以找到两个客户端地址访问的相同服务, 以帮助将它们链接到同一用户。此外, c2653 和 c2654 关于自身的注意力值均高于非两节点共同访问的其他服务。这表明 PolluTracker 还关注客户端



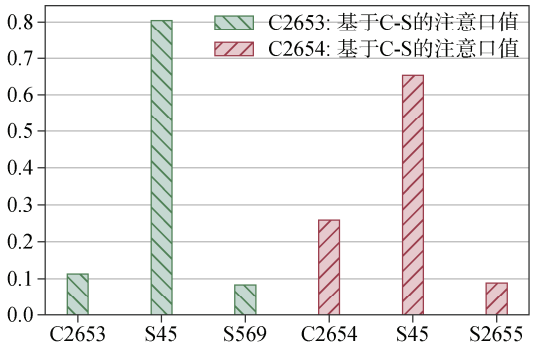
(a) 地址-服务关系图实例



(b) c2653 基于C-F元路径的注意力值



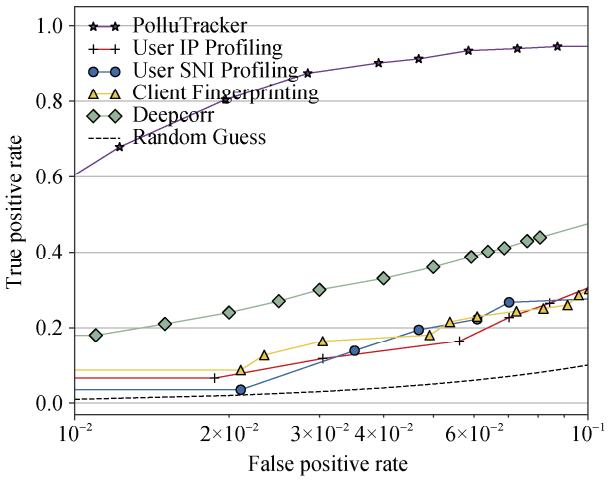
(c) s569基于S-F元路径的注意力值



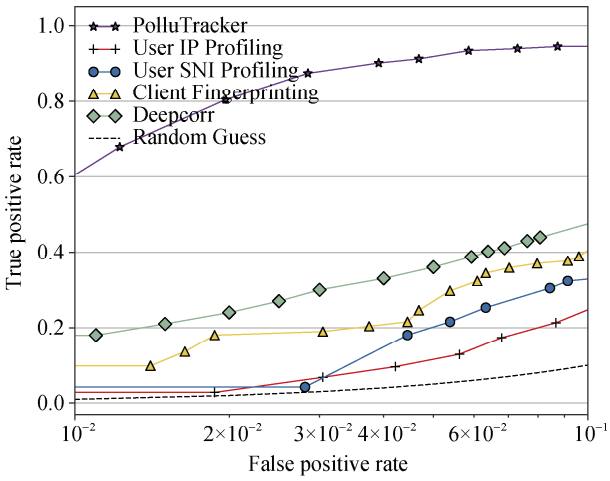
(d) c2653和c2654基于C-S元路径的注意力值

图 5　注意力可视化分析结果

Figure 5　Attention visualization analysis results



(a) 用户测试集1



(b) 用户测试集2

图 6　所有对比方法的地址关联 ROC 曲线

Figure 6　Address correlation ROC curves for all comparison methods

信息来识别用户，访问的相同服务和客户端自身信息的更多关注可以帮助去除复杂访问记录的影响，在一定程度上有助于识别和追踪具有复杂行为的恶意用户。

5.2　关联能力分析

为了探索加密流量下网络公害行为主体关联技术的有效性，我们测量了 PolluTracker 在网络公害用户和正常用户的地址形成的任意地址对上进行了关联性能测试。实验通过构建训练地址对样本和测试地址对样本来综合评估关联模型在关联任务上的表现。我们随机采样了训练集和测试集中的用户进行了更为鲁棒的实验验证，图 6 显示了两次采样数据集下模型与其他对比方法的 ROC 曲线。可以看出，PolluTracker 显著优于先前的关联算法，其 ROC 曲线

与其他方法之间存在较大差距。例如, 对于目标 FPR = 4×10^{-2}, PolluTracker 实现了 0.90 的 TPR, 而所有对比方法提供的 TPR 都小于 0.40。显著的效果提升是因为 PolluTracker 可以建模一个为成对客户端地址量身定制的关联函数, 并对地址的相关元信息知识进行适度学习。方法不但能够区分正常用户和恶意用户的访问行为, 还能够实现同一用户在不同客户端地址下的访问行为关联。表 3 显示了方法在两个指标下的实验结果。由于测试数据集包含来自训练集中未知用户的地址, 地址关联模型 PolluTracker 保持了强大的现实应用能力, 从而可以关联开放场景数据集上的未知地址。

表 3 所有方法在关联任务下的 AUC 和 Accuracy 指标

Table 3 The AUC and accuracy performance of All baselines under the correlation task

模型	用户测试集 1		用户测试集 2	
	AUC	Acc	AUC	Acc
User IP Profiling	0.785	0.711	0.683	0.630
User SNI Profiling	0.777	0.693	0.632	0.527
Client Fingerprinting	0.808	0.751	0.794	0.829
Deepcorr	0.826	0.802	0.819	0.855
PolluTracker	0.966	0.992	0.977	0.990

5.3 消融实验

除了展示针对特定任务的实验外, 本章还通过评估模型的变体进行消融实验研究, 以此来充分表明模型的优越性。表 4 显示了消融实验的所有结果。由于在构建网络行为知识图的过程中, 方法收集了客户端和服务端指纹作为流量特征关联的元信息, 为了观察两种指纹的重要性, 实验通过在构建知识图时分别移除所有客户端指纹节点或所有服务端指纹节点实现了 PolluTracker-Client 和 PolluTracker-Server 两种变体。在实验结果中, PolluTracker-Client 的效果相比 PolluTracker-Server 较差。这表明当 PolluTracker 缺少客户端指纹时, 由于性能下降更多, 客户端指纹对公害行为关联任务的贡献更大。为了探索分层注意力中每个注意力级别的有效性, 本章实验还提出了三种注意力变体, 包括 PolluTracker-CF、PolluTracker-SF、PolluTracker-CS, 这三个变体分别删除了模型基于元路径 C-F、S-F 和 C-S 的注意力值, 并为每个元路径所连接的节点分配相同的重要性权重。与 PolluTracker 相比, PolluTracker-CS 的效果急剧下降, 这表明网络访问行为信息对于关联任务至关重要。在三种注意力中, 元路径 S-F 贡献最少。 每个级别的注意力在关联效果方面都可以提供有效的提升, 最终使模型实现了极高的准确率。最后, 实验进一步研究了关联工作是否可以应用于其他带有孪生网络架构的图神经网络, 例如 GraphSAGE[22] 和 GAT[24]。结果表明, GNN 模型在学习通用的最终图表示方面具有强大的性能, 然而本文方法仍然获得了所有方法中的最优性能。

表 4 5 个月数据集上的消融实验结果

Table 4 The ablation study on the 5-month dataset

模型	用户测试集 1		用户测试集 2	
	AUC	Acc	AUC	Acc
PolluTracker-Client	0.906	0.902	0.944	0.930
PolluTracker-Server	0.920	0.911	0.953	0.949
PolluTracker-CF	0.909	0.879	0.950	0.920
PolluTracker-SF	0.912	0.906	0.968	0.943
PolluTracker-CS	0.781	0.687	0.847	0.880
Siamese GraphSAGE	0.942	0.908	0.933	0.913
Siamese GAT	0.955	0.922	0.960	0.958
PolluTracker	0.966	0.992	0.977	0.990

表 5 流量混淆方法下的 PolluTracker 的准确率

Table 5 PolluTracker's accuracy with traffic obfuscation methods

混淆方法	方法实现	用户测试集 1	用户测试集 2
C-Random	随机伪造地址	0.855	0.905
CF-Random	随机伪造浏览器参数	0.878	0.897
CF-Background	多种浏览器的背景流量	0.871	0.922
SF-Background	多种服务的背景流量	0.893	0.910
方法组合	四种混淆方法组合使用	0.705	0.769

5.4 对抗实验

尽管现有的网络公害治理技术针对一般的公害行为能够进行有效的检测, 然而面对具有逃逸意识的恶意用户常常束手无策, 恶意用户通过流量混淆等手段来实现对于模型的对抗行为。因此, 探索对抗场景下的模型关联效果也尤为重要。对此, 表 5 中实现了四种类型的流量混淆方法。C-Random 和 CF-Random 分别表示使用随机伪造的地址或浏览器参数来获取随机的客户端节点或随机的客户端指纹节点的组合。CF-Background 和 SF-Background 是添加使用不同浏览器或访问不同在线服务的背景流量的方法。所添加的背景流量与每个用户的原始流量比值为 1:1。结果表明, 由于关联模型通过关注多种类型元信息的重要性以确定地址间的关联关系, 因此每种混淆方法的单一使用都不足以实现有效的对抗效果。当组合所有四种流量混淆方法时, 模型的关联准确率才有显著降低, 这表明实现对抗关联工作

需要严格限制基于地址和流量的所有特征的使用。PolluTracker 的强大关联表现需要极高的对抗成本才能达到较为有效的对抗效果。

5.5 案例分析

我们考虑通过实际的案例分析来介绍本文方法在网络公害行为追溯任务中的实用价值。在我们对数据集流量中客户端地址访问的域名进行主动验证的过程中, 我们发现了域名为“*141tube.com”(“*”符号为匿名处理的通配符)的黄色网站和与该域有通信行为的一个客户端地址。为了溯源该公害用户的流量下所有网络行为, PolluTracker 将输入该目标地址的网络行为知识图和流量下所有待测的客户端地址的网络行为知识图进行搜索和匹配, 从而实现目标公害用户的行为关联与画像分析工作。结果表明, PolluTracker 能够正确检举出 5 个月内该用户的所有变化的客户端地址。通过深入分析我们发现, 如表 6 所示, 该用户的客户端加密套件等指纹信息基本没有改变, 同时, 不同地址的相同服务访问行为也能够帮助模型实现该用户的行为关联(如客户端地址 1 和客户端地址 2 均共同访问了相同的 SNI 域)。此外, 利用关联模型我们还找到了该网站用于逃逸审查的变体域名(如, *140tube.com 和*143tube.com), 表明 PolluTracker 能够有效检测出相似的网络服务域信息。最终, 利用有效的客户端和服务端画像学习, 关联方法能够正确关联公害主体的所有变更的客户端地址, 从而实现恶意用户的行为链分析和溯源。

表 6 网络公害行为案例分析

Table 6 Case study of cyber pollution behavior

公害地址	加密套件	访问的相似 SNI 域
客户端地址 1	[0x6a6a, 0x1303, 0x1301, 0x1302, 0xcca9, 0xcca8, …]	*141tube.com
客户端地址 2	[0x6a6a, 0x1303, 0x1301, 0x1302, 0xcca9, 0xcca8, …]	*141tube.com
客户端地址 3	[0x6a6a, 0x1303, 0x1301, 0x1302, 0xcca9, 0xcca8, …]	*140tube.com

6 讨论

在网络监管体系逐渐走向成熟之际, 网络行为分析工作仍然面临诸多困难和挑战。尽管本文提出的加密流量下网络公害行为主体关联算法一定程度上缓解了网络公害治理问题, 我们讨论本文方法可能存在的限制条件, 并探讨网络公害行为主体关联技术的未来工作。

动态地址的用户标注数据集获取。动态地址的用户标注数据集获取工作是实现地址关联技术的首要前提。本文基于 HTTP 持久化 Cookie 实现了客户端地址频繁更迭的用户数据标注工作, 然而事实证明, 该方法需要大量的时间成本和观测资源才能形成颇具规模的实验数据集。未来地址关联技术需要更加切实有效的数据集获取手段来降低这些时间和资源代价的门槛, 从而体现大规模应用价值。利用自动化工具模拟产生用户数据集是可考虑的方向之一。如何使用自动化设置来产生符合逻辑的用户行为数据是未来待解决的研究内容。

TLS 1.3 全加密场景下的流量检测。面向未来加密场景的 TLS 1.3 协议正在逐渐运用到实验和生产环境中。与先前的 TLS 加密流量场景不同的是, TLS1.3 将加密握手阶段的指纹特征并且具有与以往不同的握手模式。在全加密场景下, 利用握手报文的明文信息实现特征构建与流量检测的工作将不再有效。未来工作需要探索新加密场景的其他指纹信息或更多可靠的统计特征来实现流量检测任务。新场景带来的全新通信环境下, 流量检测模型的实际效果也需要更多的未来工作进行后续的探索、实践与评估。

7 结论

本文重点探索了网络公害治理领域的研究内容。具体而言, 我们提出了一种加密流量下网络公害行为主体关联技术。方法首先提取加密流量中的指纹特征和访问行为来为每一个客户端地址构建网络行为知识画像, 随后, 利用图神经网络和孪生网络构建的地址关联模型 PolluTracker, 方法能够判断任意两张客户端地址的网络行为知识图是否属于同一用户, 以此来解决流量加密和客户端地址更迭场景下的网络公害行为溯源问题。广泛的实验表明, 我们的方法能够以 99%的准确率实现公害主体的地址关联工作。面对实际场景的目标公害用户追踪任务, 方法能够有效检出目标用户的所有客户端地址, 从而实现网络公害行为的溯源分析工作。

参考文献

[1] Sun G L, Xue Y B, Dong Y F, et al. An Novel Hybrid Method for Effectively Classifying Encrypted Traffic[C]. *2010 IEEE Global Telecommunications Conference GLOBECOM 2010*, 2011: 1-5.

[2] Alshammari R, Zincir-Heywood A N. Can Encrypted Traffic Be Identified without Port Numbers, IP Addresses and Payload Inspection?[J]. *Computer Networks*, 2011, 55(6): 1326-1350.

[3] Taylor V F, Spolaor R, Conti M, et al. AppScanner: Automatic Fingerprinting of Smartphone Apps from Encrypted Network Traffic[C]. *2016 IEEE European Symposium on Security and Privacy*, 2016: 439-454.

[4] Sirinam P, Imani M, Juarez M, et al. Deep Fingerprinting: Undermining Website Fingerprinting Defenses with Deep Learning[C]. *The 2018 ACM SIGSAC Conference on Computer and Communications Security*, 2018: 1928-1943.

[5] Liu C, He L T, Xiong G, et al. FS-Net: A Flow Sequence Network for Encrypted Traffic Classification[C]. *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, 2019: 1171-1179.

[6] Ede T, Bortolameotti R, Continella A, et al. Flowprint: Semi-supervised mobile-app fingerprinting on encrypted network traffic [C/OL]. *27th Annual Network and Distributed System Security Symposium,* 2020.

[7] Shen M, Zhang J P, Zhu L H, et al. Accurate Decentralized Application Identification *via* Encrypted Traffic Analysis Using Graph Neural Networks[J]. *IEEE Transactions on Information Forensics and Security*, 2021, 16: 2367-2380.

[8] Lin K D, Xu X L, Gao H H. TSCRNN: A Novel Classification Scheme of Encrypted Traffic Based on Flow Spatiotemporal Features for Efficient Management of IIoT[J]. *Computer Networks*, 2021, 190: 107974.

[9] Lin X J, Xiong G, Gou G P, et al. ET-BERT: A Contextualized Datagram Representation with Pre-Training Transformers for Encrypted Traffic Classification[C]. *The ACM Web Conference 2022*, 2022: 633-642.

[10] Hou C S, Shi J Z, Kang C C, et al. Classifying User Activities in the Encrypted WeChat Traffic[C]. *2018 IEEE 37th International Performance Computing and Communications Conference*, 2019: 1-8.

[11] Wang J B, Cao Z G, Kang C C, et al. User Behavior Classification in Encrypted Cloud Camera Traffic[C]. *2019 IEEE Global Communications Conference*, 2020: 1-6.

[12] Kumpošt M, Matyáš V. User Profiling and Re-Identification: Case of University-Wide Network Analysis[C]. *Trust, Privacy and Security in Digital Business*, 2009: 1-10.

[13] Herrmann D, Banse C, Federrath H. Behavior-Based Tracking: Exploiting Characteristic Patterns in DNS Traffic[J]. *Computers & Security*, 2013, 39: 17-33.

[14] Gonzalez R, Soriente C, Laoutaris N. User Profiling in the Time of HTTPS[C]. *The 2016 Internet Measurement Conference*, 2016: 373-379.

[15] Althouse J B, Atkinson J, Atkins J. JA3. https://github.com/ salesforce/ja3. 2020.

[16] Brotherston L. FingerprinTLS. https://github.com/LeeBrotherston/. tls-fingerprinting. 2020.

[17] Majkowski M. SSL fingerprinting for pdf. https://idea.popcount.org/2012

[18] Husák M, Cermák M, Jirsík T, et al. Network-Based HTTPS Client Identification Using SSL/TLS Fingerprinting[C]. *2015 10th International Conference on Availability*, *Reliability and Security*, 2015: 389-396.

[19] Anderson B, McGrew D. OS Fingerprinting: New Techniques and a Study of Information Gain and Obfuscation[C]. *2017 IEEE Conference on Communications and Network Security*, 2017: 1-9.

[20] Nasr M, Bahramali A, Houmansadr A. DeepCorr: Strong Flow Correlation Attacks on Tor Using Deep Learning[C]. *The 2018 ACM SIGSAC Conference on Computer and Communications Security*, 2018: 1962-1976.

[21] Bahramali A, Houmansadr A, Soltani R, et al. Practical Traffic Analysis Attacks on Secure Messaging Applications[C]. *Proceedings 2020 Network and Distributed System Security Symposium*, 2020.

[22] Hamilton W L, Ying R, Leskovec J. Inductive Representation Learning on Large Graphs[C]. *The 31st International Conference on Neural Information Processing Systems*, 2017: 1025-1035.

[23] Kipf T N, Welling M. Semi-Supervised Classification with Graph Convolutional Networks[EB/OL]. 2016: arXiv: 1609.02907. https://arxiv.org/abs/1609.02907

[24] Velickovic P, Cucurull G, Casanova A, et al. Graph attention networks [C/OL]. *6th International Conference on Learning Representations*, 2018.

[25] Wang X, Ji H Y, Shi C, et al. Heterogeneous Graph Attention Network[C]. *The World Wide Web Conference*, 2019: 2022-2032.

[26] Chopra S, Hadsell R, LeCun Y. Learning a Similarity Metric Discriminatively, with Application to Face Verification[C]. *The 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition- Volume 1 - Volume 01*, 2005: 539-546.

[27] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C/OL]. *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017,* 2017: 5998-6008.

[28] Manning C D, Raghavan P, Schütze H. Introduction to Information Retrieval[M]. New York: Cambridge University Press, 2008.

[29] Breiman L. Random Forests[J]. *Machine Learning*, 2001, 45(1): 5-32.

[30] Wang X Y, Reeves D S. Robust Correlation of Encrypted Attack Traffic through Stepping Stones by Manipulation of Interpacket Delays[C]. *The 10th ACM Conference on Computer and Communications Security*, 2003: 20-29.

[31] Kingma D P, Ba J. Adam: A Method for Stochastic Optimization[EB/OL]. 2014: arXiv: 1412.6980. https://arxiv.org/abs/1412.6980.

崔天宇 2023 年中国科学院大学网络空间安全专业获得博士学位。现任中关村实验室助理研究员。主要研究方向为网络流量分析与大模型安全。Email: cuity@zgclab.edu.cn

侯承尚 2020 年中国科学院大学网络空间安全专业获得博士学位。现任中国科学院信息工程研究所工程师, CCF 会员。主要研究方向为网络行为分析与对抗。Email: houchengshang@iie.ac.cn

刘畅 2020 年中国科学院大学网络空间安全专业获得博士学位。现任中国科学院信息工程研究所工程师, CCF 会员。主要研究方向为网络流量分析和信息安全。Email: liuchang@iie.ac.cn

石俊峥 高级工程师, 硕士生导师。2013 年获北京邮电大学工程硕士学位, 2021 年获科学院大学博士学位, CCF 会员。主要研究领域为网络测量与行为分析。Email: shijunzheng@iie.ac.cn

苟高鹏 研究员, 博士生导师, 2005 年、2008 年、2014 年分别获中国北京航空航天大学工学学士、硕士、博士学位, CCF 会员。主要研究领域为网络安全和网络异常检测。Email: gougaopeng@iie.ac.cn

熊刚 教授, 博士生导师, CCF 会员。主要研究领域为网络和信息安全的理论和技术, 累计在权威期刊和会议上发表了 100 多篇论文。Email: xionggang@iie.ac.cn