匿名通信系统隐藏服务定位技术研究综述

赵 娜[1),2)] 苏金树[1)] 赵宝康[1)] 韩 彪[1)] 邹鸿程[1)]

[1)](国防科技大学计算机学院 长沙 410000)

[2)](长沙师范学院信息科学与工程学院 长沙 410100)

摘 要 匿名通信系统诞生之初是为了保护通信实体身份的匿名性和网络中通信内容的隐私性、完整性,但随着匿名通信系统的广泛使用,其匿名性不断增强,在隐藏服务技术的支持下,匿名通信系统被不法分子滥用的情况愈演愈烈,在匿名通信系统隐藏服务技术支持下的暗网平台已然成为了"法外之地".站在网络监管部门的立场上,对匿名通信系统,尤其是匿名通信系统隐藏服务及其定位技术的研究是必要且紧迫的.在对匿名通信系统的基本属性、分类方法和工作原理介绍的基础上,对其隐藏服务的定位技术按照用户位置的不同分为客户端定位和服务端定位分别进行了研究和阐述,重点介绍网络流水印技术、网站指纹攻击方法等代表性的隐藏服务定位技术,同时对现有隐藏服务定位技术的发展现状和优缺点进行总结,最后展望匿名通信系统及其隐藏服务定位技术的未来研究方向.

关键词 匿名通信系统;隐藏服务;客户端定位;服务端定位

中图法分类号 TP391 **DOI 号** 10.11897/SP.J.1016.2022.00373

A Survey on Hidden Service Location Technologies in Anonymous Communication System

ZHAO Na[1),2)] SU Jin-Shu[1)] ZHAO Bao-Kang[1)] HAN Biao[1)] ZOU Hong-Cheng[1)]

[1)](*Department of Computer Science, National University of Defense Technology, Changsha* 410000)

[2)](*Department of Information Science and Technology, Changsha Normal University, Changsha* 410100)

Abstract With the rapid development of Internet, people are more and more concerned about their privacy. Anonymous communication systems are designed for protecting the anonymity of communication entities and the privacy and integrity of communication content in networks at the beginning. Along with the improvement of anonymous communication systems, more and more people prefer anonymous communication systems to access the Internet in order to protect their identities and secret. But on the other hand, just because of anonymity, anonymous communication systems have been abused by a large number of criminals nowadays, especially with the support of hidden services technologies, which can strongly protect confidentiality of senders and receivers during communication. Criminals buy and sell pirated software, pornographic videos, drugs, guns, private information and other illegal items through the dark Web platforms. The dark Web platforms have already become a place out of laws with the technical assistance of anonymous communication systems' hidden services technologies. Therefore, the research on the hidden service location technologies of anonymous communication systems is helpful for network supervision

收稿日期:2019-09-24;在线发布日期:2020-03-15.本课题得到国家重点研发计划项目(2018YFB0204301)、国家自然科学基金(61972412,61601483)、湖南省教育厅科学研究一般项目(19C0140)、长沙市杰出创新青年培养计划(kq1905006)、长沙师范学院校级课题(2019xjzkpy16)资助.**赵 娜**,博士,讲师,中国计算机学会(CCF)会员,主要研究方向为网络空间安全、匿名通信系统.E-mail:mailfromzn@qq.com.**苏金树**(通信作者),博士,教授,博士生导师,中国计算机学会(CCF)高级会员,主要研究领域为计算机网络、网络空间安全.E-mail:sjs@nudt.edu.cn.**赵宝康**,博士,副教授,中国计算机学会(CCF)高级会员,主要研究领域为计算机网络、网络空间安全.**韩 彪**,博士,副教授,中国计算机学会(CCF)会员,主要研究方向为网络空间安全、智能感知与网络通信.**邹鸿程**,博士,主要研究方向为网络空间安全、匿名通信系统.

and law enforcement departments to effectively collect evidence and accurately crack down on dark Web crimes, it is a necessary and urgent task to locate hidden users in anonymous communication systems for network security management, scholars from domestic and foreign have carried out a lot of research on it. The location technologies of anonymous communication systems have made some effective achievements, but the anonymity is getting stronger and stronger along with the development of location technologies at the same time. Besides, new anonymous communication systems are emerging, hidden service nodes are more hidden and migrate frequently, it is no doubt that the research on hidden service location technologies of anonymous communication systems is a severe and tough challenge to researchers. This paper makes an in-depth study of the existing literatures about anonymous communication systems and hidden service location technologies. Firstly, we summarize the basic tree properties of anonymous communication systems, classify and compare the existing anonymous communication systems according to the proxy mode, communication performance and network structure, also, we introduce traffic obfuscation and anonymous routing techniques in detail, which is used in anonymous communication systems to access and routing anonymously. Then, we identify the scope of hidden services and take Tor, the most famous anonymous communication system at present, as an example to analyze how hidden services work. On the basis of the above content, we classify the location technologies of hidden services into client-side location and server-side location depending on the users' location and analyze typical location methods such as network flow watermark technologies, website fingerprinting attack methods and so on in detail. Furthermore, we summarize several representative hidden service location technologies and analyze the characteristics, advantages and disadvantages of different types of these technologies. Finally, since numerous issues are still open and challenging in the domain of anonymous communication systems' hidden service location technologies, the future research trends and challenges of anonymous communication systems and hidden service location technologies are prospected at the end of this paper.

Keywords anonymous communication system; hidden service; client-side location; server-side location

1 引 言

随着互联网的广泛应用,人们越来越关心网络通信过程中的身份和内容隐私问题,由此诞生了提供匿名通信服务的匿名通信系统,自 1981 年 Chaum[1] 提出的匿名邮件系统开始,匿名通信技术在匿名邮件、匿名投票、信息共享等方面均得到了广泛的研究与应用,目前应用最广泛的匿名通信系统是最初由美国海军实验室开发的 Tor(The onion router)网络[2],图 1 显示了到 2019 年 6 月为止,Tor 网络的直接连接用户量达到了 300 万左右,且有明显的上升趋势.

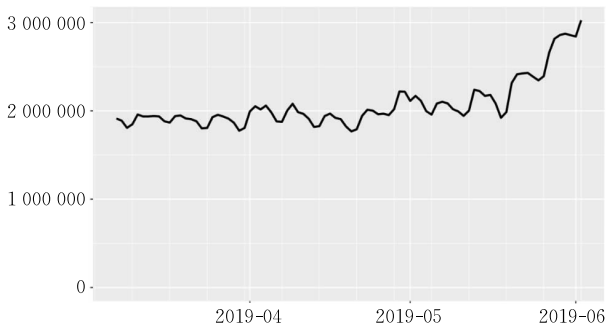


图 1 Tor 的直接连接用户量趋势图①
(其中横轴为时间,纵轴为 Tor 网络的直接连接用户量)

匿名通信系统设计的初衷是为情报人员、自由言论者等网络用户提供网络通信身份的匿名性和通信内容的隐私性、完整性服务,但随着匿名通信系统技术的不断发展和完善,系统用户量不断增大,用户群体也愈加多样化,不法分子利用匿名通信系统在

① Tor Metrix[EB/OL]. https://metrics.torproject.org

提供网络服务的同时可以帮助用户规避网络监管的这一特性，将其作为暗网(Dark Web)平台的技术支撑. 近年来暗网事件频发，这一隐藏在互联网之下的深层非法网络形式逐渐进入了大众的视野，受到了社会各界的广泛关注，各国政府对暗网的监管和打击力度不断加大，2017 年 7 月 20 日，美国执法机构宣布关闭当时全球最大的暗网交易平台“阿尔法湾”(Alpha Bay)，同时，荷兰执法机构宣布关闭全球第三大暗网交易平台“汉萨”(Hansa). 2019 年 5 月 3 日德国执法机构宣布成功摧毁全球第二大暗网交易平台“华尔街市场”(Wall Street Market)，该平台用户数超过 115 万，卖家超过 5400 个，在被关闭前，毒品、窃取数据、伪造证件和恶意软件等交易项目超过 6.3 万个. 我国对暗网治理工作也尤为重视①，公安部自 2018 年以来连续两年开展了“净网 2018”和“净网 2019”专项行动，主要整治对社会危害大、严重侵犯公民个人信息、网络攻击、网络诈骗、网络赌博和网络色情等违法犯罪行为，净网行动取得了一定的成效并在持续进行中. 腾讯安全云鼎实验室统计了到 2018 年 12 月为止的几大暗网市场商品分类，发现毒品/药物类超过 50%，其次是数字商品类，暗网市场中还充斥着各种色情、黑客、枪支、护照、假钞等违法内容，如图 2 所示.

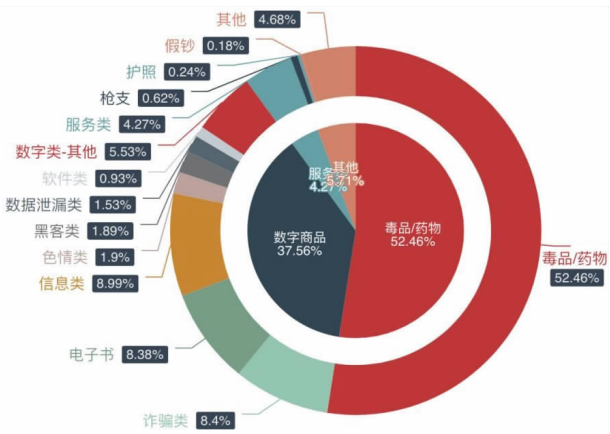


图 2　暗网主要市场商品类型分布①

暗网用户需要通过特殊的接入软件和加密的信道对暗网资源进行访问，匿名通信系统的隐藏服务技术为暗网平台的实现提供了主要的技术支撑，暗网用户通过匿名通信系统进行非法交易，可以保护交易双方的隐私，躲避执法部门的监管. 在匿名通信系统隐藏服务技术的保护下，暗网平台已然成为规避网络监管的安全场所，军火、毒品、色情和恶意软件等非法商品充斥其中，成为名副其实的“法外之地”，是网络监管的痛点和难点.

匿名通信系统隐藏服务定位技术的研究有助于帮助网络监管和执法部门对暗网犯罪进行有效取证和精准打击，国内外学者对此开展了大量的研究工作，通过网络流水印、交叉攻击、网站指纹攻击等方法可以在一定程度上破坏系统的匿名性，实现隐藏服务的定位，然而随着定位技术的发展，匿名通信系统自身的匿名性也在不断增强，新型的匿名通信系统也不断产生和发展，尤其是提供隐藏服务的节点较系统中其它节点更为隐蔽，且迁移频繁，因此，匿名通信系统隐藏服务的定位技术依然面临巨大的挑战.

国内外学者在匿名通信系统及其隐藏服务的研究过程中，形成了一系列的综述性文献，这些文章从不同的角度对匿名通信系统进行了描述：东南大学的 Luo 等人[3]从匿名通信和暗网关系的角度，首先对匿名通信和暗网的基本概念进行了明确的界定，并对二者关系进行了剖析，进而对现在主流的暗网形态、匿名通信的工作原理、关键技术进行了概括，最后分析了此领域的相关研究工作，为匿名通信与暗网治理的研究指明了方向；电子科技大学的 Zhuo[4]在其博士论文中对现有匿名通信系统进行了分类，分为基于代理的匿名通信系统、基于 Mix 的匿名通信系统、基于广播/组播的匿名通信系统和基于 P2P 的匿名通信系统；北京邮电大学 Lu 等人[5]对匿名通信系统的匿名性测量问题开展了全面且深入的研究；Kelly 等人[6]对现有匿名通信系统从有线、无线和混合式接入方式利用立体分类法进行了分类介绍，在分类方法上进行了创新；Erdin 等人[7]讨论了匿名通信系统的攻击方式，分为基于应用的攻击和基于网络层的攻击两种，同时文章还讨论了如何应对这些攻击的方法；Nepal 等人[8]针对 Tor 网络的攻击方式进行了研究并分类对比；Shirazi 等人[9]介绍了现有的匿名通信系统路由协议，并按 Mixnets、DC-nets、洋葱路由和基于 DHT 的路由协议进行了分类对比.

本文在上述文献的基础之上对匿名通信系统领域的文献进行了广泛、深入地调研，总结了现有匿名通信系统的 3 个基本属性，分别根据代理方式、通信性能、网络结构对现有匿名通信系统进行了全面的分类对比，详细介绍了匿名通信系统所使用的流量混淆技术和匿名路由技术；明确了匿名通信系统隐

① 腾讯云鼎实验室[EB/OL]. https://cloud.tencent.com/developer/article/1372920

藏服务涵盖的内容，以 Tor 网络为例介绍了隐藏服务的工作原理；将现有隐藏服务定位技术研究从用户位置的不同分为客户端定位和服务端定位，并对其中典型的定位方法进行了详细分析；最后对匿名通信系统隐藏服务定位技术的未来研究方向进行了展望.

2 匿名通信系统

随着互联网的普及，人们对网络通信过程中通信实体身份和内容匿名性的需求促使了匿名通信系统的产生和发展，匿名通信系统在向用户提供网络服务的前提下，通过匿名接入、匿名路由等技术达到保护通信实体和通信内容隐私的目的. 本节对匿名通信系统进行概述，首先在现有文献的基础上对匿名通信系统的基本属性进行抽象，然后从代理方式、通信性能和网络结构 3 种不同的角度对匿名通信系统进行详细的分类和对比，接下来对现有的匿名接入技术和匿名路由工作原理进行介绍，最后明确匿名通信系统隐藏服务的含义和范围，并以 Tor 的隐藏服务机制为例对隐藏服务工作原理进行介绍.

2.1 匿名通信系统基本属性

现有的大量匿名通信系统针对不同的应用场景提供不同程度的匿名服务，目前针对匿名通信系统尚无统一的精确定义，为更好地对匿名通信系统进行研究，本节通过分析各类现有匿名通信系统的特点，结合文献[6]和文献[7]的内容，将现有匿名通信系统的基本属性抽象为以下 3 个：

(1) 不可关联性(Unlinkability). 不可关联性指匿名通信系统中的通信实体在访问系统资源时，其它通信实体或网络攻击者无法通过其观测到的匿名通信流量关联到具体的消息发送方或者接收方，即通信流量与通信实体身份是不可关联的，如图 3 所示.

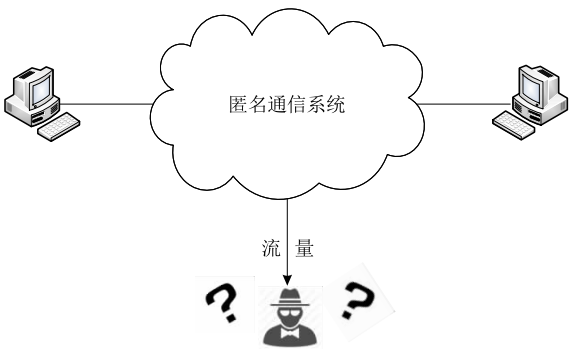


图 3 不可关联性示意图

(2) 不可辨识性(Unidentifiability). 不可辨识性指匿名通信系统的观察者/攻击者无法在一组相似的代理或组的集合中识别出真正的代理或组的身份等内容，不可辨识性可分为发送方不可辨识性、接收方不可辨识性和通信关系不可辨识性，如图 4 所示.

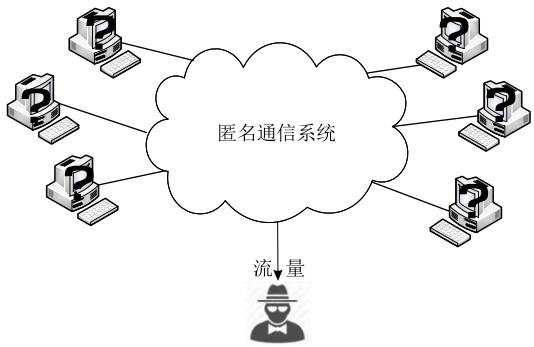


图 4 不可辨识性示意图

(3) 不可观测性(Unobservability). 不可观测性指观察者/攻击者不能从匿名通信系统中的任何其它实体观测到感兴趣的内容，不可观测性同样可分为发送方不可观测性、接收方不可观测性和通信关系不可观测性，如图 5 所示.

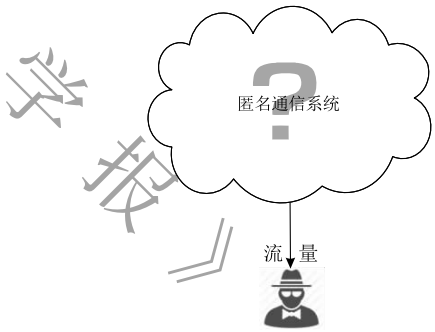


图 5 不可观测性示意图

(4) 匿名(Anonymity). 匿名指消息的发送者或接收者身份不可被识别的状态，根据匿名身份的不同又可分为发送方匿名(sender anonymity)、接收方匿名(receiver anonymity)和通信关系匿名(unlinkability of sender and receiver)[10]，不同的匿名通信系统根据不同的应用场景提供不同级别的匿名性.

图 6 显示了不可关联性、不可辨识性和不可观测性这 3 种属性与匿名的关系，匿名包含了不可关联性和不可辨识性两个属性，而匿名的目的是达到系统的不可观测性. 为了便于表述，全文中提及的“匿名性”不区分具体属于以上哪种基本属性，具有普适性意义.

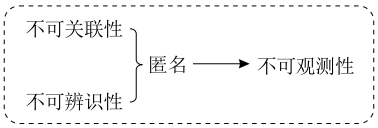


图 6　匿名属性关系示意图

2.2　匿名通信系统分类

国内外不同的学者由于其具体研究场景的侧重点不同，对匿名通信系统的分类方式不尽相同，本节综合现有文献的分类方式，从代理方式、通信性能和网络结构这 3 个方面对现有匿名通信系统进行分类.

(1) 根据代理方式分类

匿名通信系统可以根据其代理方式的不同分为基于单代理的匿名通信系统和基于多代理的匿名通信系统. 基于单代理的匿名通信系统利用单一的代理服务器，结合加密技术实现匿名通信，主要提供发送方的匿名性. 单代理匿名通信系统特点是简单易用，用户只需要简单的网络知识即可在本地进行部署和使用，典型的单代理匿名通信系统有 Anonymizer[11] 等；基于多代理的匿名通信系统则利用多重代理服务器组建匿名通信系统，主要提供收发双方的匿名性，基于多代理的匿名通信系统部署难度要比单代理匿名通信系统相对高、部署难度大(需要特殊的代理软件和系统参数配置)，但其匿名性更强，不仅可以提供发送方匿名服务，还可提供接收方和收发双方通信关系的匿名性，因此暗网平台往往选择基于多代理的匿名通信系统作为其技术支撑. 典型的多代理匿名通信系统包括 Mixnets[1]、Mixmaster①、Tor、Tarzan[12] 和 Freenet[13]、HORNET[14] 等. 基于多代理的匿名通信系统不依赖于某一个中间节点，往往通过多跳路由的方式提供匿名网络访问服务，图 7 为基于多代理的匿名通信系统原理图.

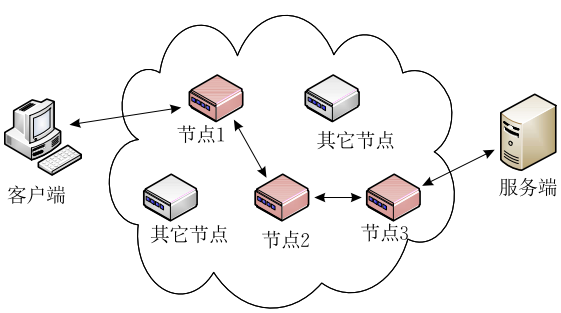


图 7　基于多代理的匿名通信系统原理图

(2) 根据通信性能分类

增加匿名性功能势必会影响整个系统的通信性能，而系统的通信性能主要体现在消息延迟上，因此，根据不同的“匿名/延迟”需求的应用场景，常将现有匿名通信系统分为高延迟匿名通信系统和低延迟匿名通信系统. 高延迟匿名通信系统可为用户提供强匿名性服务，适用于诸如匿名邮件系统等对延迟要求不高的非交互式网络应用，高延迟匿名通信系统的延迟可达几小时到几天不等，前文提到的 Mixnets 和 Mixmaster 等可归为此类，此外还有 Cyberpunk[15]、Miminion[16]、DC-Net[17] 等高延迟匿名通信系统；低延迟匿名通信系统的匿名性低于高延迟匿名通信系统，但是消息延迟时间短，一般只有几秒钟，因此低延迟匿名通信系统适用于网页浏览、实时通话等实时性要求高的网络应用，前文提到的 Tor、Freenet 等可归为此类，低延迟匿名通信系统还有 Crowds[18]、PANEL[19] 和 Loopix[20] 等，2017 年 Chaum 在 Mixnets 的基础之上提出的 cMix[21] 是一种轻量级的低延迟匿名通信系统，适用于移动端的匿名通信，Chen 等人[22] 提出的 TARANET 在网络层实现了高效的匿名通信.

(3) 根据网络结构分类

匿名通信系统根据所构建网络的组织结构不同又可分为集中式(非 P2P 结构)的匿名通信系统和分布式(P2P 结构)的匿名通信系统. 集中式结构的匿名通信系统，以 Tor 为例，需要在网络中部署全局的目录服务器以保存系统中所有中间路由节点的信息，系统的安全性很大程度上依赖于目录服务器的安全；随着 P2P 技术的兴起，无中心的分布式 P2P 的匿名通信系统也应运而生，基于 P2P 的匿名通信系统可以更充分地利用系统用户资源，平衡网络负载，具有更好的动态性和适用性，Crowds、Freenet、I2P、Herbivore[23] 等可归属此类，利用 BitTorrent 和 Bitcoin 技术的 ZeroNet 系统自 2015 年提出以来也得到了研究者们的广泛关注，此外值得一提的还有我国中国科学院自主研发的 P2P 全文信息检索系统 WonGoo[24-25]，Tan 等人[26] 提出的 StegoP2P 可以在现有 P2P 技术的基础上实现抵御常见流量审查的隐蔽通信.

根据上述 3 种对匿名通信系统不同的分类方式，表 1 对现有国内外典型匿名通信系统进行了汇总.

① Network Working Group Internet-Draft: Mixmaster Protocol Version2draft-sassaman-mixmaster-03. txt[EB/OL]. https://tools. ietf. org/html/draft-sassaman-mixmaster-03

表 1 国内外现有典型匿名通信系统研究汇总

研究工作	代理方式		通信性能		网络结构		匿名性		
	单代理	多代理	低延迟	高延迟	集中式	分布式	发送者匿名	接收者匿名	通信关系匿名
Mixnets	√			√	√		√	√	√
Mixmaster		√		√	√		√		√
Miminion		√		√	√		√	√	√
DC-Net		√		√		√	√	√	
cMix		√	√		√		√	√	√
Anonymizer	√		√		√		√		
Crowds		√	√			√	√		
WonGoo		√	√			√	√		
Tor(Onion2)		√	√		√		√	√	√
HORNET		√	√		√		√	√	√
I2P		√	√			√	√	√	√
Freenet		√	√			√	√	√	
Freegate		√	√			√	√		
Herbivore		√		√		√	√	√	
ZeroNet		√	√			√	√	√	√
StegoP2P		√	√			√	√	√	√
TARANET		√	√		√		√	√	√
PANEL		√	√		√		√	√	√
Loopix		√	√		√		√	√	√

2.3 匿名通信系统工作原理

匿名通信系统主要采用流量混淆技术和匿名路由技术为用户提供匿名接入和流量的匿名转发服务，以实现通信实体的匿名和通信内容的隐藏.本节针对流量混淆技术和匿名路由技术分别进行介绍.

(1)流量混淆(Trafic Obfucation)技术

为抵抗流量分析攻击，规避执法部门的网络监管，匿名通信系统通常会在接入网络之前采用流量混淆技术对用户流量进行混淆处理，再将混淆后的数据发送至网络中.Dixon 等人[27]将现有流量混淆技术归类为加密、随机化、拟态和隧道 4 种，并对比了每种方法的性能等特点，中国科学院的 Tan 等人[28]针对匿名通信系统提出了基于相对熵的不可观测性度量方法，Yao 等人[29]在文献[27]和[28]的基础之上进一步将随机化、拟态和隧道这 3 种流量混淆技术进行了形式化定义，并针对流量混淆技术的隐蔽性、计算开销和部署难度提出了具体的评价指标.

① 加密技术(Encryption).加密技术被认为是最传统的流量混淆技术，通用性强，使用范围广，但多数加密协议报文头部依然是明文，无法满足匿名通信系统不可观测性的要求，即达不到系统的匿名性需求.

② 随机化技术(Randomization).随机化技术利用流密码对数据包有效载荷的每一个字节进行处理，随机化技术处理后的流量不像任何现有的网络协议流量.由于随机化后的流量不像任何网络协议，因此随机化技术在使用黑名单场景时效果较好，相反地，当在白名单场景时，随机化技术将很容易失效.Tor 网络使用的 obfs 系列协议(obfs2/3[①]，obfs4[②])、Dust[30]、ScrambleSuit[31]等均使用了随机化技术对通信流量进行混淆处理.

③ 拟态技术(Mimicry).与随机化技术效果相反，拟态技术可以将匿名通信系统流量伪装成普通网络协议流量，最常用的是伪装成 HTTP 协议流量，拟态技术适用于白名单场景，但其缺点是系统性能欠佳.使用拟态技术的 SkypeMorph 可以将流量伪装成 Skype 流量[32]，FTE(Format Transforming Encryption)[33]技术利用正则表达式指定密文格式，可以使得深度报文检测系统(Deep Packet Inspection，DPI)[34]对协议进行错误分类，典型的拟态技术还有 Marionette[35]和 StegoTorus[36]等.

④ 隧道技术(Tunneling)."隧道"并不能算是匿名通信系统独有的概念，广义来讲，VPN、HTTPS 代理等使用加密通道技术的系统都可称之为隧道.但是在匿名通信系统中的隧道技术侧重于以抵抗网络监管为目的，比较典型的是 Tor 网络中最新使用的 Meek 技术[37]，Meek 的实现基于域名前置(Domain Fronting)技术，目的是让审查者误以为系统客户端在访问正常网站，达到躲避网络监管的

① Tor: pluggable-transports/obfsproxy[EB/OL]. https://gitweb.torproject.org/pluggable-transports/obfsproxy.git/tree/doc

② obfs4: The obfourscator[EB/OL]. https://github.com/Yawning/obfs4

目的，典型的隧道技术还有 CloudTransport[38]、Flashproxy[39]和 Telex[40]等.

(2) 匿名路由(Anonymous Routing)技术

为保证通信的匿名性，匿名通信系统在对流量混淆处理的基础上还在消息转发过程中进行匿名路由操作. 匿名路由技术在保证系统匿名性的同时还需保证通信系统的性能. 结合文献[9]的分类方法和匿名通信系统的使用现状，下面分别从基于 Mix 节点的匿名路由技术、基于洋葱路由的匿名路由技术和基于分布式哈希表的匿名路由技术三个方面进行介绍.

① 基于 Mix 节点的路由技术. 1981 年 Chaum 提出了基于 mix 节点混淆机制的匿名通信系统 Mixnets，并由此引发了匿名通信系统的研究热潮[41-42]. Mixnets 匿名性强但系统延迟高，主要用于匿名邮件和电子投票等服务，在 Mixnets 中，mix 节点首先对传输的消息进行加密处理，接下来当消息经过 mix 节点时进行乱序处理，目的是隐藏消息的输入输出关系，使得匿名通信系统的攻击者无法通过观测到的消息获得通信实体间的通信关系. 此外，mix 节点为增强系统的匿名性还会对消息进行一定的延迟处理. 为了达到系统节点信任分配的目的，Chaum 进一步提出了 mix 级联策略，该方法通过多个 mix 节点串联形成 mix 链，如图 8 所示为两个 mix 节点级联的原理示意图，消息在通过 mix 链时，每个 mix 节点按层次利用节点公钥对消息进行加密，同时每个节点利用私钥进行解密，目的是提取内层信息和下一跳的地址. 基于 Mix 节点的路由技术由于其节点级联机制，对于交叉攻击可以达到较好的抵御效果[43]，但固定节点的 mix 级联存在单点失效问题，因此 Mixmaster 等系统采用了 free-route 策略，free-route 策略中路由并不固定，网络中所有的 mix 节点均可以组成节点序列进行路由. 此外，针对 Mixnets 路由延迟高问题，Chaum 进一步提出了 cMix 方案[21]，cMix 在核心的实时通信阶段使用预计算来避免计算复杂度高的公钥加密操作，适用于轻量级的客户机上的应用程序.

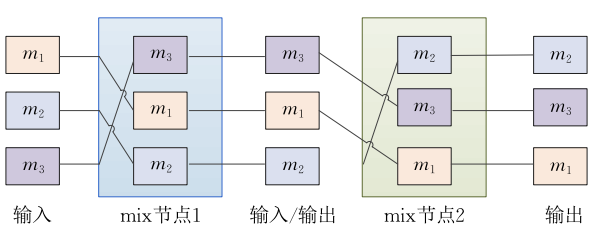


图 8　两个 mix 节点级联的原理示意图[9]

② 洋葱路由(Onion Routing). 1996 年美国海军实验室提出了基于“洋葱路由”的匿名通信架构[44]，现在使用范围最广、用户量最多的开源匿名通信系统 Tor(The onion router)使用的就是洋葱路由架构. 洋葱路由以防止流量分析、抵抗网络监管为目的，通过多跳代理机制对消息进行转发. 用户利用客户端代理选择一系列洋葱路由节点(通常为 3 跳)建立双向通信信道，层层加密的消息每经过一个通信链路的中继节点进行一次解密，最后经由出口节点解密完成后将明文发送至接收端. 消息传输的过程中每一个中间节点只知道链路中前序节点和后续节点的信息，从而保证了通信系统的匿名性. 图 9 为包含 3 跳洋葱路由的匿名通信过程示意图.

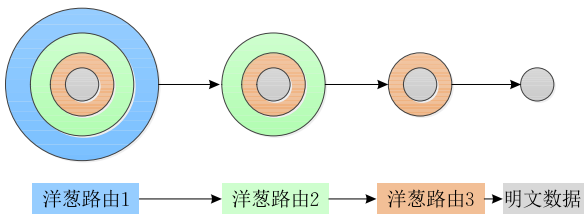
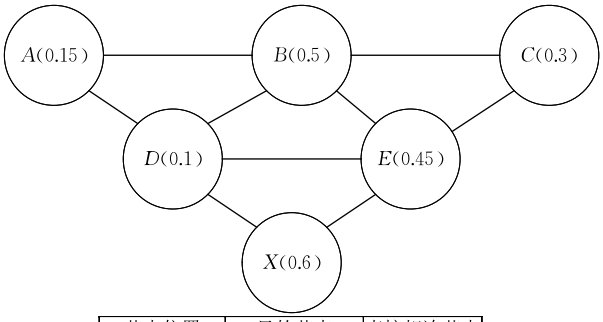


图 9　洋葱路由的匿名通信过程

③ 基于 DHT 的路由技术. 分布式哈希表(Distributed Hash Table，DHT)是一种用于存储(value，key)对的数据管理模型，可基于键值查找来定位相应的数据资源. 2001 年诞生于麻省理工的 Chord 协议[45]可以说是 DHT 的典型代表，随后 Pastry[46]等著名的 DHT 协议也相继诞生并在分布式系统中得到了广泛应用. DHT 路由技术因其无中心、分布式、结构化等特点，I2P、Freenet、Torsk[47]等匿名通信系统均采用了基于 DHT 的匿名路由技术，本节以 DHT 路由技术的典型代表 Freenet 系统路由为例进行说明. Freenet 是一个分布式的匿名信息存储和检索系统，在 Freenet 网络中，每个节点既是分布式文件的存储节点也是路由节点，如图 10 所示为 Freenet 网络拓扑和节点 X 的路由表[48]，X 的路由表中包含 X 所有可达的目的节点和对应的直接相连节点，网络中所有节点的位置由[0，1)区间内的 16 位有效数字的浮点数表示(所以 Freenet 中的节点在逻辑上构成一个环形)，X 在路由的过程中，根据数据包中的文件位置值索引下一跳(直接相连节点)，例如节点 X 收到文件位置值为 0.31 的文件请求，根据路由表可知 C 节点与该请求文件位置最近，因此 X 接下来会将该请求转发至节点 E.



节点位置	目的节点	直接相连节点
0.15	*A*	*D*
0.5	*B*	*D*、*E*
0.3	*C*	*E*
0.1	*D*	*D*
0.45	*E*	*E*

图 10 Freenet 网络拓扑和节点 *X* 路由表

2.4 匿名通信系统隐藏服务

匿名通信系统的隐藏服务，广义上指构建于匿名通信系统之上的服务，狭义上特指 Tor 提供的隐藏服务，即 Tor Hidden Service. 暗网用户可利用匿名通信系统发布隐藏服务（Web 浏览、文件共享等）或通过匿名通信系统访问其他用户提供的隐藏服务，因此从用户位置的角度，客户端隐藏和服务端隐藏均属于隐藏服务范围.

Tor 网络是目前拥有用户量最多的匿名通信系统，Tor 从 2004 年开始提供隐藏服务，传统的 Tor 网络采用如 2.3 节所述的“3 跳”工作方式，而当用户使用隐藏服务时，客户端需要通过“引入节点”和“汇聚节点”与隐藏服务器建立“6 跳”连接，形成如图 11 所示的通信链路. 具有隐藏服务功能的 Tor 网络组成要素及工作过程如下[49]：

(1) 客户端代理（Onion Proxy，OP）. 客户端代理指的是在客户端使用 Tor 浏览器时，用户必须通过 Tor 浏览器才能接入 Tor 网络. 客户端代理负责发起通信、创建链路、选择路径并进行通信节点间的密钥协商. 客户端代理将待发送的应用数据进行划分并封装成信元（Cell，Tor 的基本传输单元），每个信元大小均为 512 字节.

(2) 洋葱路由（Onion Router，OR）. 洋葱路由负责在网络中对信元进行匿名转发，目前 Tor 网络中包含中继（Relay）和网桥（Bridge）两种功能的转发节点，Bridge 节点可提供比 Relay 节点更强的匿名性. 另外，根据洋葱路由位置的不同又可分为入口节点、中间节点和出口节点. 客户端代理对待发送的数据进行层次加密（若链路为 3 跳，则进行 3 层加密），数据每经过一个中间节点解密一次，到出口节点后数据全部解密完成并发往目的端.

(3) 目录服务器（Directory Server，Dir）. 可以说目录服务器是 Tor 网络中最重要的组成部分，目录服务器负责记录并管理整个网络中所有的洋葱路由信息、节点公钥，也包括隐藏服务器的注册信息和公钥.

(4) 隐藏服务器（Hidden Server，HS）. 隐藏服务器指提供隐藏服务内容的服务器，隐藏服务包括 Web 服务、文件共享、实时通话等服务，隐藏服务器生成并发布特定格式的资源定位符（〈z.onion〉的形式，其中“z”由 16 位的随机字符组成）供用户访问.

(5) 引入节点（Introduction Point）. 引入节点在通信链路建立时使用，由隐藏服务器选择的若干中继节点组成，隐藏服务器将选中的引入节点发布至目录服务器，以供客户端选择并与之建立连接，客户端接下来将选中的汇聚节点信息发送至引入节点，引入节点再将此信息转发至隐藏服务器. 可见，引入节点的功能可理解为链路建立时隐藏服务器的“代理”.

(6) 汇聚节点（Rendezvous Point）. 汇聚节点由客户端代理选择，隐藏服务器需通过引入节点获取汇聚节点的信息. 汇聚节点在链路建立的过程中可理解为客户端的“代理”，在通信链路建立完成后作为转发节点参与数据转发.

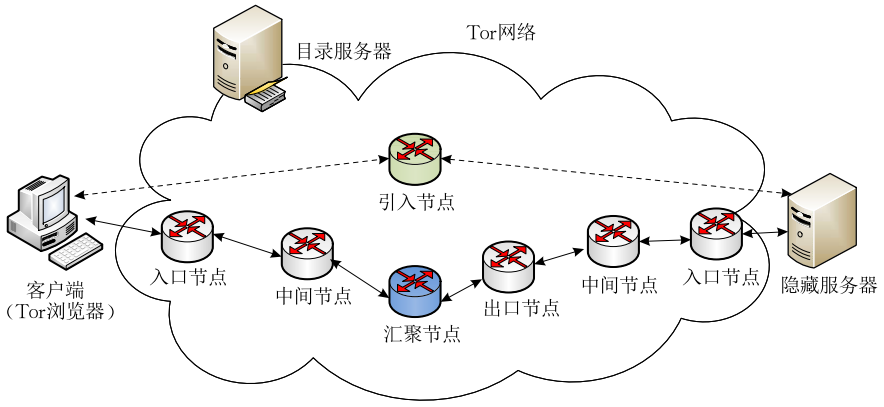


图 11 Tor 客户端和隐藏服务器之间的通信链路

在 Tor 的隐藏服务工作模式下，客户端仅知道引入节点信息，隐藏服务器端仅知道汇聚节点信息，客户端和隐藏服务器均无法获知对方信息，单独的引入节点和汇聚节点也无法定位客户端和服务器端. Tor 由此实现了客户端、服务端和通信关系的匿名性服务.

3　隐藏服务定位技术研究

匿名通信系统可以有效地为用户提供匿名服务，但系统的匿名性同时也使得执法者难以追踪基于匿名通信系统的网络犯罪，由于使用了流量混淆技术，基于端口号和基于深度包检测的传统流量识别技术在匿名通信系统中失去了用武之地. 网络监管部门很难对匿名通信系统，尤其是具有隐藏服务功能的匿名通信系统的流量进行识别、分析以及进一步对隐藏服务用户进行追踪定位. 针对匿名通信系统的隐藏服务被滥用的情况愈加严重，匿名系统流量难以识别和定位的问题，许多学者对匿名通信系统隐藏服务的定位问题进行了研究，本章将现有匿名通信系统隐藏服务定位的技术按用户位置的不同分为隐藏服务的客户端定位和服务端定位，并分别进行介绍.

3.1　客户端定位技术

匿名通信系统中隐藏服务的客户端指系统隐藏服务的请求者，对隐藏服务的客户端进行定位可直接且有效地帮助网络监管和执法部门对使用隐藏服务的用户进行溯源取证. 隐藏服务客户端的定位主要使用网络流水印、交叉攻击等技术方法，本节重点对网络流水印技术进行分类介绍，随后介绍几种典型的交叉攻击和其它定位客户端的方法.

3.1.1　网络流水印技术

水印技术[50]广泛应用于信息隐藏、版权保护等领域，近年来，有部分学者通过在网络流量中嵌入水印信息的方法进行流量分析[51-52]. 在匿名通信系统流量分析领域，利用网络流水印技术可以提高隐藏服务客户端定位的准确率，因此网络流水印技术成为了匿名通信系统隐藏服务定位研究领域的热点[53-55]. 图 12 为网络流水印技术的通用模型，水印嵌入系统在流量进入匿名通信系统之前进行水印信息的嵌入，在服务端（或服务端的入口节点）进行水印信息的提取和检测，一旦检测到包含水印信息的流量，即可通过流量中数据包的信息对用户身份（IP 地址）进行定位. 需要指出的是，网络流水印技术既

图 12　网络流水印技术的通用模型

可以对匿名通信系统的客户端进行定位，也可以定位服务端和收发双方的通信关系.

网络流水印技术是一种主动流量分析技术，根据水印嵌入的载体不同可将现有文献中的网络流水印技术分为基于数据包自身、基于时间间隔和基于流速率 3 种[54-55]，如图 13 所示. 下面分别进行介绍.

（1）基于数据包自身的网络流水印技术. 根据数据包自身特点，研究者们分别提出了基于包的有效载荷（Packet Payload）、基于包头部（Packet Header）、基于包大小（Packet Size）和基于数据包计数（Packet Counting）的网络流水印方案. Wang 等人提出的 SWT（Sleepy Watermark Tracing）框架[56]，通过在流量中嵌入虚拟空字符串（virtual null string）的方法进行流量的分析和追踪；Ramsbrock 等人[57]提出可通过填充字符的方法改变数据包的大小进行信号的嵌入，达到实时追踪数据流的目的. 直接在数据包中嵌入水印信息的方法通常为了保证水印信息可被识别而对嵌入的水印信息不进行加密处理，因此易被检测和过滤；文献[58]针对 Freedom 等匿名通信系统提出了数据包计数攻击的方法，攻击者同时观测发送方和接收方之间的进出数据包并进行计数，如果能够观察到通信双方之间数据包流的匹配模式则可以对它们之间的通信链路进行识别.

（2）基于时间间隔的网络流水印技术. 根据网

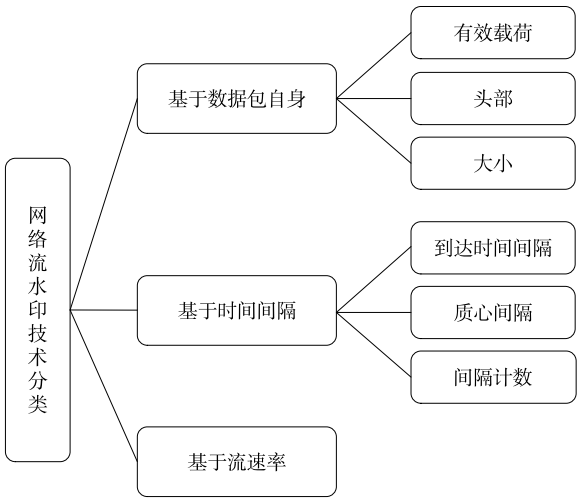


图 13　网络流水印技术分类

络中数据包的时间间隔特点，可通过控制网络中的特定节点，嵌入时间相关的水印信息，此类方法可分为基于包到达时间间隔(Inter-packet Delay)、基于质心间隔(Interval Centroid)和基于包间隔计数(Interval Packet Counting) 3 种网络流水印嵌入技术. Houmansadr 等人[59-60]提出的 RAINBOW 方案可通过调整一个包到达时间间隔的值嵌入水印信息，具有较好的隐蔽性，但由于要保存原始数据流的包时间间隔，该方案实时性效果欠佳；Wang 等人[61]提出了针对单代理匿名通信系统 Anonymizer 的网络流水印攻击方法，该方法利用包间时间间隔嵌入水印，仅需十分钟即可渗透 www.anonymizer.com 的网络屏蔽. Zhang 等人[62]在扩频流水印技术[63]的基础之上，提出了基于包的时隙质心间隔流水印的匿名通信追踪技术，该方法可用于追踪交互式和非交互式的匿名通信系统流量；Pyun 等人[64]提出了通过改变相邻时序对的数据分组数进行水印信息嵌入的方法，该方法适用于交互式流量追踪，方案可有效抵抗时间扰动，但难以抵抗多流攻击[65].

(3) 基于流速率(Traffic Rate)的网络流水印技术. 该方法通过流量注入的方式控制流量速率，以进行水印信号的嵌入. 该思想主要源于 Yu 等人[63]提出的 DSSS(Direct Sequence Spread Spectrum，直接序列扩频)方法，通过在网络中部署的流量干扰器注入无用流量以改变流速率，并将此作为水印信号对匿名系统的通信者进行追踪；Chan 等人[66]提出了针对 Tor 网络的中继识别水印方案，利用部署的恶意流量服务器造成 Tor 网络流量突发的状态，该方案通过定位网络中的中继节点以降低客户端的匿名性；Ling 等人[67]利用软件定义网络(Software Defined Networking，SDN)[68]交换机控制流速率的方式进行水印信息的嵌入以对通信关系进行确认，该方案在 SSH、OpenVPN 和 Tor 这 3 个系统中进行了真实环境的实验验证，检测率分别为 100%、100%和 95%，假阳性(False Positive)达到了 0，这为此方向的研究提供了一个有效的解决方案和新的研究思路.

3.1.2 交叉攻击方法

交叉攻击(Intersection Attack)方法是定位匿名通信系统隐藏服务客户端的一种有效手段. 文献[69]提出的交叉攻击方法，分析者通过观测到的流量对匿名通信系统的用户行为(访问的网站、上下线时间、邮件服务等)进行分类，建立不同的用户集合，通过取交集的方式逐步缩小集合范围，最终定位通信服务的发起者. 文献[70]提出的前序交叉攻击(Predecessor Intersection Attack)方法，攻击者利用网络中的一组中继协同工作，以记录通信服务的可能发起者，通过多轮的观测和记录最后将可疑发起者集合收敛至一个元素，完成定位. 文献[71]针对 Tarzan 和 Crowds 系统提出了时间交叉攻击(Timing Intersection Attack)的方法. 为抵抗此类时间交叉攻击，Tor 网络可选择每十分钟重新建立转发路径. 图 14 为交叉攻击节点收敛过程示意图.

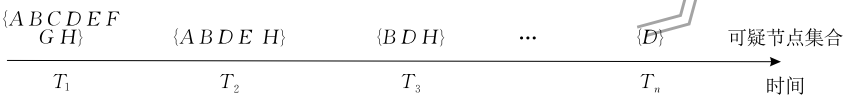


图 14 交叉攻击节点收敛过程示意图

3.1.3 其它方法

除网络流水印技术和交叉攻击方法外，还有一些其它有效的客户端定位方案，比如 Ling 等人提出的 TorWard 系统[72-73]可通过在网络中部署带宽较大的 Tor 出口节点以吸引 Tor 客户端的选择，进而捕获 Tor 流量，该系统进一步利用重放攻击的方式对客户端进行追踪；Zhuo 等人[74]针对 Freegate 系统的匿名通信流量进行了数据包粒度、数据流粒度等多粒度特征提取，以较低的误报率对匿名通信系统的客户端进行定位；Levine 等人[75]利用贝叶斯框架进行建模，提出了针对 Freenet 下载用户的定位方法，该方法只需要一个对等点以被动地分析邻居节点发送的流量，在实际 Freenet 环境中进行测试的结果假阳性率仅为 2%.

3.2 服务端定位技术

匿名通信系统中隐藏服务的服务端指的是系统隐藏服务的提供者，隐藏服务的服务端定位一直是匿名通信系统领域的研究热点，本节首先重点对该领域中网站指纹攻击方法进行分类和介绍，随后介绍几种典型的隐藏服务中的暗网平台数据采集与分析技术，最后对其它有价值的服务端定位方法进行介绍.

3.2.1 基于网站指纹攻击方法

网站指纹攻击(Website Fingerprinting Attack)[76]方法在匿名通信系统研究中是一种典型的被动流量分析技术，在攻击的过程中，攻击者需要预先收集目标网站的指纹信息(报文大小、方向、顺序等)，接下

来通过机器学习等方法对收集到的指纹信息进行建模，随后攻击者通过在匿名通信系统中部署的流量监控器捕获网络流量，最后将捕获到的流量进行预处理后与指纹特征库进行比对，从而实现对用户访问的网站进行追踪，达到定位匿名系统服务端的目的. 网站指纹攻击模块一般部署在客户端到匿名通信系统入口代理之间，图 15 为匿名通信系统网站指纹攻击过程示意图.

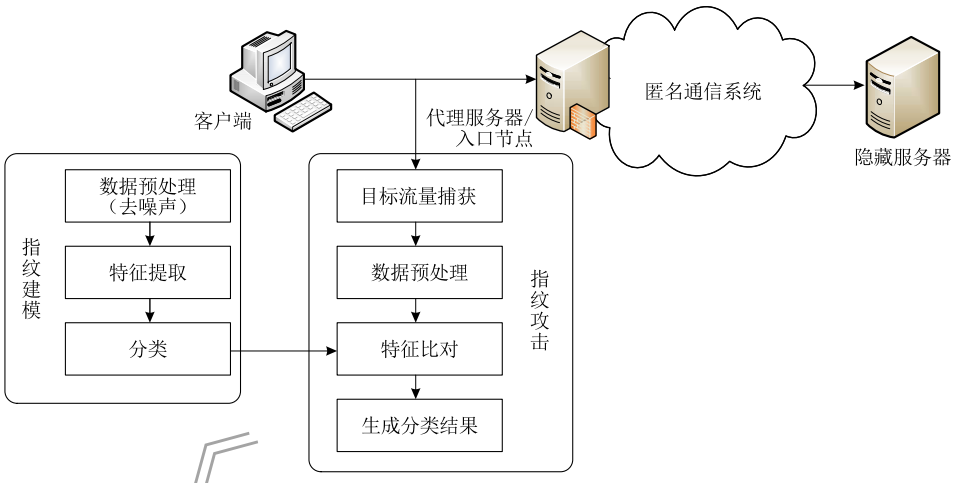


图 15　匿名通信系统网站指纹攻击过程示意图

网站指纹攻击方法根据不同类型的匿名通信系统特点，在对其服务端定位时需使用不同的指纹特征和分类方法，本节将现有的针对匿名通信系统的网站指纹攻击方法分别从基于代理节点和基于多代理匿名通信系统两个角度进行详细介绍.

(1) 基于代理节点的网站指纹攻击方法. 网站指纹攻击方法在 2002 年由德克萨斯大学的 Hintz[76] 提出，利用报文大小和方向特征对不同网页进行区分，但准确率较低. 此后大量学者对网站指纹攻击的流量识别技术进行了研究，在匿名通信系统的服务端定位领域也涌现了一批有价值的文献，2006 年 Liberatore 等人[77] 将网站指纹攻击方法引入到单代理匿名通信系统，利用朴素贝叶斯算法 (Naive Bayes)[78] 对 2000 个不同网站的超过 40 万条数据流进行仿真实验，验证了该方法的有效性，但基于报文大小的方法不适用于信元长度固定的匿名通信系统 (比如 Tor)；Zhuo 等人[79] 提出了针对 Shadowsocks 和 SSH(Secure Shell) 代理的网站指纹攻击方法，利用 Profile 隐马尔可夫模型 (Profile Hidden Markov Model，Profile HMM)[80] 对网站的超链接进行识别，此方案既可识别单个网页，也可识别整个网站，在实际匿名通信系统中取得了较好的攻击效果；Zeng 等人[81] 提出了一种基于流上下文和主机行为的 Shadowsocks 代理检测方法，从流量关系、主机流量行为、主机 DNS 行为三个方面提取 12 维特征建立检测模型，并在大数据平台进行了实验，实验证明了该方法对 Shadowsocks 流量的识别准确度达 93% 以上；Gu 等人[82] 针对 SSH 代理基于上下行流量的不同特性，利用隐马尔可夫模型对目标网站进行建模，实验显示该方案对网站识别的准确率可达到 96.8%.

(2) 基于多代理匿名通信系统的网站指纹攻击方法. Panchenko 等人[83] 利用支持向量机 (Support Vector Machine，SVM)[84] 方法对报文数量、时间和方向等特征进行提取分类，在对封闭世界 (closed-world) 的 775 个网站进行了实验的基础上在开放世界 (open-world) 成功对多代理匿名通信系统 (Tor，JAP[85]) 的网站进行了识别；He 等人[86-87] 分别对 TLS 指纹和报文长度分布特征进行提取，利用支持向量机对 Tor 流量进行识别和分类；Cai 等人[88] 利用改进核函数的支持向量机对网页进行建模，结合隐马尔可夫模型使每个网页对应一个 HMM 状态，通过马尔科夫链建立网页跳转模型构建网站模型，对 Tor 网络中网站的识别准确率达到了 90%；Wang 等人[89] 在 Cai 的方案的基础上进行了改进，使用 K 近邻 (K-Nearest Neighbor)[90] 算法进一步提高了识别的准确率，但此方法需要的流量特征比前者多；Hayes 等人[91] 提出的 k-fingerprinting 网站指纹方案，利用随机决策森林 (Random Decision Forests)[92] 对提供 Tor 隐藏服务的网站进行识别，并针对 Tor 的包填充、流量变形等防御手段进行了对比实验；Kwon 等人[93] 针对 Tor 的隐藏服务通信链路与普通链路不同的特点提出了电路指纹识别攻击方法，能够以 88% 的正确率对受监控的 50 个 Tor

隐藏服务器进行定位;Rimmer 等人[94]利用深度学习方法对收集的 300 多万个网络痕迹组成的数据集进行自动特征提取,并分别在封闭世界和开放世界进行了实验,验证了该方法的准确性和效率;Sirinam 等人[95]针对新提出的 Tor 网络防御方案 WTF-PA[96]和 Walkie-Talkie[97],利用深度学习中的卷积神经网络(Convolutional Neural Networks,CNN)[98]提出了 DF(Deep Fingerprinting)方案,该方案对 WTF-PA 和 Walkie-Talkie 的攻击精度分别为 90% 和 49.7%,Bhat 等人[99]利用 CNN 并结合包间隔时间方法,提出了 Var-CNN 网站指纹攻击方案,与 DF 相比,在开放世界中,Var-CNN 的真实阳性率(TPR)提高了 1%,假阳性率(FPR)低 4 倍,而在低数据情况下,TPR 提高了 13%,FPR 降低了 3.12%,这意味着,Var-CNN 大幅减少成功实施网站指纹攻击所需的训练数据量.这缩短了数据收集所需的时间,并降低了出现数据过时问题的可能性.

表 2 分别对现有典型基于代理节点和多代理匿名通信系统的网站指纹攻击技术进行了分类总结.

表 2 现有典型匿名通信系统网站指纹攻击方法分类

攻击目标	文献来源	主要指纹特征	方法	时间
基于代理节点	Liberatore 等人[77]	报文大小、方向	朴素贝叶斯	2006
	Zhuo 等人[79]	报文大小、方向	隐马尔可夫模型	2018
	Zeng 等人[81]	上下文流量特性、主机行为	大数据统计和关联技术	2019
	Gu 等人[82]	上下行流量特性	朴素贝叶斯、隐马尔可夫模型	2015
基于多代理匿名通信系统	Panchenko 等人[83]	报文数量、时间、方向	支持向量机	2011
	He 等人[86]	TLS 指纹、报文长度	支持向量机	2013
	He 等人[87]	报文时间	支持向量机	2014
	Cai 等人[88]	报文长度、访问频率	支持向量机、隐马尔可夫模型	2012
	Wang 等人[89]	数据包数量、方向、密度、突发流、总时间	*K* 近邻	2013
	Hayes 等人[91]	报文统计特征、顺序等	随机森林	2016
	Kwon 等人[93]	报文统计特征	*K* 近邻	2015
	Rimmer 等人[94]	自动提取	深度学习	2017
	Sirinam 等人[95]	自动提取	深度学习	2018
	Bhat 等人[99]	自动提取	深度学习	2019

3.2.2 隐藏服务平台数据采集与分析

在匿名通信系统隐藏服务领域的研究中,值得一提的还有服务端的数据采集与分析技术,这对网络监管和执法取证具有重要意义,国内外不少学者在此方向进行了研究与实践.服务端的数据采集与分析主要是利用网络爬虫技术对匿名通信系统隐藏服务端的网络资源进行采集和分析.网络爬虫(Web crawler)是一种网站自动索引程序[100],将网络爬虫引入暗网数据采集与分析领域,可实现匿名通信系统服务端资源的自动获取,从而对网站进行进一步的分析研究.Christin 等人[101]对隐藏服务网站“丝绸之路”进行了长达 6 个月的观测,得到了此暗网平台详细的数据分析;Yin[102]设计并实现了暗网扫描系统,利用该系统发现并分析了 Tor 隐藏服务内容、隐藏服务生命周期和隐藏服务网站内容的流行性;Bernaschi 等人[103]在 5 个月的时间内对 Tor 进行了 3 次爬取,并对 Tor 隐藏服务的波动性、结构性进行了分析; He[104]首先利用爬虫框架对 Tor 隐藏服务域名和网页进行收集,然后根据法律文本提取和构造关键词,并将其作为隐藏服务非法活动的分类依据,通过特征权重和贝叶斯分类器的方法可以达到 93.5% 的准确率;Park 等人[105]提出了 Tor 隐藏服务爬取和分析的系统框架,系统在微软 Azure 或者亚马逊 EC2 云平台利用 Docker 容器[106]对自动获取 Tor 隐藏服务的爬虫系统和网页内容分类器进行加速,对获取到的 1 万 6 千余条隐藏服务地址进行监测和分析.服务端数据采集与分析技术在匿名通信系统隐藏服务的服务端定位的研究中占据尤为重要的位置.

3.2.3 其它方法

匿名通信系统隐藏服务的服务端定位除基于网站指纹攻击方法外还有一些其它有效的匿名通信系统隐藏服务的定位方案.Ling 等人[107]提出了通过控制客户端、汇聚节点和入口节点进行合谋攻击的方法,可实现 Tor 隐藏服务器定位;Iacovazzi 等人在网络流水印技术的基础上提出了 INFLOW 方案[108],该方案利用 Tor 网络的拥塞控制机制首次提出“逆流水印”技术,方案可实现对隐藏服务器的定位;Elices 等人[109]针对 Tor 的隐藏服务提出了基于时间间隔的流量关联攻击方法,可有效定位 Tor 网络中的隐藏服务器;Tan 等人[110]对 Tor 隐藏服务的 Eclipse 攻击进行了深入的研究,并在实际 Tor

网络中进行了阻断隐藏服务器的实验；Matic 等人[111]提出了一个自动识别隐藏服务中的位置泄漏的工具 CARONTE，可以识别隐藏服务内容中的敏感信息或暴露服务器 IP 地址的配置，在 1974 个隐藏服务中应用 CARONTE，可以完全恢复其中 101 个(5%)IP 地址；Biryukov 等人[112]通过控制客户端到汇聚节点的通信链路，对 Tor 隐藏服务器的入口节点进行攻击，以实现隐藏服务定位.

3.3　隐藏服务定位技术分析

匿名通信系统隐藏服务定位技术从用户位置的角度可分为客户端定位和服务端定位，本节对具有代表性的隐藏服务定位技术进行汇总、对不同类别隐藏服务定位技术的特点、优势与不足进行分析，如表 3 所示.

表 3　隐藏服务定位技术分析

用户位置	定位技术	优势	不足	主动/被动
客户端	网络流水印	实时性强、定位精度高、误报率低、易部署	需控制节点多、隐蔽性差、易被识别	主动
	交叉攻击	定位精度高、隐蔽性强、误报率低、易部署	观测时间长、实时性差、不适用于 Tor、I2P	被动
	其它(TorWard 为例)	隐蔽性强、定位精度高、实时性强	需控制节点多、部署难度大	主动
服务端	网站指纹攻击	隐蔽性强、需控制节点少、易部署	实时性差、误报率高	被动

在隐藏服务的客户端定位方面主要包括网络流水印技术和交叉攻击等方法. 网络流水印技术属于匿名通信系统主动攻击技术，通过控制部分网络节点(一般为入口节点和出口节点)，在匿名通信流量中利用数据包特征、包间隔、流速等载体嵌入并检测水印信息，不仅可以实现客户端定位，也可实现节点间通信关联和服务端定位，网络流水印技术具有实时性强、定位精度高、误报率低等优点，但该技术需要同时进行水印信息的嵌入和检测，较其它方法需控制节点多，并且水印技术需要对原始流量进行调整，因此隐蔽性差、易被识别，数据包重组、乱序、填充等处理可在一定程度上抵御此类攻击；交叉攻击通过不断缩小所观测的可疑节点集合的方法对隐藏服务客户端进行定位，但交叉攻击需要被观测对象之间持续进行通信. Tor、I2P 为抵御交叉攻击将一次会话时间设置为 10 min，因此交叉攻击方法在此类短会话系统中的效果不佳；除网络流水印技术和交叉攻击方法外，还有一些高效的客户端定位方案，表 3 中以 TorWard 为例，与其它定位方法进行了对比，TorWard 系统已在 3.1.3 节介绍，所以此处不再赘述.

在隐藏服务的服务端定位方面以网站指纹攻击的研究为主，网站指纹攻击通过被动采集流量，对流量特征进行提取，形成指纹特征库，进而利用机器学习、深度学习等算法对捕获的目标流量进行比对分析，实现服务端定位. 3.2.1 节详细介绍并分析、对比了典型的网站指纹攻击方案，本节以网站指纹攻击技术作为整体与客户端定位技术进行综合对比. 从网站指纹攻击过程可以看出，该方法一般部署在客户端和匿名通信系统入口节点之间，需控制节点少，且该方法被动获取网络流量，不需对原流量进行调整，具有较高的隐蔽性，但对于采集到的流量需要进行预处理、特征比对等计算，因此实时性不高，且现有算法对数据依赖性强，因此降低网站指纹攻击的误报率依然是该方向研究的热点问题.

4　未来研究方向展望

匿名通信系统及其隐藏服务的相关研究一直是网络安全领域国内外学术界关注的热点问题，而隐藏服务定位技术的研究有助于网络监管和网络空间治理，如何设计并实现行之有效的匿名通信系统隐藏服务用户定位方案是现阶段的难点问题. 目前此领域陆续涌现出了一些有价值的技术方法，取得了一定的效果，但依然有许多关键问题值得更深层次的探索. 结合实际，本文总结了以下几点未来可能的研究方向：

(1) 隐蔽网络流水印技术的研究. 网络流水印技术在匿名通信系统隐藏服务定位方面已取得了一定的成效，但如何在保证定位准确性的同时提高水印信息的隐蔽性是目前该技术面临的一个难点问题. Iacovazzi 等人[113]在 2019 年 RAID 会议上提出的 DUSTER 攻击方案是对隐蔽网络流水印技术的一次有效尝试，该方案利用 Tor 网络拥塞控制机制的脆弱性嵌入水印信息，并创新性地提出了“Detection and Cancelling”方法，水印检测器在检测到水印信息之后将其删除，使得接收方察觉不到水印信息的存在.

(2) 开放世界网站指纹识别精度的优化问题研究. 现有多数网站指纹攻击的工作致力于追求最大化系统召回率(maximize recall)，此类攻击方案在封闭世界表现优异，但是难以适应大规模的开放世

界，Wang 等人[114]通过对已知网站指纹攻击方案的精确度(precision)指标研究，提出在开放世界网站指纹攻击的研究中，精确度指标要比召回率更重要，而目前相关研究的最高精度仅为 78%. 由于开放世界网站指纹攻击的研究对于匿名通信系统隐藏服务定位具有更现实的意义，也更具挑战，开放世界网站指纹识别精度的优化问题研究将会是一个具有更实际意义的研究方向.

(3)“少样本”网站指纹攻击技术研究. 现有的网站指纹攻击模型需要依赖的训练集数据量大，并且需要定期持续地更新数据，大多数相关研究都假设测试和训练数据具有相似的分布，几乎在同一时间从同一类型的网络中收集，然而匿名通信系统的隐藏服务发布方式隐蔽，且迁移频繁，大规模数据采集和维护工作是本身就是难点问题，所以研究利用少量样本训练网络指纹攻击模型更具现实意义. Sirinam 等人[115]在 2019 年的 CCS 会议上提出的 TF(Triplet Fingerprinting)方案，利用了当下热门技术 N-Shot 学习，通过训练少量样本即可达到目标分类的目的，且准确率(accuracy)在 85%以上，该方法为“少样本”网站指纹攻击技术的研究提供了新的思路.

5 总 结

匿名通信系统隐藏服务的研究随着频频曝光的暗网安全事件得到了国内外学者们的广泛关注，从网络监管和执法取证的角度，匿名通信系统隐藏服务定位技术的研究十分必要且紧迫. 本文首先归纳了现有匿名通信系统的基本属性，从代理方式、通信性能和网络结构 3 个方面对国内外具有代表性的匿名通信系统进行了分类，研究了匿名通信系统及其隐藏服务机制的工作原理，重点根据隐藏服务的用户位置对隐藏服务定位技术进行了分类研究，总结了现有隐藏服务定位技术的发展现状和优缺点，在隐藏服务的客户端定位方面，针对匿名通信系统的网络流水印技术和交叉攻击等文献分别进行了研究和总结，在隐藏服务的服务端定位方面，针对隐藏服务网站指纹攻击方法、暗网数据采集与分析技术等文献分别进行了研究和总结.

总体而言，目前对匿名通信系统隐藏服务定位的研究是多角度、多技术途径的，本文从隐藏服务用户位置的角度对隐藏服务定位技术进行了分类研究，定位隐藏服务的客户端有助于执法部门和网络监管部门对使用隐藏服务的用户进行追踪取证，而定位隐藏服务的服务端有助于直接从源头阻断隐藏服务内容，净化网络空间. 然而，现有匿名通信系统隐藏服务定位技术虽然取得了一定的进展，但匿名通信系统自身匿名性增强技术也在不断发展进化，新的匿名通信系统也在不断涌现，随着互联网规模的不断扩大，匿名通信系统隐藏服务用户量也势必会随之增大，互联网环境也会愈加复杂，如何设计并实现更高效的、轻量级的匿名通信系统隐藏服务定位方案依然会是网络安全领域研究的热点和难点.

参 考 文 献

[1] Chaum D L. Untraceable electronic mail, return addresses, and digital pseudonyms. Communications of the ACM, 1981, 24(2): 84-90

[2] Dingledine R, Mathewson N, Syverson P. Tor: The second-generation Onion router. Journal of the Franklin Institute, 2004, 239(2): 135-139

[3] Luo Jun-Zhou, Yang Ming, Ling Zhen, et al. Anonymous communication and darknet: A survey. Journal of Computer Research and Development, 2019, 56(1): 103-130(in Chinese)
(罗军舟，杨明，凌振等. 匿名通信与暗网研究综述. 计算机研究与发展，2019，56(1)：103-130)

[4] Zhuo Zhong-Liu. Research on the Key Technologies of Network Tracing in the Anonymous Network[Ph. D. dissertation]. University of Electronic Science and Technology, Chengdu, 2018(in Chinese)
(卓中流. 匿名网络追踪溯源关键技术研究[博士学位论文]. 电子科技大学，成都，2018)

[5] Lu Tian-Bo, Du Ze-Yu, Wang Z-Jane. A survey on measuring anonymity in anonymous communication systems. IEEE Access, 2019, 7: 70584-70609

[6] Kelly D, Raines R, Baldwin R, et al. Exploring extant and emerging issues in anonymous networks: A taxonomy and survey of protocols and metrics. IEEE Communications Surveys & Tutorials, 2011, 14(2): 579-606

[7] Erdin E, Zachor C, Gunes M H. How to find hidden users: A survey of attacks on anonymity networks. IEEE Communications Surveys & Tutorials, 2015, 17(4): 2296-2316

[8] Nepal S, Dahal S, Shin S. Deanonymizing schemes of hidden services in Tor network: A survey//Proceedings of the 2015 International Conference on Information Networking (ICOIN). Siem Reap, Cambodia, 2015: 468-473

[9] Shirazi F, Simeonovski M, Asghar M R, et al. A survey on routing in anonymous communication protocols. ACM Computing Surveys, 2018, 51(3): 1-39

[10] Pfitzmann A, Waidner M. Networks without user observability. Computers & Security, 1987, 6(2): 158-166

[11] Boyan J. The anonymizer-protecting user privacy on the Web. Computer-Mediated Communication Magazine, 1997, 4(9): 1-6

[12] Freedman M J, Morris R. Tarzan: A peer-to-peer anonymizing network layer//Proceedings of the 9th ACM Conference on Computer and Communications Security. Washington, USA, 2002: 193-206

[13] Clarke I, Sandberg O, Wiley B, et al. Freenet: A distributed anonymous information storage and retrieval system//Proceedings of the Designing Privacy Enhancing Technologies. Berlin, Heidelberg, Germany, 2001: 46-66

[14] Chen C, Asoni D E, Barrera D, et al. HORNET: High-speed Onion routing at the network layer//Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. Denver, USA, 2015: 1441-1454

[15] Kelly D J. A taxonomy for and analysis of anonymous communications networks [Ph. D. dissertation]. Air Force Institute of Technology, USA, 2009

[16] Danezis G, Dingledine R, Mathewson N. Mixminion: Design of a type III anonymous remailer protocol//Proceedings of the 2003 Symposium on Security and Privacy. Berkeley, USA, 2003: 2-15

[17] Chaum D. The dining cryptographers problem: Unconditional sender and recipient untraceability. Journal of Cryptology, 1988, 1(1): 65-75

[18] Reiter M K, Rubin A D. Crowds: Anonymity for Web transactions. ACM Transactions on Information and System Security, 1998, 1(1): 66-92

[19] Moghaddam H M, Mosenia A. Anonymizing masses: Practical light-weight anonymity at the network level. arXiv preprint arXiv:1911.09642, 2019

[20] Piotrowska A M, Hayes J, Elahi T, et al. The Loopix anonymity system//Proceedings of the 26th USENIX Security Symposium. Vancouver, Canada, 2017: 1199-1216

[21] Chaum D, Das D, Javani F, et al. cMix: Mixing with minimal real-time asymmetric cryptographic operations//Proceedings of the International Conference on Applied Cryptography and Network Security. Kanazawa, Japan, 2017: 557-578

[22] Chen C, Asoni D E, Perrig A, et al. TARANET: Traffic-analysis resistant anonymity at the network layer//Proceedings of the 2018 IEEE European Symposium on Security and Privacy (EuroS&P). London, UK, 2018: 137-152

[23] Goel S, Robson M, Polte M, et al. Herbivore: A scalable and efficient protocol for anonymous communication. Cornell University, USA: Technical Report: TR2003-1890, 2003

[24] Lu Tian-Bo. WonGoo: A peer-to-peer protocol for anonymous communication//Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications. Las Vegas, USA, 2004: 1102-1106

[25] Lu Tian-Bo. Research on WonGoo—A Peer-to-Peer Anonymous Communication Protocol [Ph. D. dissertation]. Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 2006(in Chinese)
(陆天波. P2P 匿名通信协议 WonGoo 研究[博士学位论文]. 中国科学院计算技术研究所, 北京, 2006)

[26] Tan Qing-Feng, Fang Bin-Xing, Shi Jin-Qiao, et al. StegoP2P: A Hidden Communication Approach in P2P Networks. Journal of Computer Research and Development, 2014, 51(8): 1695-1703(in Chinese)
(谭庆丰, 方滨兴, 时金桥等. StegoP2P: 一种基于 P2P 网络的隐蔽通信方法. 计算机研究与发展, 2014, 51(8): 1695-1703)

[27] Dixon L, Ristenpart T, Shrimpton T. Network traffic obfuscation and automated internet censorship. IEEE Security & Privacy, 2016, 14(6): 43-53

[28] Tan Qing-Feng, Shi Jin-Qiao, Fang Bin-Xing, et al. Towards measuring unobservability in anonymous communication systems. Journal of Computer Research and Development, 2015, 52(10): 2373-2381(in Chinese)
(谭庆丰, 时金桥, 方滨兴等. 匿名通信系统不可观测性度量方法. 计算机研究与发展, 2015, 52(10): 2373-2381)

[29] Yao Zhong-Jiang, Ge Jing-Guo, Zhang Xiao-Dan, et al. Research review on traffic obfuscation and its corresponding identification and tracking technologies. Journal of Software, 2018, 29 (10): 3205-3222(in Chinese)
(姚忠将, 葛敬国, 张潇丹等. 流量混淆技术及相应识别, 追踪技术研究综述. 软件学报, 2018, 29(10): 3205-3222

[30] Wiley B. Dust: A blocking-resistant internet transport protocol. Technical rep ort. http://blanu.net/Dust.pdf, 2011

[31] Winter P, Pulls T, Fuss J. ScrambleSuit: A polymorph network protocol to circumvent censorship. arXiv preprint arXiv:1305.3199, 2013

[32] Mohajeri Moghaddam H, Li B, Derakhshani M, et al. Skypemorph: Protocol obfuscation for Tor bridges//Proceedings of the 2012 ACM Conference on Computer and Communications Security. Raleigh, USA, 2012: 97-108

[33] Dyer K P, Coull S E, Ristenpart T, et al. Protocol misidentification made easy with format-transforming encryption//Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security. Berlin, Germany, 2013: 61-72

[34] Xu C, Chen S, Su J, et al. A survey on regular expression matching for deep packet inspection: Applications, algorithms, and hardware platforms. IEEE Communications Surveys & Tutorials, 2016, 18(4): 2991-3029

[35] Dyer K P, Coull S E, Shrimpton T. Marionette: A programmable network traffic obfuscation system//Proceedings of the 24th USENIX Security Symposium. Washington, USA, 2015: 367-382

[36] Weinberg Z, Wang J, Yegneswaran V, et al. StegoTorus: A camouflage proxy for the Tor anonymity system//Proceedings of the 2012 ACM Conference on Computer and Communications Security. Raleigh, USA, 2012: 109-120

[37] Fifield D, Lan C, Hynes R, et al. Blocking-resistant communication through domain fronting. Proceedings on Privacy Enhancing Technologies, 2015, 2015(2): 46-64

[38] Brubaker C, Houmansadr A, Shmatikov V. Cloudtransport: Using cloud storage for censorship-resistant networking//Proceedings of the International Symposium on Privacy Enhancing Technologies Symposium. Amsterdam, The Netherlands, 2014: 1-20

[39] Fifield D, Hardison N, Ellithorpe J, et al. Evading censorship with browser-based proxies//Proceedings of the International Symposium on Privacy Enhancing Technologies Symposium. Berlin, Heidelberg, Germany, 2012: 239-258

[40] Wustrow E, Wolchok S, Goldberg I, et al. Telex: Anticensorship in the Network Infrastructure//Proceedings of the USENIX Security Symposium. San Francisco, USA, 2011: 45

[41] Jakobsson M, Juels A, Rivest R L. Making mix nets robust for electronic voting by randomized partial checking//Proceedings of the USENIX Security Symposium. San Francisco, USA, 2002: 339-353

[42] Berthold O, Pfitzmann A, Standtke R. The disadvantages of free MIX routes and how to overcome them//Proceedings of the International Workshop on Design Issues in Anonymity and Unobservability. Heidelberg, Berlin, 2001: 30-45

[43] Danezis G, Serjantov A. Statistical disclosure or intersection attacks on anonymity systems//Proceedings of the International Workshop on Information Hiding. Berlin, Heidelberg, Germany, 2004: 293-308

[44] Goldschlag D M, Reed M G, Syverson P F. Hiding routing information//Proceedings of the International Workshop on Information Hiding. Berlin, Heidelberg, Germany, 1996: 137-150

[45] Stoica I, Morris R, Karger D, et al. Chord: A scalable peer-to-peer lookup service for internet applications. ACM SIGCOMM Computer Communication Review, 2001, 31(4): 149-160

[46] Rowstron A, Druschel P. Pastry: Scalable, decentralized object location, and routing for large-scale peer-to-peer systems//Proceedings of the IFIP/ACM International Conference on Distributed Systems Platforms and Open Distributed Processing. Berlin, Heidelberg, Germany, 2001: 329-350

[47] McLachlan J, Tran A, Hopper N, et al. Scalable Onion routing with Torsk//Proceedings of the 16th ACM Conference on Computer and Communications Security. Chicago, USA, 2009: 590-599

[48] Baumeister T, Dong Y, Tian G, et al. Using randomized routing to counter routing table insertion attack on Freenet//Proceedings of the 2013 IEEE Global Communications Conference (GLOBECOM). Atlanta, USA, 2013: 754-759

[49] Ling Z, Luo J, Yu W, et al. Protocol-level attacks against Tor. Computer Networks, 2013, 57(4): 869-886

[50] Van Schyndel R G, Tirkel A Z, Osborne C F. A digital watermark//Proceedings of the 1st International Conference on Image Processing. Austin, USA, 1994: 86-90

[51] Wang W, Zhang X, Shi W, et al. Network traffic monitoring, analysis and anomaly detection [Guest Editorial]. IEEE Network, 2011, 25(3): 6-7

[52] Biersack E, Callegari C, Matijasevic M. Data traffic monitoring and analysis: From measurement, classification, and anomaly detection to quality of experience. Lecture Notes in Computer Science, 2013, 5(23): 12561-12570

[53] Guo Xiao-Jun, Cheng Guang, Zhu Chen-Gang, et al. Progress in research on active network flow watermark. Journal on Communications, 2014, 35(7): 178-192(in Chinese)
(郭晓军, 程光, 朱琛刚等. 主动网络流水印技术研究进展. 通信学报, 2014, 35(7): 178-192)

[54] Zhang Lian-Cheng, Wang Yu, Kong Ya-Zhou, et al. Survey on security threats and countermeasures of network flow watermarking. Journals of Computer Research and Development, 2018, 55(8): 1785-1799(in Chinese)
(张连成, 王禹, 孔亚洲等. 网络流水印安全威胁及对策综述. 计算机研究与发展, 2018, 55(8): 1785-1799)

[55] Iacovazzi A, Elovici Y. Network flow watermarking: A survey. IEEE Communications Surveys & Tutorials, 2016, 19(1): 512-530

[56] Wang X, Reeves D S, Wu S F, et al. Sleepy watermark tracing: An active network-based intrusion response framework//Proceedings of the IFIP International Information Security Conference. Boston, USA, 2001: 369-384

[57] Ramsbrock D, Wang X, Jiang X. A first step towards live botmaster traceback//Proceedings of the International Workshop on Recent Advances in Intrusion Detection. Berlin, Heidelberg, Germany, 2008: 59-77

[58] Back A, Möller U, Stiglic A. Traffic analysis attacks and trade-offs in anonymity providing systems//Proceedings of the International Workshop on Information Hiding. Berlin, Heidelberg, Germany, 2001: 245-257

[59] Houmansadr A, Kiyavash N, Borisov N. Non-blind watermarking of network flows. IEEE/ACM Transactions on Networking, 2013, 22(4): 1232-1244

[60] Houmansadr A, Kiyavash N, Borisov N. RAINBOW: A robust and invisible non-blind watermark for network flows//Proceedings of the NDSS. San Diego, USA, 2009

[61] Wang X, Chen S, Jajodia S. Network flow watermarking attack on low-latency anonymous communication systems//Proceedings of the 2007 IEEE Symposium on Security and Privacy (SP'07). Oakland, California, USA, 2007: 116-130

[62] Zhang Lu, Luo Jun-Zhou, Yang Ming, et al. Interval centroid based flow watermarking technique for anonymous communication traceback. Journal of Software, 2011, 22(10): 2358-2371(in Chinese)
(张璐, 罗军舟, 杨明等. 基于时隙质心流水印的匿名通信追踪技术. 软件学报, 2011, 22(10): 2358-2371)

[63] Yu W, Fu X, Graham S, et al. DSSS-based flow marking technique for invisible traceback//Proceedings of the 2007 IEEE Symposium on Security and Privacy (SP'07). Oakland, USA, 2007: 18-32

[64] Pyun Y J, Park Y H, Wang X, et al. Tracing traffic through intermediate hosts that repacketize flows//Proceedings of the IEEE INFOCOM 2007-26th IEEE International Conference on Computer Communications. Washington, USA, 2007: 634-642

[65] Kiyavash N, Houmansadr A, Borisov N. Multi-flow attacks against network flow watermarking schemes//Proceedings of the USENIX Security Symposium. San Jose, USA, 2008: 307-320

[66] Chan-Tin E, Shin J, Yu J. Revisiting circuit clogging attacks on Tor//Proceedings of the 2013 International Conference on Availability, Reliability and Security. Regensburg, Germany, 2013: 131-140

[67] Ling Z, Luo J, Xu D, et al. Novel and practical SDN-based traceback technique for malicious traffic over anonymous networks//Proceedings of the IEEE INFOCOM 2019-IEEE Conference on Computer Communications. Paris, France, 2019: 1180-1188

[68] Nunes B A A, Mendonca M, Nguyen X N, et al. A survey of software-defined networking: Past, present, and future of programmable networks. IEEE Communications Surveys & Tutorials, 2014, 16(3): 1617-1634

[69] Berthold O, Federrath H, Köhntopp M. Project "anonymity and unobservability in the Internet"//Proceedings of the 10th Conference on Computers Freedom and Privacy. Toronto, Canada, 2000: 57-65

[70] Wright M K, Adler M, Levine B N, et al. The predecessor attack: An analysis of a threat to anonymous communications systems. ACM Transactions on Information and System Security, 2004, 7(4): 489-522

[71] Wright M, Adler M, Levine B N, et al. Defending anonymous communications against passive logging attacks//Proceedings of the 2003 Symposium on Security and Privacy. Berkeley, USA, 2003: 28-41

[72] Ling Z, Luo J, Wu K, et al. TorWard: Discovery of malicious traffic over Tor//Proceedings of the IEEE INFOCOM 2014-IEEE Conference on Computer Communications. Toronto, Canada, 2014: 1402-1410

[73] Ling Z, Luo J, Wu K, et al. TorWard: Discovery, blocking, and traceback of malicious traffic over Tor. IEEE Transactions on Information Forensics and Security, 2015, 10(12): 2515-2530

[74] Zhuo Z, Zhang X, Li R, et al. A multi-granularity heuristic-combining approach for censorship circumvention activity identification. Security and Communication Networks, 2016, 9(16): 3178-3189

[75] Levine B N, Liberatore M, Lynn B, et al. Statistical detection of downloaders in Freenet//Proceedings of the 3rd IEEE International Workshop on Privacy Engineering. San Jose, USA, 2017: 25-32

[76] Hintz A. Fingerprinting websites using traffic analysis//Proceedings of the International Workshop on Privacy Enhancing Technologies. Berlin, Germany, 2002: 171-178

[77] Liberatore M, Levine B N. Inferring the source of encrypted HTTP connections//Proceedings of the 13th ACM Conference on Computer and Communications Security. New York, USA, 2006: 255-263

[78] Rish I. An empirical study of the naive Bayes classifier//Proceedings of the IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence. Seattle, USA, 2001: 41-46

[79] Zhuo Z, Zhang Y, Zhang Z, et al. Website fingerprinting attack on anonymity networks based on profile hidden Markov model. IEEE Transactions on Information Forensics and Security, 2018, 13(5): 1081-1095

[80] Eddy S R. Profile hidden Markov models. Bioinformatics (Oxford, England), 1998, 14(9): 755-763

[81] Zeng X, Chen X, Shao G, et al. Flow context and host behavior based Shadowsocks's traffic identification. IEEE Access, 2019, 7: 41017-41032

[82] Gu Xiao-Dan, Yang Ming, Luo Jun-Zhou, et al. Website fingerprinting attack based oil hyperlink relations. Chinese Journal of Computers, 2015, 38(4): 833-845(in Chinese)
(顾晓丹，杨明，罗军舟等. 针对 SSH 匿名流量的网站指纹攻击方法. 计算机学报，2015，38(4)：833-845)

[83] Panchenko A, Niessen L, Zinnen A, et al. Website fingerprinting in Onion routing based anonymization networks//Proceedings of the 10th Annual ACM Workshop on Privacy in the Electronic Society. Chicago, USA, 2011: 103-114

[84] Suykens J A K, Vandewalle J. Least squares support vector machine classifiers. Neural Processing Letters, 1999, 9(3): 293-300

[85] Berthold O, Federrath H, Köpsell S. Web MIXes: A system for anonymous and unobservable Internet access//Proceedings of the Designing Privacy Enhancing Technologies. Berlin, Heidelberg, Germany, 2001: 115-129

[86] He Gao-Feng, Yang Ming, Luo Jun-Zhou, et al. Online identification of Tor anonymous communication traffic. Journal of Software, 2013, 24(3): 540-556(in Chinese)
(何高峰，杨明，罗军舟，等. Tor 匿名通信流量在线识别方法. 软件学报，2013，24(3)：540-556)

[87] He G, Yang M, Gu X, et al. A novel active website fingerprinting attack against Tor anonymous system//Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design (CSCWD). Hsinchu, China, 2014: 112-117

[88] Cai X, Zhang X C, Joshi B, et al. Touching from a distance: Website fingerprinting attacks and defenses//Proceedings of the 2012 ACM Conference on Computer and Communications Security. Raleigh, USA, 2012: 605-616

[89] Wang T, Goldberg I. Improved website fingerprinting on Tor//Proceedings of the 12th ACM Workshop on Privacy in the Electronic Society. Berlin, Germany, 2013: 201-212

[90] Cover T, Hart P. Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 1967, 13(1): 21-27

[91] Hayes J, Danezis G. k-fingerprinting: A robust scalable website fingerprinting technique//Proceedings of the 25th USENIX Security Symposium. Vancouver, Canada, 2016: 1187-1203

[92] Ho T K. Random decision forests//Proceedings of the 3rd International Conference on Document Analysis and Recognition. Montreal, Canada, 1995: 278-282

[93] Kwon A, AlSabah M, Lazar D, et al. Circuit fingerprinting attacks: Passive deanonymization of Tor hidden services//Proceedings of the 24th USENIX Security Symposium. Washington, USA, 2015: 287-302

[94] Rimmer V, Preuveneers D, Juarez M, et al. Automated website fingerprinting through deep learning. arXiv preprint arXiv:1708.06376, 2017

[95] Sirinam P, Imani M, Juarez M, et al. Deep fingerprinting: Undermining website fingerprinting defenses with deep learning//Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. Toronto, Canada, 2018: 1928-1943

[96] Juarez M, Imani M, Perry M, et al. Toward an efficient website fingerprinting defense//Proceedings of the European Symposium on Research in Computer Security. Heraklion, Greece, 2016: 27-46

[97] Wang T, Goldberg I. Walkie-talkie: An efficient defense against passive website fingerprinting attacks//Proceedings of the 26th USENIX Security Symposium. Vancouver, Canada, 2017: 1375-1390

[98] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature, 2015, 521(7553): 436

[99] Bhat S, Lu D, Kwon A, et al. Var-CNN: A data-efficient website fingerprinting attack based on deep learning. Proceedings on Privacy Enhancing Technologies, 2019, 2019(4): 292-310

[100] Kobayashi M, Takeda K. Information retrieval on the Web. ACM Computing Surveys, 2000, 32(2): 144-173

[101] Christin N. Traveling the silk road: A measurement analysis of a large anonymous online marketplace//Proceedings of the 22nd International Conference on World Wide Web. Rio de Janeiro, Brazil, 2013: 213-224

[102] Yin Shuai. Design and Implements of Dark Web Scanning System Based on Injection [M. S. dissertation]. Beijing University of Posts and Telecommunications, Beijing, 2018 (in Chinese)
(殷帅. 基于节点注入的暗网扫描系统的设计与实现[硕士学位论文]. 北京邮电大学, 北京, 2018)

[103] Bernaschi M, Celestini A, Guarino S, et al. Spiders like Onions: On the network of Tor hidden services//Proceedings of the World Wide Web Conference. San Francisco, USA, 2019: 105-115

[104] He Si-Yu. Classification of Didden Service Content[M. S. dissertation]. Beijing Jiaotong University, Beijing, 2019(in Chinese)
(何思雨. 隐藏服务内容分类研究[硕士学位论文]. 北京交通大学, 北京, 2019)

[105] Park J, Lee Y. POSTER: Probing Tor hidden service with dockers//Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. New York, USA, 2017: 2571-2573

[106] Merkel D. Docker: Lightweight linux containers for consistent development and deployment. Linux Journal, 2014, 2014(239): 2

[107] Ling Z, Luo J, Wu K, et al. Protocol-level hidden server discovery//Proceedings of the 2013 Proceedings IEEE INFOCOM. Turin, Italy, 2013: 1043-1051

[108] Iacovazzi A, Sarda S, Elovici Y. Inflow: Inverse network flow watermarking for detecting hidden servers//Proceedings of the IEEE INFOCOM 2018-IEEE Conference on Computer Communications. Honolulu, USA, 2018: 747-755

[109] Elices J A, Pérez-González F. Locating Tor hidden services through an interval-based traffic-correlation attack//Proceedings of the 2013 IEEE Conference on Communications and Network Security (CNS). Washington, USA, 2013: 385-386

[110] Tan Q, Gao Y, Shi J, et al. Toward a comprehensive insight into the eclipse attacks of Tor hidden services. IEEE Internet of Things Journal, 2018, 6(2): 1584-1593

[111] Matic S, Kotzias P, Caballero J. Caronte: Detecting location leaks for deanonymizing Tor hidden services//Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. Denver, USA, 2015: 1455-1466

[112] Biryukov A, Pustogarov I, Weinmann R P. Trawling for Tor hidden services: Detection, measurement, deanonymization//Proceedings of the 2013 IEEE Symposium on Security and Privacy. San Francisco, USA, 2013: 80-94

[113] Iacovazzi A, Frassinelli D, Elovici Y. The {DUSTER} Attack: Tor Onion Service Attribution Based on Flow Watermarking with Track Hiding//Proceedings of the 22nd International Symposium on Research in Attacks, Intrusions and Defenses. Beijing, China, 2019: 213-225

[114] Wang T. Optimizing precision for open-world website fingerprinting. arXiv preprint arXiv:1802.05409, 2018

[115] Sirinam P, Mathews N, Rahman M S, et al. Triplet fingerprinting: More practical and portable website fingerprinting with N-shot learning//Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security. London, United Kingdom, 2019: 1131-1148

ZHAO Na, Ph. D. candidate. Her current research interests include cyberspace security and anonymous communication systems.

SU Jin-Shu, Ph. D., professor, Ph. D. supervisor. His current research interests include computer networks and cyberspace security.

ZHAO Bao-Kang, Ph. D., associate professor. His current research interests include computer networks and cyberspace security.

HAN Biao, Ph. D., associate professor. His current research interests include cyberspace security and intellisense and network communication.

ZOU Hong-Cheng, Ph. D. candidate. His current research interests include cyberspace security and anonymous communication systems.

Background

Anonymous communication systems and location issues are hot topics for researchers. With the rapid development of anonymous communication systems, especially with the support of hidden service technologies, anonymous communication systems are being increasingly abused by criminals, the dark Web platforms have already become a place out of laws with the technical assistance of anonymous communication systems' hidden services which can strongly protect confidentiality of senders and receivers during communication.. From the standpoint of network supervision department, it is necessary and urgent to study anonymous communication systems and the location technologies of anonymous communication systems' hidden services.

Scholars from domestic and foreign have carried out a lot of research on anonymous communication systems and location technologies. At present, the related works mainly introduce from the aspects of anonymous routing protocols, dark network relationship and attack modes. The location technologies of anonymous communication systems have made some effective achievements, but the anonymity is getting stronger and stronger along with the development of location technologies at the same time. Besides, new anonymous communication systems are emerging, hidden service nodes are more hidden and migrate frequently, it is no doubt that the research on hidden service location technologies of anonymous communication systems is a severe and tough challenge to researchers. This paper focus on the location technologies of anonymous communication systems' hidden services. We make an in-depth study of the existing literatures about anonymous communication systems and hidden service location technologies and divide the existing location technologies into client-side location and sever-side location. In this paper we mainly introduce the typical hidden services' location technologies, such as network flow watermark technologies, website fingerprinting attack methods and so on in detail. The development status and the advantages and disadvantages of existing representative hidden services' location technologies are summarized in the end of this paper.

This work is supported in part by the National Natural Science Foundation of China (No. 61972412, No. 61601483), the Excellent Youth Scientific Research Project of Education Department of Hunan Province (No. 16B023), the General Scientific Research Project of Education Department of Hunan Province (No. 19C0140), and the Training Program for Excellent Young Innovators of Changsha (No. kq1905006).