DOI:10.7544/issn1000-1239.2019.20180769
56(1):103-130,2019

匿名通信与暗网研究综述

罗军舟　杨　明　凌　振　吴文甲　顾晓丹
(东南大学计算机科学与工程学院　南京　211189)
(jluo@seu.edu.cn)

Anonymous Communication and Darknet: A Survey

Luo Junzhou, Yang Ming, Ling Zhen, Wu Wenjia, and Gu Xiaodan
(*School of Computer Science and Engineering, Southeast University, Nanjing* 211189)

Abstract　An anonymous communication system is an overlay network built on top of the Internet that integratedly uses various anonymous technologies such as data forwarding, content encryption and traffic obfuscation to conceal the relationships between the communication entities and hide their communication content from being revealed. Since it is quite difficult to traceback and locate the communication entities, considerable abuse problems have sprung up. Specifically, the hidden services of anonymous network are most often abusively used to establish darknets where diverse illegal activities take place, bringing huge harm to the individuals and society. At present, it lacks comprehensive and deep technical analysis and research survey in the field of anonymous communication system and darknet study. Based on the status quo, this paper first elaborates on the basic concepts of the two terms and their relationships, then continues to demonstrate the work mechanisms and three key technologies of the anonymous communication system in detail including anonymous access, anonymous routing and darknet services with an exemplification of the four mainstream darknets, which are Tor, I2P, Freenet and ZeroNet. On this basis, the paper summarizes the state-of-the-art anonymous communication attacks and defense technology, and introduces the research work on current darknet governance as well. Finally, the developing trend of the next generation anonymous communication system is prospected, and the challenges of the darknet governance and relevant countermeasures are also discussed.

Key words　anonymous communication; darknet; hidden services; attack and defense technologies; darknet governance

摘　要　匿名通信系统是一种建立在Internet之上综合利用数据转发、内容加密、流量混淆等多种技术隐藏通信实体关系和内容的覆盖网络.由于匿名通信主体难以被追踪定位,匿名通信网络中各类匿名滥用问题层出不穷,而其中隐藏服务机制更是被用于构建充斥着各种非法活动的暗网,给社会和人们带来了巨大的危害和影响.鉴于目前在匿名通信与暗网领域尚缺乏全面、深入的技术剖析与研究综述,重点对两者的基本概念及相互关系进行阐述,并围绕Tor,I2P,Freenet,ZeroNet这4种主流暗网形态,详细介绍了匿名通信的工作原理,以及匿名接入、匿名路由和暗网服务三大关键技术.在此基础上,分析总结

收稿日期:2018-11-16;**修回日期**:2018-11-28
基金项目:国家重点研发计划项目(2017YFB1003000,2018YFB0803400);国家自然科学基金项目(61572130,61502100,61532013,61632008);江苏省科技成果转化基金项目(BA2016052)
This work was supported by the National Key Research and Development Program of China (2017YFB1003000, 2018YFB0803400), the National Natural Science Foundation of China (61572130, 61502100, 61532013, 61632008), and the Jiangsu Provincial Scientific and Technological Achievements Transfer Fund (BA2016052).
通信作者:杨明(yangming2002@seu.edu.cn)

了匿名通信攻击与防御技术的研究现状,并介绍了暗网治理方面的相关研究工作.最后,对下一代匿名通信系统的发展趋势进行展望,并初步探讨了暗网治理面临的挑战和思路.

关键词 匿名通信;暗网;隐藏服务;攻防技术;暗网治理

中图法分类号 TP391

匿名通信是一种通过采用数据转发、内容加密、流量混淆等措施来隐藏通信内容及关系的隐私保护技术.为了提高通信的匿名性,这些数据转发链路通常由多跳加密代理服务节点构成,而所有这些节点即构成了匿名通信系统(或称匿名通信网络).匿名通信系统本质上是一种提供匿名通信服务的覆盖网络,可以向普通用户提供 Internet 匿名访问功能以掩盖其网络通信源和目标,向服务提供商提供隐藏服务机制以实现匿名化的网络服务部署.作为匿名通信系统的核心功能,隐藏服务机制通常利用多跳反向代理或通过资源共享存储来掩盖服务提供商的真实地址,可以保证匿名服务不可追踪和定位.

由于匿名通信系统具有节点发现难、服务定位难、用户监控难、通信关系确认难等特点,利用匿名通信系统隐藏真实身份从事恶意甚至于网络犯罪活动的匿名滥用现象层出不穷.由于用户必须通过特殊软件或进行特殊配置才能访问服务,隐藏服务机制更是被用于部署"丝绸之路"(silk road)、AlphaBay等网络黑市,形成了潜藏于 Internet 中充斥着毒品枪支交易、人口贩卖、恐怖活动、反社会言论、谣言散布以及敲诈勒索等非法活动的暗网:臭名昭著的 Tor 黑市站点"丝绸之路"在其上线的 2 年半时间中总交易额高达 12 亿美金,并拥有 15 万匿名用户和 4 000 余非法商户;而勒索病毒 WannaCry,Skynet 等各类僵尸主机的命令控制(command-and-control, C&C)服务器直接部署于暗网中以防止被追踪溯源.此外,国内在暗网黑市中的各类违法事件也是频频曝光:2018 年 6 月 AcFun 视频站点的近千万条用户信息数据在 Tor 暗网中进行兜售;同年 9 月华住集团旗下酒店约 1.2 亿条客户个人信息在暗网中出售.除了暗网安全事件的不断攀升,暗网站点数量和用户量也呈现出逐年增长的趋势.根据 Tor Metrics 官方网站的统计信息[1],截至 2018 年 10 月 Tor 暗网中已有超过 11 万个站点,而用户数超过 200 万.与 Tor 相比,其他暗网系统,如 I2P,Freenet,ZeroNet 等,由于其网络拓扑特点,很难发现并统计其站点和用户规模.

在学术界,研究人员针对匿名通信系统进行了多年的研究工作,而暗网在近年来也逐渐受到关注.这些工作大致可分为 2 类:1)专注于匿名通信系统本身,分析其安全性并提出相应的隐私增强技术,包括匿名接入、匿名路由、隐藏服务等机制的优化,以及流量混淆与协议伪装等技术的设计和应用;2)侧重于针对匿名通信系统与暗网的攻击及监管,利用流量分析、系统/协议漏洞分析等手段,提出了一系列的去匿名化攻击方法,可用于对其进行有效监管,具体包括暗网隐藏节点发现和隐藏服务定位、暗网用户的网络行为分析,以及暗网流量追踪和通信关系确认等.

鉴于匿名通信与暗网对网络空间安全及治理的重要性,以及在研究领域尚缺乏完整、全面评估的现状,本文针对匿名通信与暗网的基本概念及相互关系、关键技术和攻防研究以及暗网治理方面的研究工作进行综述.本文首先介绍了典型匿名通信系统的基本工作原理,涵盖了 Tor,I2P,Freenet 和 ZeroNet 这 4 种主流暗网形态.其次,分别从匿名接入、匿名路由、暗网服务这 3 项关键技术的角度阐述了各系统的特点.在此基础上,分析了针对匿名通信的各类攻击与防御技术的研究现状,并介绍了暗网治理方面的相关研究工作.本文最后对下一代匿名通信系统的发展趋势进行了展望,并初步探讨了我国在暗网治理中面临的挑战和可采取的治理措施.

1 匿名通信系统

近年来,匿名通信系统与暗网引起了人们的广泛关注,但其概念及相互关系一直缺乏统一的定义.因此,本节给出匿名通信系统的基本分类,并阐明与暗网的关系.随后,具体介绍 Tor,I2P,Freenet,ZeroNet 这 4 种典型的匿名通信系统.

1.1 基本分类

根据通信延迟性能,可将匿名通信系统分为低延迟系统和高延迟系统.在低延迟系统中,又可根据网络结构进一步分成 P2P 匿名网络和非 P2P 匿名网络,而其中 P2P 匿名网络进一步包含结构化与非结构化 2 种网络模型.

高延迟匿名通信系统源于早期的 Mixnet 匿名通信网络,最初由 Chaum 于 1981 年提出[2]. 它的核心思路是基于 mix 节点对输入的消息进行加密和混淆,使得攻击者无法追踪消息的网络传播过程. 因此,mix 节点隐藏了各个消息之间的输入和输出关系,可以防止攻击者对输入和输出消息进行关联. Danezis[3]提出了结合 mix 级联模型和自由路由的受限路由 mix 网络,以保证消息的匿名性. Dingledine 等人[4]提出基于信誉系统的 mix 节点选择技术,从而为 mix 网络提供更高的可靠性和效率. Dingledine 等人[5]提出了一种具有分布式信任的 mix 级联协议,基于信誉度量重新排列 mix 级联,以提高可靠性.

Mixnet 协议的典型应用场景是高延迟邮件转发服务(Remailer),用于提供匿名电子邮件收发功能:Babel[6]引入 mix 路径,旨在通过延迟一些批量消息来抵御流量分析攻击. Mixmaster[7]通过在每个数据包的末尾添加随机数据,将消息加密转换为统一的大小. 为保证所有发送者的路由信息相同,Mixminion[8]部署了一组冗余和同步的目录服务器系统,旨在为电子邮件消息提供发件人和收件人匿名.

相对于高延迟匿名系统,低延迟系统因其应用领域宽泛,受到了研究人员更多的关注. 结构化 P2P 匿名网络主要包括 Tor,I2P 和 Freedom[9],它们均采用洋葱路由或类洋葱路由协议进行通信. 其中 Freedom 为网络中每个用户分配一个唯一身份标识以屏蔽 IP 地址信息,同时利用类似于洋葱路由的加密通信方式为用户提供匿名网页浏览等服务. 非结构化 P2P 匿名网络有 Torsk,Freenet,Crowds 等,主要采用基于 DHT(distributed Hash table)的路由协议和随机游走协议. Torsk[10]通过使用 DHT 取代 Tor 的节点选择和目录服务,以获得更好的扩展性. Crowds[11]是为匿名网页浏览而设计的匿名通信系统之一,其主要特征为随机节点选择. Crowds 中所有节点都被分群,群中每个节点可以连接任何其他节点作为通信中继.

在综合 mix 技术和洋葱路由技术的基础上,新型匿名通信技术相继被提出,包括 TARANET[12],Riffle[13]和 Loopix[14]等,可进一步提高匿名通信系统的匿名性、扩展性及性能.

1.2 与暗网的关系

匿名通信技术的提出源于网络用户的隐私保护需求,但是随之而来的是各种匿名滥用问题. 尤为突出的是,一些不法用户基于 Tor,I2P,Freenet 等匿名通信系统提供的隐藏服务机制,架设必须通过特殊软件或进行特殊配置才能访问的网络服务,形成了无法被直接检索的暗网. 由于可以有效隐藏网络服务提供者的 IP 地址等信息,暗网中充斥着毒品枪支交易、人口贩卖等非法活动.

在学术界,暗网的概念并没有一个明确的定义,术语“Dark Web”和“Darknet”经常被混用指代暗网. 为了更好地厘清这 2 个术语之间的关系和差异,并界定本文所讨论的暗网的内容,给出如下定义:

1) Dark Web. 必须通过特殊的软件、特殊的配置才能访问的拥有特殊域名的 Web 站点,搜索引擎无法对其进行检索.

2) Darknet. 必须通过特殊的软件、特殊的配置才能访问,包含 Web、IRC、文件共享等各类资源和服务的匿名网络,搜索引擎无法对其进行检索.

两者的关系如图 1 所示:其中 Dark Web 重点在于不可检索、不可直接访问,并不强调匿名,因此既包含匿名 Web 站点(如 Tor 隐藏服务站点),也包含非匿名 Web 站点(如 ZeroNet 提供的 Web 站点);而 Darknet 则强调匿名性,即所有的资源都需要基于相应的匿名通信系统才能进行访问,其中的资源不限于 Web 服务.

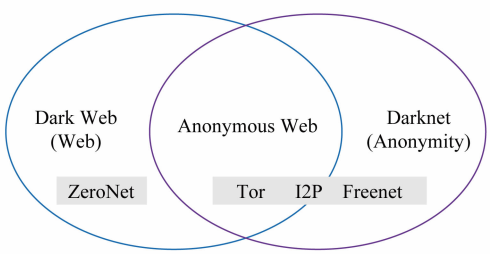

Fig. 1 The differences between Darknet and Dark Web

图 1 Darknet 和 Dark Web 的区别

由于 Tor,I2P,Freenet 等匿名通信系统不仅允许发布隐藏 Web 服务,还能提供其他类型资源的匿名访问,同时满足了 Dark Web 和 Darknet 两个定义,并且隐藏 Web 服务又是其中应用最为广泛的服务. 因此,很多时候 Darknet 被混淆为 Dark Web.

本文所讨论的暗网强调匿名性,并且不限于 Web 服务,因此采用 Darknet 术语. 鉴于 ZeroNet 虽然本身并不提供匿名性保护,但它通常和 Tor 打包发布,并利用 Tor 提供服务部署和资源获取的功能,因此也将其纳入讨论的范畴. 1.3 节将对 Tor,I2P,Freenet,ZeroNet 展开介绍.

1.3 典型匿名通信系统

1.3.1 Tor 匿名通信系统

Tor(the second-generation onion router)[15]是目前使用范围最广的匿名通信系统之一,其核心技术"洋葱路由"在 20 世纪 90 年代中期由美国海军研究实验室提出,并在 1997 年交由美国国防高级研究计划局(DARPA)进一步开发. 2003 年 Tor 正式版发布,2004 年 Tor 设计文档在第 13 届 USENIX 安全讨论会上正式发表,并在同年由美国海军研究实验室公开其源码. Tor 使用多跳代理机制对用户通信隐私进行保护:首先,客户端使用基于加权随机的路由选择算法分别选择 3 个中继节点,并逐跳与这些中继节点建立链路. 在数据传输过程中,客户端对数据进行 3 层加密,由各个中继节点依次进行解密. 由于中继节点和目的服务器无法同时获知客户端 IP 地址、目的服务器 IP 地址以及数据内容,从而保障了用户隐私.

Tor 于 2004 年开始支持隐藏服务,为 Tor 暗网的出现提供了技术支撑. Tor 暗网是目前规模最大的暗网之一,其中包含大量的敏感内容与恶意内容. Tor 隐藏服务是仅能在 Tor 暗网中通过特定形式的域名(〈z〉.onion)访问的网络服务. Tor 暗网的基本组件包括客户端、目录服务器(directory server)、隐藏服务目录服务器(hidden service directory)、洋葱路由器(onion router, OR)和隐藏服务器(hidden server),如图 2 所示:

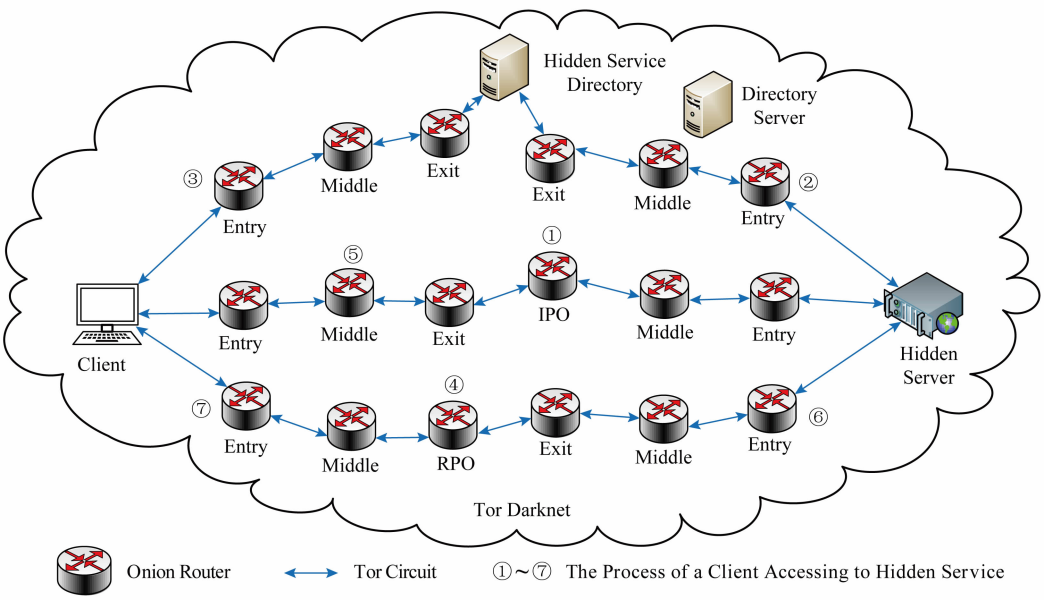

Fig. 2 Tor hidden service

图 2 Tor 隐藏服务

1) 客户端. 客户端是运行在用户操作系统上的本地程序,称为洋葱代理(onion proxy, OP). OP 将用户数据封装成 Tor 信元并层层加密,为各类 TCP 应用程序提供匿名代理服务.

2) 洋葱路由器. Tor 暗网中的数据中继节点. Tor 默认匿名链路由 3 个 OR 组成,分别为入口节点(Entry)、中间节点(Middle)和出口节点(Exit),其中入口节点一般选择可信度较高的守护节点(Guard).

3) 隐藏服务器. 提供 Web,IRC 等 TCP 应用服务. 隐藏服务器受到 Tor 匿名性的保护,必须使用 Tor 客户端才能够访问其 TCP 应用服务.

4) 目录服务器. 目录服务器保存了所有洋葱路由器的 IP 地址、带宽等信息. 客户端在首次启动后向目录服务器请求洋葱路由器信息,以便完成节点选择和链路建立.

5) 隐藏服务目录服务器. 隐藏服务目录服务器存储并为客户端提供隐藏服务器的引入节点(introduction point, IPO)、公钥等节点信息.

客户端、洋葱路由器、目录服务器、隐藏服务目录服务器和隐藏服务器的功能都集成在 Tor 软件包中,用户可以通过配置文件对具体功能进行配置.

Tor 隐藏服务器在启动时会选择 3 个引入节点作为其前置代理,并将引入节点及其公钥信息上传至隐藏服务目录服务器. 客户端访问隐藏服务时,首先建立 3 跳链路访问隐藏服务目录服务器,获取引入节点和公钥信息. 客户端选择一个汇聚节点(ren-

dezvous point, RPO)作为客户端和隐藏服务器通信链路的汇聚点,并将汇聚节点的信息通过引入节点告知隐藏服务器.客户端和隐藏服务器各自建立到达汇聚节点的链路,完成 6 跳链路的搭建后即可开始通信.Tor 用户通过 6 跳链路访问隐藏服务器,在此过程中任意节点无法同时获知 Tor 客户端 IP 地址、隐藏服务器 IP 地址以及数据内容,保障了 Tor 客户端与隐藏服务器的匿名性.

1.3.2 I2P 匿名网络

I2P(invisible Internet project)[16]是一种使用单向加密隧道的 P2P 匿名通信系统,采用称为大蒜路由的扩展洋葱路由技术,使得隧道中每一跳节点只掌握相邻节点的信息,无法获知通信双方的通信关系,从而保证通信的匿名性.该项目在 2003 年 2 月启动[17],2004 年 7 月发布首个 PC 端稳定版本 0.3.2.1,2014 年 8 月发布 I2P Android 客户端.用户可以配置 I2P 的隧道跳数、带宽和延迟等参数,以满足特定匿名性能需求.

I2P 中节点的类型分为 Floodfill 和 Nonfloodfill 两类.节点默认初始身份为 Nonfloodfill,其中满足性能要求的节点会自适应地成为 Floodfill 节点,其数量约占所有 I2P 节点的 6%[18].Floodfill 节点保存 RouterInfo 和 LeaseSet 两类数据信息:RouterInfo 包括节点的 ID、公钥、签名、通信协议及端口等内容;LeaseSet 包括服务 Hash 值、隧道入口节点和起止有效时间等信息.I2P 系统根据 Kademlia 算法[19]来组织所有的 Floodfill 节点,形成 I2P 的网络数据库(network database, netDb)以提供对所有 RouterInfo 和 LeaseSet 信息的保存、查询等功能.

I2P 拓扑结构如图 3 所示,每个节点根据 256 b 路由值(router key)分布在逻辑环中,其中路由值是采用 SHA256 算法对节点 ID 和当前日期计算获得.I2P 通过异或 2 个节点的路由值来度量节点间的距离,每个节点会选择异或距离最近的 Floodfill 节点上传 RouterInfo 和 LeaseSet 信息.因此,节点每天会由于日期的改变而呈现不同的逻辑分布,并将信息上传至不同的 Floodfill 节点,以此来抵御 Sybil 攻击[20].

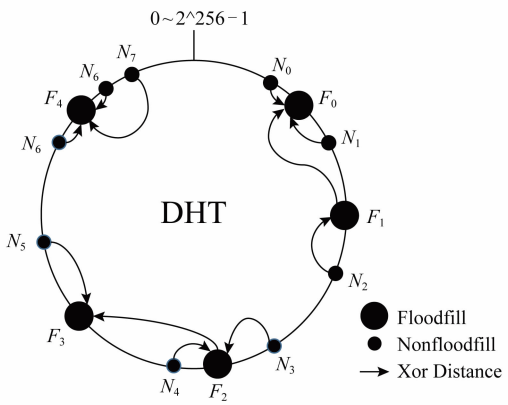

Fig. 3 The topology of I2P

图 3 I2P 拓扑结构图

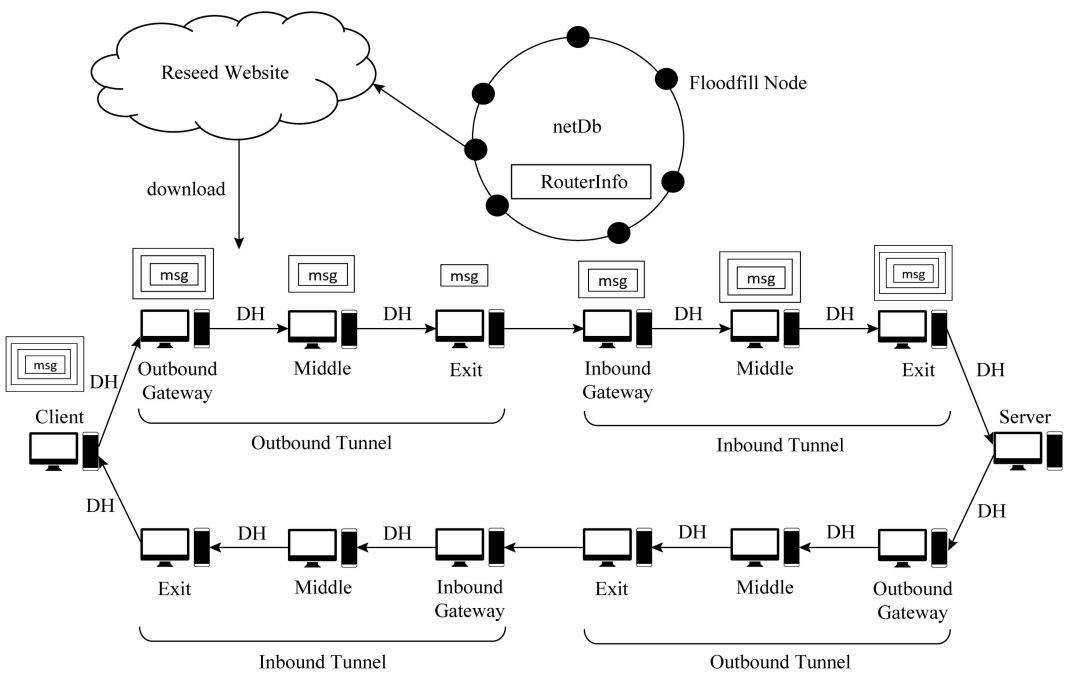

Fig. 4 Tunnels in I2P

图 4 I2P 隧道机制

I2P 的客户端和服务器均利用多跳单向加密隧道进行通信,以保护双方通信隐私,其过程如图 4 所示. 节点初次加入 I2P 网络时,从官方补种网站(reseed website)进行补种,获取部分 RouterInfo 信息,并根据路由选择机制建立单向隧道. 隧道根据用途可以分为探测隧道(exploratory tunnel)和客户隧道(client tunnel),根据数据传输方向可以分为输入隧道(inbound tunnel)和输出隧道(outbound tunnel). 其中,探测隧道用于辅助构建、测试客户隧道和查询 netDb 中信息等,客户隧道用于应用服务如 Web 浏览、聊天室、邮件和文件共享等. 客户端和服务器在通信过程中分别建立各自的输入和输出客户隧道,默认隧道长度为 3 跳,一次完整的通信过程需要 4 条隧道参与. 客户端发生的数据先采用端到端加密,然后在客户端进行 3 次洋葱式加密后发送到其输出客户隧道的 Gateway 节点,并在各个节点上分别进行解密转发到服务器的输入客户隧道 Gateway 节点,然后依次在各个节点上加密转发到服务器,最终由服务器通过 4 次解密得到明文数据,数据反向发送过程与此类似. 多跳隧道中的节点只知道其前驱和后继节点信息,从而隐藏通信双方的通信关系,同时单向隧道通过增加参与通信的节点数量提升了通信的匿名性.

1.3.3 Freenet 匿名网络

Freenet 是一个分布式的匿名信息存储与检索系统,在为用户提供文件上传、下载与检索功能的同时能够保障文件发布者与查阅者的匿名性[21-22]. 该系统源于爱丁堡大学的 Ian Clarke 在 1999 年的毕业设计,后由其发起了 Freenet 开源项目,并于 2000 年 3 月发布了首个版本. Freenet 在设计上主要追求 5 个目标:保护文件的发布者和查阅者的匿名性;本地存储的可否认性;能够抵抗第三方对信息可访问性的破坏;高效的分布式存储与路由;完全的去中心化.

所有 Freenet 节点分布在周长为 1 个单位的逻辑环上,这些节点按功能可以分为种子节点和非种子节点,默认新加入的节点均为非种子节点,种子节点具有辅助发现节点的功能. Freenet 中新节点随机产生一个介于[0,1)的实数,用于标识其在环上的位置. 每个节点会贡献一定大小的本地硬盘空间共同构成 Freenet 的存储空间.

Freenet 节点有 2 种工作模式:Opennet 模式和 Darknet 模式[23]. Opennet 模式下节点可以和任何其他节点建立连接. 与自身节点一跳直接相连的节点称为邻居节点,邻居节点的数量与该节点的带宽成正相关,最多支持 85 个邻居节点(版本号 0.7.5,内部版本 1483). 节点间的距离是由节点间不大于 0.5 的弧长定义. 在 Opennet 模式下,Freenet 定义距离不超过 0.01 的邻居节点为近邻居,且每个节点的远邻居节点数量不超过邻居节点数的 30%,其余为近邻居节点,如图 5 所示. 新节点在加入网络时通过种子节点获取其他节点的信息,然后选择邻居节点建立连接从而加入 Freenet 网络. 而在 Darknet 模式下,节点只能和用户添加的信任节点建立连接,以此保证安全性.

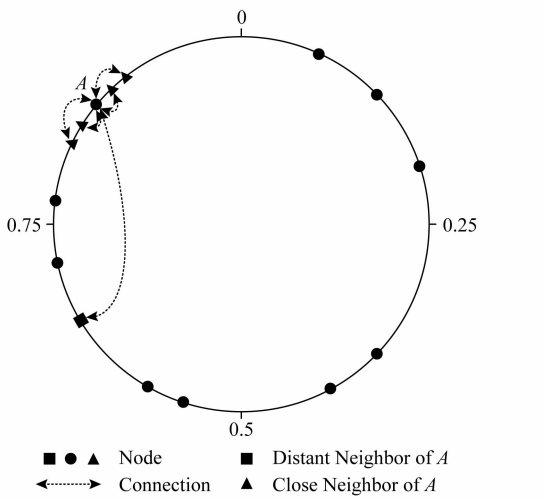

Fig. 5 The topology of Freenet

图 5 Freenet 拓扑结构

Freenet 节点通过 FCP(Freenet client protocol)协议向上层应用提供匿名文件存储与检索服务,并集成基于 FCP 的 Fproxy 服务,从而提供 HTTP 代理、Freenet 节点配置、文件上传与检索、Freesite 浏览等功能. 在文件上传前,Freenet 将文件按照 32 KB 大小切块,分别计算每块的唯一标识符. 文件标识符主要分为内容 Hash 键值(content Hash keys, CHK)和签名子空间键值(signed subspace keys, SSK),文件的索引值可根据该键值通过 SHA256 Hash 运算获得,索引值可转化为逻辑环上的对应位置. 在文件存储的过程中,节点根据文件的位置查找距离最近的邻居节点,由其进一步迭代查找目标存储节点,最终通过该多跳查找路径将文件上传. 文件检索的过程中,根据文件对应的位置利用深度优先的最近距离贪心算法查找目标存储节点. 除 Fproxy 外,还有一些基于 FCP 开发的第三方应用,如 Frost,FMS(Freenet message system),Sone 等.

Freenet 在不同的工作模式下安全性和匿名性不同. 在 Opennet 模式下,相对更容易面临恶意节点的威胁. Freenet 多跳的文件传输和检索机制保障了文件发布者与查阅者的匿名性;文件分块加密存储保证了本地存储的可否认性;文件的冗余存储机制保证了部分节点离开网络情况下文件依然具有较高可访问性;利用 DHT 提供了高效存储与路由;非结构化的 P2P 架构避免了系统对于中心节点的依赖. 而在 Darknet 模式下,由于仅与可信的邻居节点建立连接并通信,从而提供了更高的安全性.

1.3.4 ZeroNet 网络

ZeroNet① 是一个基于 BitTorrent 技术的开源项目,由匈牙利开发者 Tamas Kocsis 创建. ZeroNet 采用早期的 BitTorrent 架构,强依赖于 BitTorrent Tracker 服务器,如图 6 所示. ZeroNet 的用户数据采用 MessagePack 协议封装,并建立 TCP 长连接进行数据传输,每个用户同时只能和 1 000 个其他用户进行连接.

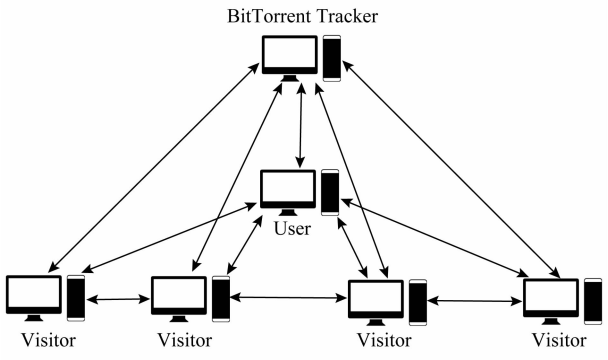

Fig. 6 The architecture of ZeroNet

图 6 ZeroNet 架构

ZeroNet 使用与 Bitcoin 钱包相同的椭圆曲线加密技术生成网站地址并对站点内容进行签名和验证,并基于 BIP32(bitcoin improvement proposals)方案为访问站点的用户生成标识. 每个新网站都会生成对应的公私密钥:私钥仅被网站所有者掌握,用于对发布内容进行签名;而公钥作为网站的地址,任何人都可以用来对网站内容的完整性进行验证,以防止恶意代码插入或任何修改.

当用户请求访问 ZeroNet 上的网站资源时,客户端软件将基于 BitTorrent 技术从拥有该资源的站点上下载文件,并在本地浏览器进行解析. 网站所有者将访问用户信息存储在 ZeroNet 客户端的内置 SQL 数据库中,在站点更新后会及时推送至用户端进行增量更新,以提高访问者的访问速度. ZeroNet 默认不提供匿名通信功能,但用户也可以通过内置的 Tor 功能实现通信匿名.

由于 ZeroNet 是一个由访问者提供服务的 P2P 网络,所以可以比集中式服务器更稳定地提供内容,甚至在互联网不可用的情况下,用户也可以离线访问本地存储的站点内容.

2 匿名通信关键技术

匿名通信系统的目标在于实现通信实体关系和内容的隐藏,具体涉及匿名接入、匿名路由、暗网服务等一系列关键技术. 本节从匿名接入、匿名路由、暗网服务 3 个方面,介绍相关技术及其研究进展.

2.1 匿名接入技术

匿名接入技术是匿名通信网络为了规避网络监管而提供的客户端隐蔽接入技术. 网络监管者出于安全审计的考虑,会检测匿名网络的中继节点并阻止用户接入. 为了规避此类网络审查,研究人员提出了各种匿名接入技术,包括 bridge,obfs,meek,FTE,Flashproxy 等.

2.1.1 bridge 节点

Tor 的设计者在 2011 年提出了 bridge 节点的概念[24]以突破网络监管. 由于此类节点的信息不公开,因此 Tor 客户端可以将其作为入口节点从而接入匿名网络.

bridge 是匿名网络为了提升自身隐蔽性和规避基于 IP 阻断的网络监管而开发的一种接入机制. bridge 节点功能上与一般的 Tor 中继节点相同,但其信息并未被目录服务器公开,从而降低了被网络监管者发现和阻断的风险. 为了方便用户获取此类节点信息,Tor 项目提供了 2 种获取方式:1)在 BridgeDB 数据库中保存了 bridge 节点信息,用户可通过直接访问官网站点获取 3 个 bridge 节点;2)用户还可通过 Google 和 Yahoo 邮箱向指定邮箱发送请求,邮件服务器会自动回复 3 个 bridge 节点信息. 由于 Google 和 Yahoo 申请虚假邮件地址比较困难,所以可有效防止无限制获取 bridge 节点信息等恶意攻击.

2.1.2 obfs 混淆代理

虽然 Tor 可以使用 bridge 节点来规避基于 IP 阻断的监管,但监管者仍能通过深度包检测(deep packet inspection, DPI)技术来识别出 Tor 流量[25]. 鉴于此,Tor 项目开发了第 1 个传输插件 obfs 混淆

① https://zeronet.readthedocs.io/en/latest/

代理. Obfs 后续又经历了 3 次版本改进,分别为 obfs2[26],obfs3[27],obfs4[28],当前最新版本的 Tor 使用的是 obfs4. 其中,obfs2 设计于 2012 年,采用分组密码(AES-CTR-128)对 Tor 的传输数据进行加密,擦除了 Tor 流量相关的标识信息,从而实现了有效混淆. 但是 obfs2 在设计上存在 2 个问题:1)握手阶段的连接容易被识别. 通信双方在传输数据之前需要交换秘密信息并协商出会话密钥,而在这一阶段的密钥种子是通过明文传输且报文格式固定,导致这种 obfs2 流量容易被中间审查机制识别并计算出双方的会话密钥. 2)可利用 Tor 客户端探测 bridge 节点以检查其是否可进行 obfs2 握手,从而实施主动的 bridge 探测攻击. 鉴于上述设计缺陷,Tor 后续又推出了 obfs3 和 obfs4 方案.

作为 obfs2 的后续设计,obfs3 使用 Diffie-Hellman 协议来计算通信双方的共享秘密,并由此生成对称密钥. 相较于 obfs2,obfs3 在握手阶段有了很大的改进,但其在密钥交换阶段仍然缺乏对 bridge 身份的验证,存在中间人攻击和主动探测攻击的风险.

为了抵御上述攻击,需要能够有效保证双方通信真实性、完整性和机密性的优化方案. ScrambleSuit[29]提出利用带外方式交换共享秘密实现通信双方的互认证,能够有效防御主动探测攻击. 在此基础上,obfs4 进一步利用 BridgeDB 实现了基于 bridge 身份验证的密钥交换. 客户端通过 BridgeDB 查询 bridge 节点,获得其 IP 地址、节点 ID 和公钥信息. 只有同时匹配这 3 个信息才能通过 obfs4 节点的身份验证并建立连接. obfs4 是对 obfs3 的重大升级,既有效地混淆了 Tor 流量,又实现了 bridge 身份验证,达到能够抵御主动探测攻击和中间人攻击的效果.

2.1.3 meek 隐蔽通道构建技术

meek 是一种基于前置域(domain fronting)[30]的隐蔽通道构建技术,其工作原理如图 7 所示. 首先,meek-client 向前置域名服务器发送一个 TLS Client Hello 消息,并将该服务器域名填入 SNI(server name indication)字段. 在 TLS 握手成功后,meek-client 把实际传输的 Tor 流量封装在 HTTP POST 载荷中,并将目标 bridge 地址写入 HTTP HOST 字段. 由于监管者无法查看加密后的 HTTP POST 内容,因此无法识别真正的目标 bridge 地址. 前置域服务器接收到 meek-client 数据后,根据 HTTP HOST 字段值将数据转发到目标 bridge 节点. 该节点上运行的 meek-server 对 HTTP 报头处理后将封装的 Tor 流量转发给后续中继节点. 在当前 Tor 网络中,meek 的实现主要依赖于 Google,Amazon 和 Azure 等大型服务提供商提供的白名单前置域名服务器,造成 Tor 客户端在访问正常网站的假象,从而规避网络流量监控.

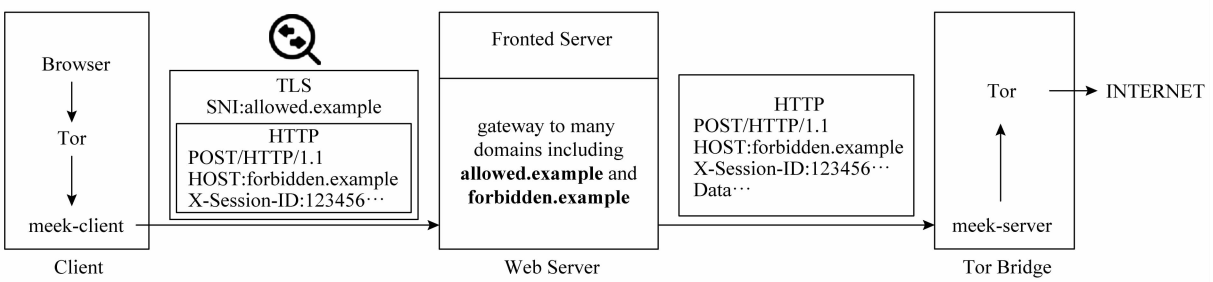

Fig. 7 The architecture of meek

图 7 meek 体系结构

2.1.4 FTE 加密流量转换技术

FTE(format-transforming encryption)[31]由 Dyer 等人在 2013 年 ACM 计算机与通信安全会议(ACM Conference on Computer and Communications Security, CCS)上发表,其主要原理是通过扩展传统对称加密,将密文转换为指定的传输格式. 与 obfs 流量混淆方式不同,FTE 没有将原协议流量转换为无序的未知流量,而是根据用户输入的正则表达式,输出具有一定协议格式的流量. 其中,用户输入的正则表达式可以从 DPI 系统源码中直接提取或根据应用层流量自动学习得到. 这使得基于正则表达式的 DPI 技术会将其误识别为用户选定的协议流量,从而实现规避审查的目的. 由于大部分 Tor 流量为 HTTP 流量,默认情况下采用 HTTP 正则表达式将之转化为 HTTP 协议,从而实现流量伪装.

2.1.5 Flashproxy 传输插件

Flashproxy[32]是运行在浏览器中的传输插件,其原理是利用不断变化的代理主机接入到匿名网络,

且保证代理更新的速度比监管机构的检测、跟踪和阻断速度快. 在 2013—2016 年期间,Flashproxy 项目一直被部署在 Tor Browser 中,直到 2017 年才弃用,被更有效的 obfs4 取代.

在使用 Flashproxy 转发数据的过程中,需要 3 个组件协同工作,包括客户端浏览器中的传输插件、代理主机 Flash Proxy 和第三方服务 Facilitator. 其中,Facilitator 负责客户端的注册和 Flash Proxy 代理的分配,当客户端需要使用 Flashproxy 代理服务时,首先通过安全集合点连接到 Facilitator 以告知客户端需要代理服务,然后客户端传输插件开始侦听远程连接.

每个空闲的 Flash Proxy 会主动轮询 Facilitator,以获得正在请求代理服务的客户端. 当 Flash Proxy 获得客户端在 Facilitator 的注册信息后,便主动发起和客户端的连接,再发起对中继节点的连接,最终客户端和目标中继节点通过 Flash Proxy 进行通信,其原理如图 8 所示:

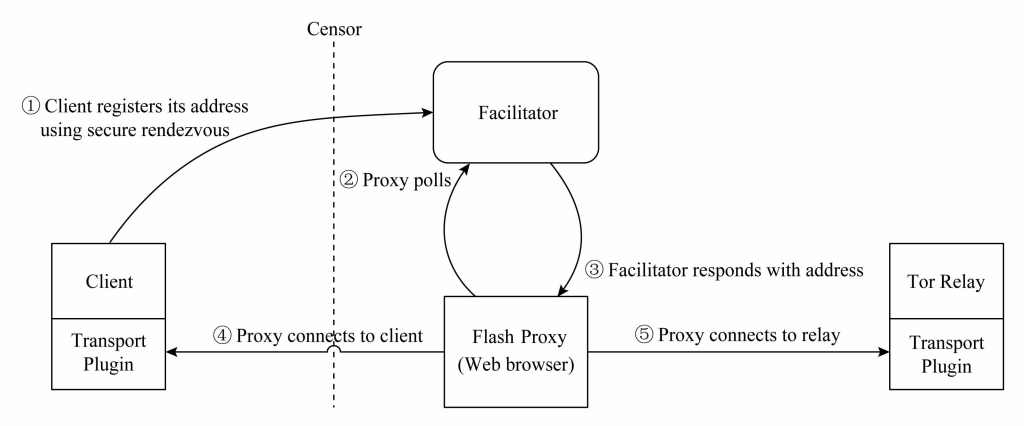

Fig. 8 The architecture of Flashproxy

图 8 Flashproxy 体系结构

通过对上述 5 种主要的匿名接入技术的调研,可以发现匿名接入技术重点针对暗网流量特征和接入节点 IP 进行混淆,从而抵御网络监管. 其中,对于暗网流量特征,主要通过加密传输消除原有流量特征或者将其伪装成其他协议流量,例如 obfs 和 FTE. 至于对接入节点的阻断,主要通过代理主机转发流量进行规避,如 meek 依赖大型服务提供商提供的白名单转发服务器. 而 Flashproxy 则是通过提高转发主机的更新频率来保证监管者无法枚举并阻断所有连接.

2.2 匿名路由技术

匿名路由技术是匿名通信系统为用户提供匿名服务时所使用的一种关键技术. 该技术具体涉及路由节点选择、链路建立、数据传输等多个阶段. 通过匿名路由技术,匿名通信系统的安全和性能都能够得到保障. 目前,低延迟匿名通信系统所使用的路由技术主要包括洋葱路由、大蒜路由和基于 DHT 的路由.

2.2.1 洋葱路由

洋葱路由技术的核心思想是通过多跳代理与层层加密的方法为用户的通信隐私提供保护. Tor 是最典型的使用洋葱路由技术的匿名通信系统,因此本节以 Tor 为例,对洋葱路由技术进行介绍.

在 Tor 暗网中,客户端与隐藏服务器之间需要各自选择 3 个节点建立多跳链路才能进行通信. 目前,Tor 主要采用基于加权随机的路由选择算法选择洋葱路由器构建链路. 该算法依据服务器描述符(server descriptor)与共识文档(consensus document)中的带宽信息与放缩因子计算各节点的加权值,并按照出口节点、入口节点和中间节点的顺序选择链路节点. 值得注意的是,链路中任意 2 个洋葱路由器应来自不同的 C 类网段.

在节点选择完成之后,OP 从入口节点开始逐跳建立匿名链路,节点之间均采用 TLS/SSLv3 对链路进行认证. 为增强可读性,下面将入口节点、中间节点、出口节点用 OR_1,OR_2,OR_3 表示. 首先,OP 与 OR_1 建立 TLS 链接,即 OP 发送 CELL_CREATE 信元,OR_1 进行响应以完成 Diffie-Hellman 握手并协商会话密钥 k_1,从而建立第 1 跳链路. 其中,CELL_CREATE 信元是 Tor 中一种数据传输的基本单元,长度为 512 B. 然后,OP 向 OR_1 发出与 OR_2 建立链路的 Tor 信元,OR_1 收到后与 OR_2 建立 TLS 链路,并通过 Diffie-Hellman 协议协商 OP 与 OR_2 之间的会话密钥 k_2. OR_1 将会话密钥通过加密报文

告知 OP,完成第 2 跳链路的建立.以此类推,Tor 建立多跳链路实现与通信目标的安全连接.

在匿名链路建立后,用户可以通过 OP 访问公共网络进行数据传输.当 OP 获得目标服务的 IP 地址和端口后,使用 k_1,k_2,k_3 对数据信元进行层层加密封装,即$\{\{\{\langle IP:port\rangle\}_{k_3}\}_{k_2}\}_{k_1}$,这种加密方式称为"洋葱加密".该信元经过每个 OR 节点时,都会被使用对应的密钥对最外层进行解密并转发.当到达 OR_3 后进行最后一次解密,即可识别出目标服务器的 IP 地址和端口,从而建立 TCP 连接.在此过程中,用户的上行数据经过 OP 层层加密,由各 OR 逐层解密并转发至目标服务器;与此相反,目标服务器的下行数据经过各 OR 加密,由 OP 逐层解密并最终返回给应用程序.对于目标服务器来说,Tor 用户是透明的,其始终认为自己在和 OR_3 通信,而各 OR 无法同时获得 Tor 用户 IP、目的服务器 IP 和应用数据,从而保证了通信的匿名性.

近年来,对洋葱路由技术的研究主要集中在负载均衡、拥塞控制、调度策略等方面:

1) 在负载均衡方面.通过正确评估洋葱路由器的有效带宽、综合考虑当前负载等方式不仅可以提升 Tor 网络性能,还可以提高安全性.例如 Snader 等人[33]认为仅根据洋葱路由器提交的带宽进行路由选择易受到恶意节点谎报虚假带宽的攻击,因此提出了一种基于洋葱路由器采样带宽中值的机会路由选择算法.在此基础上,Panchenko 等人[34]提出了一种基于节点当前负载和带宽的路由选择算法,进一步提高了性能和安全性.而 Rochet 等人[35]提出 Waterfiling 算法,通过在用户路径节点上均衡网络负载,从而降低被端到端流量关联的风险.

2) 在拥塞控制方面.通过防止拥塞发生或避开拥塞的洋葱路由器能够提升暗网性能.例如,Wang 等人[36]认为 Tor 在进行路由选择时将带宽作为中心特征,忽略了当前负载,导致部分节点未被充分利用或发生拥塞.针对该问题,他们提出了一种拥塞感知的路由选择算法,即客户端通过主动测量评估节点的拥塞状态,避免使用出现拥塞的节点,并通过实验验证了其方法的有效性.针对 Tor 网络中存在大量僵尸网络流量的问题,Liu 等人[37]提出了隐私保护访问控制框架 TorPolice,该框架允许服务提供商和 Tor 网络实施预设的访问策略,在向合法 Tor 用户提供服务的同时拒绝僵尸网络流量,缓解了网络拥塞.

3) 在调度策略方面.主要通过分析流量所属应用类型来进行链路调度,从而提升性能.例如,Tang 等人[38]认为 Tor 用户体验差的主要原因之一是在链路调度算法中,允许具有连续流量的繁忙链路(如批量传输)抢占突发链路(如网页浏览).针对此缺陷,该工作通过赋予突发链路更高优先级,使得 Tor 网络中大多数突发链路性能得到改善,同时产生的开销较小.Alsabah 等人[39]则利用机器学习方法对 Tor 加密链路进行分类,识别出每个应用程序所属的服务类型.通过为每个服务类型分配优先级,显著改善了 Tor 用户体验.在后续工作[40]中,他们提出了一种名为 Conflux 的动态分流方法,根据测量延迟将流量分配到不同的路径中以增强负载均衡特性,提升了连接到低带宽 bridge 节点的用户体验.

此外,对各匿名路由替代算法的性能与安全性评估也是一个重要的研究分支.Wacek 等人[41]探讨了路由技术在性能和安全性之间的权衡,并通过高精度拓扑模型对 Tor 网络的报文延迟、洋葱路由器转发性能等特征进行采集,进而对整个 Tor 网络进行评估.Backes 等人[42]提出了一种 Tor 网络安全性评估框架 MATor,能够评估路由算法、Tor 共识数据等对用户匿名性的实际影响.在此基础上,该文还提出了一种新的路由算法 DistribuTor,能够提供更强的匿名保障.在后续的研究中,该团队[43]对 MATor 进行完善,增加了对路由技术抵御各类 Structural 攻击能力的评估功能,并对 DistribuTor,LASTor 等路由替代算法进行了实验.Matic 等人[44]则对 bridge 节点的安全性进行了讨论.更多的此类工作可以参考文献[45].

2.2.2 大蒜路由

大蒜路由(garlic routing)具有分层加密、多消息绑定、端到端加密 3 个主要技术特点,典型代表为 I2P,因此本节以 I2P 为例来介绍大蒜路由.

I2P 制定了独特的路由算法挑选节点建立隧道[46-47].在隧道建立前,每个 I2P 节点只知道局部的节点信息,具体的节点选择过程可分为 4 个步骤:

1) I2P 节点会测量已知节点的网络性能,测量周期分为 1 min,1 h 和 24 h 等,测量项目包括带宽、netDb 查询时延和隧道建立成功率等信息,并将产生的详细描述文件保存于本地.

2) I2P 节点根据速度和容量对测量到的目标节点进行分类.其中,速度为 1 min 内通过该目标节点的数据量.为了保证结果的稳定性,通常会取 3 条最快隧道的速度平均值.容量为一段时间内该节点建立隧道的数量,对其估计:

$$capacity=0.4\times r(10\,\text{min})+0.3\times r(30\,\text{min})+0.2\times r(1\,\text{h})+0.1\times r(24\,\text{h}),\quad(1)$$

其中,$r(t)$代表最近时间 t 内成功建立隧道的数目;$capacity$ 为容量估计.

根据速度和容量可将目标节点分为 3 类:高速节点、高容量节点和标准节点,3 类节点的性能递减.高容量节点为容量超过所有节点容量平均值的节点,而高速节点为速度超过所有节点速度平均值的高容量节点,剩余的节点均为标准节点.

3) 不同的隧道类型会选择不同的节点,例如客户隧道优先选择高速节点,探测隧道优先选择高容量节点或标准节点.当同一类型节点充足时,该类型每个节点被选择的概率相同;当同一类型的节点数量不足时,会使用较低性能节点作为替代.其中探测隧道选择高容量节点和标准节点的比例 $ratio$ 可计算为

$$ratio=1-\frac{\text{探测隧道建立成功率}}{\text{客户隧道建立成功率}}.\quad(2)$$

4) 隧道建立者随机生成目标值,被选择节点按照与目标值的异或距离进行排序来确定节点在隧道中的位置.

I2P 使用 VTB(variable tunnel build)消息建立通信隧道,该消息主要由建立请求记录(build request record, BRR)组成,而 BRR 由 tooPeer 字段和 req 字段组成,其中 tooPeer 字段为节点标识,req 字段包含分层密钥 key、回复密钥 replykey、下一跳节点 ID 等信息.利用 VTB 建立隧道的过程如图 9 所示,可分为 5 个步骤:

① 客户端由 netDb 获知各节点公钥,利用公钥加密封装包含 3 个 BRR 的 1-VTB 消息,转发至节点 A;

② 节点 A 根据明文 $toPeer_A$ 字段识别对应 1-BRR,利用节点 A 私钥解密 req_A 获得分层密钥 key_A、回复加密密钥 $replykey_A$、下一跳节点 B 的 ID 等信息,通过 $replykey_A$ 加密回复数据 $response_A$ 并替换 1-BRR 内容,利用 key_A 解密 2-BRR 和 3-BRR 后封装成 2-VTB,转发至节点 B;

③ 节点 B 根据明文字段 $toPeer_B$ 识别对应 2-BRR,利用节点 B 私钥解密 req_B 获得明文数据,通过 $replykey_B$ 加密 $response_B$ 数据替换 2-BRR 内容,利用 key_B 解密 1-BRR 和 3-BRR 后封装成 3-VTB,转发至节点 C;

④ 节点 C 重复②和③中类似操作得到 4-VTB,并将 4-VTB 转发至客户端输入隧道的 Gateway 节点 D;

⑤ 客户端分层解密获得 4-VTB 中各节点响应明文数据,从而判断隧道是否建立成功.

经过上述 5 个步骤之后,客户端与节点 A,B,C 分别协商得到分层密钥 key_A,key_B,key_C 用于隧道加密通信.

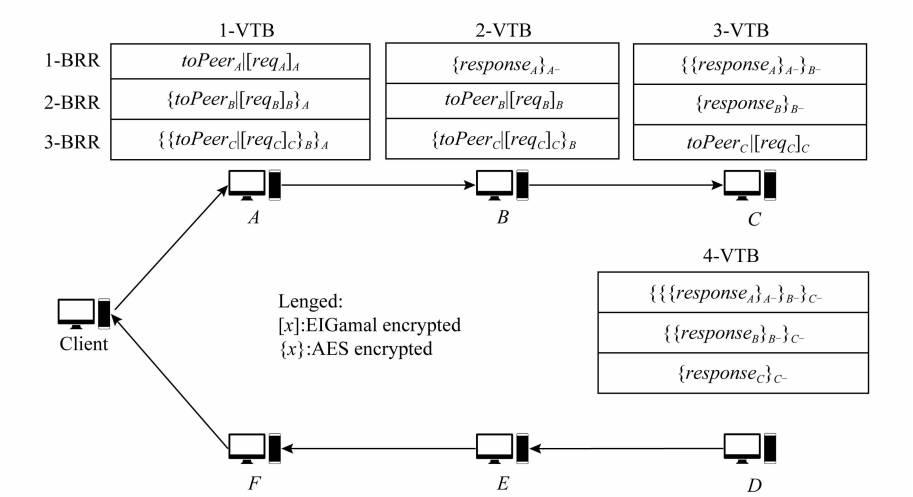

Fig. 9 Build I2P tunnel

图 9 I2P 隧道建立

大蒜路由中多消息绑定和端到端加密机制如图 10 所示.当客户端访问隐藏服务端时,使用隐藏服务 LeaseSet 中公钥信息实现端到端加密.客户端每个单独的消息称为一个蒜瓣(clove),每个蒜瓣有对应的指示信息,用于指示该蒜瓣的类型和用途.客户端将多个蒜瓣封装成一个大蒜消息进行转发,隐藏服务端分层解密并按照指示处理每个蒜瓣.

2.2.3 基于 DHT 的路由

基于 DHT 的路由技术的核心思想是利用 DHT 组织数据,进而使用对应的搜索策略提供高效路由,

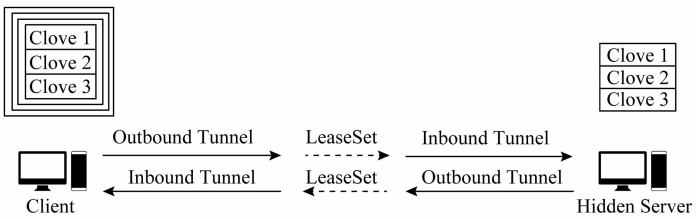

Fig. 10 I2P garlic message

图 10 I2P 大蒜消息

并通过多跳代理保障通信的匿名性[48]. Freenet 是使用基于 DHT 的路由技术的典型匿名通信系统. 本节将以 Freenet 为例,对于基于 DHT 的路由技术进行介绍.

Freenet 路由是一种面向请求消息的深度优先节点选择机制,主要应用于文件上传和文件检索[49]. 请求消息中除了包含文件索引,还包含 UID(unique identifier),HTL(hops to live)等信息. 文件索引值对应一个节点位置值,表示该文件最可能的存储位置,是请求消息路由的重要依据. UID 唯一地标识一个请求,能够有效避免消息路由过程中出现死循环. HTL 用来限制消息所允许转发的跳数,避免消息无限地被路由下去. 目前 HTL 默认的最大值为 18,通常情况下随着消息的转发 HTL 会逐跳减 1,当 HTL 减至 0 时,该消息将被丢弃. 此外,出于安全性考虑,Freenet 中特别增加了一些针对 HTL 的混淆机制,对于 HTL 为 18 的请求,节点在转发时只有 50%的概率会使 HTL 减 1;而对于 HTL 为 1 的请求,节点在转发时只有 25%的概率会使 HTL 减 1. 综上,请求消息由发起者节点产生后,会按照深度优先原则被转发到下一跳,然后由下一跳继续迭代路由. 上传请求会在检测出键值冲突或 HTL 下降至 0 时停止转发,而检索请求则会在检索成功或 HTL 下降至 0 时停止转发.

Freenet 路由选择机制还结合了贪心策略. 每个节点维护自身的路由表,并根据该表对消息进行路由. 为了便于区分,Freenet 将与节点直接通信的邻居节点称为直接邻居,距离两跳的节点称为间接邻居. 值得注意的是,Freenet 中直接通信的节点之间会交换双方直接邻居的位置等信息,从而保证任意节点都拥有所有间接邻居的位置等信息. 图 11 展示了节点 A 的直接邻居、间接邻居以及其路由表. 其中,节点 A 的路由表中目的地节点包括所有直接邻居和间接邻居,而路由选择的结果始终是直接邻居. 每次路由时,节点将根据消息中文件索引值确定对应的位置,并比较路由表中目的地节点的位置与该位置的距离,迭代地选择距离最近的目的地节点对应的直接邻居作为消息的下一跳节点. 对于需要重复路由的消息,节点继续按照贪心策略选择不重复的下一跳节点.

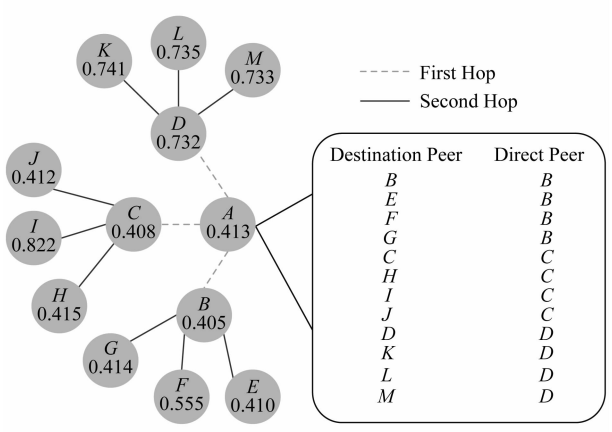

Fig. 11 Routing table in Freenet

图 11 Freenet 路由表

2.3 暗网服务技术

暗网服务是指必须通过特殊的软件、特殊的配置才能访问,并且搜索引擎无法对其进行检索的服务,包括了 Web,IRC、文件共享等多种服务类型. 由于匿名通信技术的引入,暗网服务器的位置信息通常都得到了有效的隐藏,使得大部分的暗网服务都具有较好的匿名性. 下文分别对典型匿名系统提供的暗网服务技术进行介绍.

2.3.1 Tor 隐藏服务

Tor 隐藏服务[①]是通过隐藏服务域名(即洋葱域名)唯一标识和查找的. 服务器首次运行后将生成一个隐藏服务域名,其域名形式为〈z〉.onion. 其中〈z〉是长度为 16 B 的字符串,由 RSA 公钥 Hash 值的前 80 b 进行 base32 编码获得.

① https://gitweb.torproject.org/torspec.git/tree/rend-spec-v2.txt

Tor 隐藏服务器在启动过程中会将其信息上传至隐藏服务目录服务器,Tor 客户端能够通过目录服务器获取足够的信息与隐藏服务器建立双向链路. 图 2 展示了客户端访问隐藏服务的具体过程:

① 隐藏服务器选择 3 个洋葱路由器作为其引入节点,并与引入节点建立 3 跳链路;

② 隐藏服务器将其隐藏服务描述符(hidden service descriptor)上传至隐藏服务目录服务器,描述符中包含引入节点的信息与自身 RSA 公钥;

③ 客户端通过隐藏服务域名(〈z〉.onion)进行访问时,从隐藏服务目录服务器获取引入节点的相关信息;

④ 客户端选择一个洋葱路由器作为汇聚节点并与该节点建立 3 跳链路;

⑤ 客户端建立到达引入节点的 3 跳链路,并通过引入节点将汇聚节点的信息发送到隐藏服务器;

⑥ 隐藏服务器建立到达汇聚节点的 3 跳链路,并对该链路进行认证;

⑦ 经过汇聚节点,客户端与隐藏服务器通过 6 跳链路进行交互.

Tor 隐藏服务器运行时,将自身的 RSA 公钥及引入节点信息上传至通过式(3)选定的 6 个隐藏服务目录服务器上,其中 $H(x)$是结果长度为 160 b 的 Hash 函数,‖ 为拼接函数.

$$Descriptor\text{-}id = H(Public\text{-}key\text{-}id \parallel Secret\text{-}id\text{-}part), \tag{3}$$

其中,*Public-key-id* 即为 RSA 公钥的 SHA1 摘要的前 80 b. *Secret-id-part*,可计算为

$$Secret\text{-}id\text{-}part = H(Time\text{-}period \parallel Descriptor\text{-}cookie \parallel Replica\text{-}index), \tag{4}$$

其中,*Descriptor-cookie* 为可选字段,该字段可以用来防止未授权客户端访问隐藏服务;*Replica-index* 有 2 种取值,其作用为选择不同的隐藏服务目录服务器. *Time-period* 字段用于分散隐藏服务重新选择隐藏服务目录服务器的时间,可计算为

$$Time\text{-}period = \frac{t + P[0:1] \times \frac{86\,400}{256}}{86\,400}, \tag{5}$$

其中,t 为当前的 UNIX 时间. 该字段将一天划分为 256 个时段,并在其中某个时刻重新选择隐藏服务目录服务器上传信息.

隐藏服务目录服务器利用 DHT 进行组织,每个目录服务器都是 DHT 中的一个节点,DHT 结构如图 12 所示:

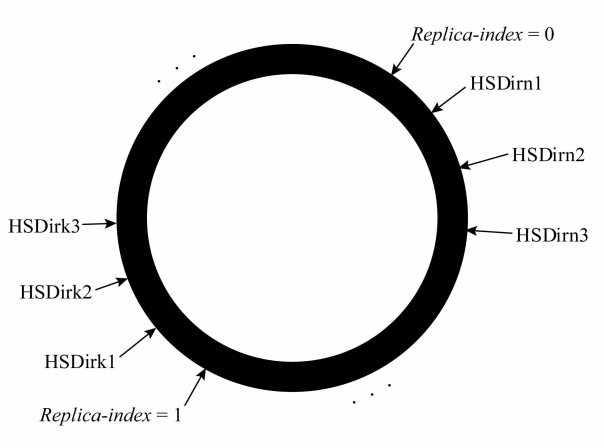

Fig. 12 The topology of hidden service directories

图 12 隐藏服务目录服务器结构图

对隐藏服务技术的研究主要集中在 Tor 暗网安全性分析及性能分析 2 个方面. Øverlier 等人[50]为增加隐藏服务器抵御 DoS 攻击的能力,提出了代客节点(valet service node)的设计. 客户端欲访问隐藏服务首先需要通过联系信息凭证获取代客节点信息,并且需要通过代客节点认证才能访问引入节点,从而保证了引入节点的安全性,最终增强隐藏服务器防御 DoS 攻击的能力. Loesing 等人[51]评估了在客户端访问隐藏服务过程中获取隐藏服务描述符、建立到达引入节点的链路等步骤的开销,并提出了提前建立更多链路、同时连接 2 个引入节点等建议提高 Tor 暗网性能. Lenhard 等人[52]评估了在使用蜂窝移动网络等低带宽接入启动阶段和隐藏服务访问阶段的性能,在此基础上,提出了在低带宽情况下设置较大网络超时等建议以提升用户体验.

2.3.2 I2P 隐藏服务

I2P 提供的隐藏服务① 类型包括 Web、邮件、聊天室和文件共享等,其中 Web 是主流的隐藏服务,称为 eepSite,支持公开和非公开 2 种访问方式. 公开的 eepSite 的访问链接是 *Hostname*,非公开的 eepSite 可以通过 *Address* 访问:

$$Destination = 256\,\mathrm{B}\ public\ key + 128\,\mathrm{B}\ signing\ key + 3\,\mathrm{B}\ certificate. \tag{6}$$

$$LSHash = SHA256(Destination). \tag{7}$$

$$Address = Base32(LSHash) + \text{`.}b32.i2p\text{'}. \tag{8}$$

I2P 中的隐藏服务通过 *Destination* 标识,如式(6)所示,*Destination* 由 256 B 公钥、128 B 签名公钥

① http://www.i2pproject.net/zh/docs/how/tech-intro

和 3 B 的证书类型组成. 对于非公开的隐藏服务, I2P 根据式(7)由 *Destination* 计算 LeaseSet 的 Hash 值(简称为 LSHash),再根据式(8)编码获得隐藏服务的 *Address* 信息. 对于公开的隐藏服务,服务拥有者需将 *Destination* 以及便于记忆的站点别名 *Hostname* 提交到 I2P 隐藏服务管理站点,经过其审核后公布给 I2P 用户,用户可以通过别名 *Hostname* 访问隐藏服务站点.

隐藏服务访问者通过 LSHash 查询 Floodfill 节点得到对应隐藏服务的 LeaseSet 信息,LeaseSet 提供访问隐藏服务的 Gateway 节点信息,以此构建隧道访问隐藏服务.

如图 13 所示,在考虑默认 3 跳隧道的情况下, I2P 隐藏 Web 服务的通信过程包括 7 个步骤:

① 隐藏服务根据节点选择机制,选择合适的节点建立隐藏服务的输入和输出隧道,并根据输入隧道入口节点信息生成隐藏服务的 LeaseSet 信息.

② 隐藏服务建立探测隧道,并通过探测隧道将①中生成的 LeaseSet 信息公布到 netDb 中.

③ 客户端根据路由选择机制,选择合适的节点建立客户端的输入和输出隧道,并根据输入隧道入口节点信息生成客户端的 LeaseSet 信息.

④ 客户端建立探测隧道,并通过探测隧道向 netDb 查询获得隐藏服务的 LeaseSet 信息.

⑤ 请求 Web 隐藏服务的消息包含 3 个 Clove,其中 Clove1 为 HTTP 请求;Clove2 为数据存储消息(database store message),即客户端的 LeaseSet 信息,用于指示隐藏服务端到客户端输出隧道和输入隧道的转发;Clove3 为传输状态消息(deliver status message),用于隐藏服务端响应并确认客户端消息.

⑥ 客户端数据经过输出隧道 3 次解密后得到端到端的加密数据,同时输出隧道末端点得到步骤③中隐藏服务 LeaseSet 所指示的入口节点信息,完成通信双方输出隧道到输入隧道的数据转发.

⑦ 隐藏服务端对数据进行 4 次解密得到明文数据,同时利用明文数据中的 Clove2 信息实现隐藏服务输出隧道至客户端输入隧道的转发.

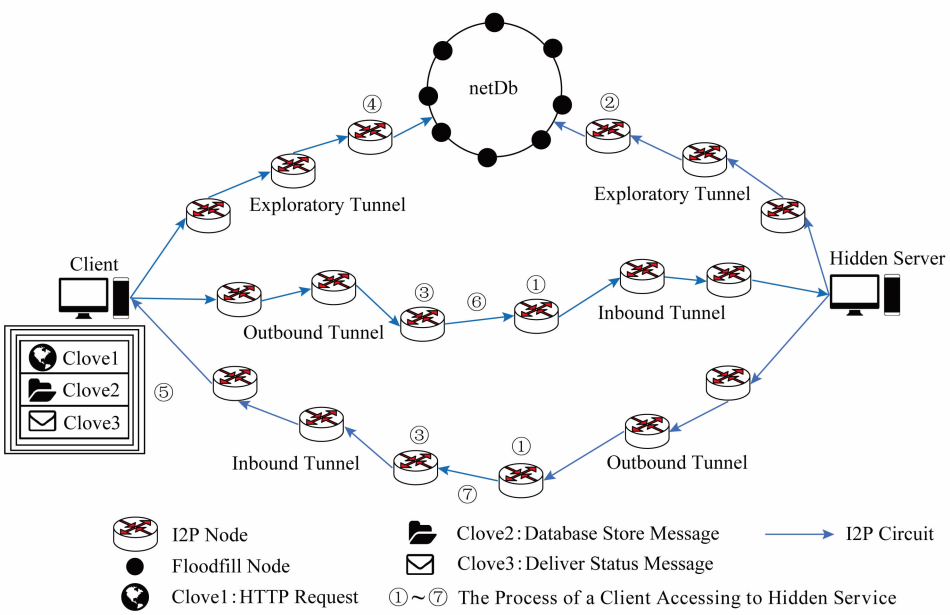

Fig. 13 Hidden Web service communication in I2P

图 13 I2P 隐藏 Web 服务的通信过程

经过以上 7 个步骤,完成了客户端和隐藏 Web 服务端的通信,双方可得知输入隧道入口节点的信息. 由于隧道中节点只知道前驱和后继节点信息,从而保证了通信的匿名性.

2.3.3 Freenet 暗网服务

Freenet 基于其匿名文件存储与检索功能提供了暗网 Freesite 服务. Freesite 是 Freenet 中的“网站”,本质上是存储在 Freenet 中的静态网页文件. 该隐藏服务基于 Freenet 文件存储与检索中的统一资源定位符(URL)以及路由机制进行部署和访问. Freesite 利用 Freenet 的文件键值作为其统一资源定位符,对于任意一个 Freesite 站点,只有获取其键值才能够访问该网站查看相关信息.

键值是 Freenet 中文件的统一资源定位符,根据不同的文件分享需求,Freenet 提供了不同种类的键值. 键值类型包括 CHK(content Hash keys)和

SSK(signed subspace keys)两大类,而 USK(updatable subspace keys)和 KSK(keyword signed keys)是 2 种特殊的 SSK[53]. Freesite 利用 SSK 或 USK 作为其键值类型将静态网页文件上传至 Freenet 中,这 2 种键值类型的文件很容易进行更新,便于所有者维护自己的 Freesite. 用户获取了文件的键值,即可在 Freenet 中利用 Fproxy 代理功能在浏览器中访问该 Freesite.

CHK 是最基础的一种键值,适合静态文件资源使用. 这种键值包括文件的 Hash 值、解密密钥和加密设置等信息,其格式为"CHK@Hash 值,解密密钥,加密设置". CHK 通过对二进制文件内容进行 SHA256 Hash 运算生成一个由文件内容决定的 Hash 值,该 Hash 值也作为对应文件在 Freenet 中的索引,生成该索引的同时会生成一个随机密钥用来对文件内容进行对称加密,文件被加密存储在 Freenet 中. 键值中的加密设置主要包括使用的加密算法等信息.

SSK 适用于 Freesite 站点等需要文件频繁更新操作的资源,可以保证只有文件所有者才能更新. 这种键值由 5 部分组成:公钥 Hash、文件解密密钥、加密设置、文件描述符以及版本号. 其格式为"SSK@公钥 Hash,文件解密密钥,加密设置/文件描述符-版本号". 文件所有者在上传 SSK 类型的文件时,首先需要生成一对公私钥,其中公钥的 Hash 值将作为该文件在 Freenet 中的索引,而私钥将用来对文件密文进行签名从而提供完整性校验. 另外,文件所有者需要提供一个文件对称加密密钥,并将原始文件加密后的密文、利用私钥对密文的签名以及公钥数据,共同存储在节点中,以便用户对其进行校验. Freesite 拥有者还可以为该文件指定一个简短的文本字符串作为文件描述符,同时用版本号指明了当前文件的版本,用以区分文件的不同版本.

USK 与 KSK 是 2 种特殊的 SSK. USK 主要用于链接到最新版本的 SSK,这种键值本质上就是对 SSK 进行了一层封装,向用户隐藏了对于最新版本 SSK 的搜索过程. 这种键值格式为"USK@公钥 Hash,文件解密密钥,加密设置/文件描述符/版本号/". KSK 是简化的 SSK,仅仅由一个描述性的文本字符串构成,其格式为"KSK@文本字符串". 当选择向 Freenet 中上传这种键值类型的文件时,用户只需要提供一个描述性的文本字符串. 节点首先根据用户提供的字符串生成一对公私钥,然后同样对公钥进行 Hash 产生该文件的索引,私钥则用来对文件进行签名以提供一定的完整性校验. 文件最终利用该文本字符串作为对称密钥加密后与签名一同存储在 Freenet 中. KSK 存在的主要问题是容易发生冲突,当 2 个用户同时选择了同一个文本描述符,两者存储的对应文件则具有相同的索引.

用户利用键值检索对应文件资源的过程是节点选择过程. Freenet 中一种典型的资源请求过程如图 14 所示,节点 A 作为请求者在其本地存储未检索到对应文件后,向节点 B 发起请求. 节点 B 收到该请求后,在本地未检索到文件,于是向与文件索引值最近的节点 C 发起文件请求. 节点 C 无法检索到该文件返回了请求失败消息. 节点 B 收到该消息后继续向次近的节点 D 发起请求,节点 D 先向节点 F 发起请求,节点 F 又向其邻居节点 B 发起请求. 此时在节点 B 处检测到请求的循环,节点 B 返回请求失败给节点 F. 节点 F 没有其他邻居节点,向节点 D 返回请求失败消息. 节点 D 此时向次近的邻居节点 E 发起请求,当节点 E 检索到该文件,将沿路径 $E\rightarrow D\rightarrow B\rightarrow A$ 反馈结果给节点 A.

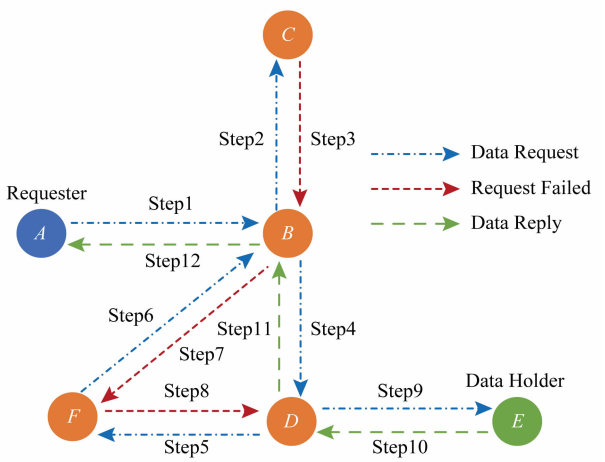

Fig. 14 A typical request sequence in Freenet

图 14 一种典型的 Freenet 资源请求过程

2.3.4 ZeroNet 暗网服务

ZeroNet 基于 BitTorrent 技术部署用户站点,并通过 Bitcoin 加密技术保证站点文件的完整性. 当用户访问一个 ZeroNet 的网站资源时,首先查找本地资源,若不存在则通过 BitTorrent 网络查找该网站资源的拥有者,将拥有资源的部分访问者的 IP 地址以及端口号返回给用户. 因为站点的每个访问者均存储了站点文件,使得用户可以通过 TCP 连接从这些访问者中下载网站内容. 具体步骤如下:

① 用户查找网站资源的拥有者. 该操作是在 BitTorrent Tracker 和用户之间进行的,如图 15 所示. 用户向 BitTorrent Tracker 请求网站资源时,Tracker 将用户注册成为访问者,并将拥有该资源的部分可访问网站的 IP 及端口号返回给用户.

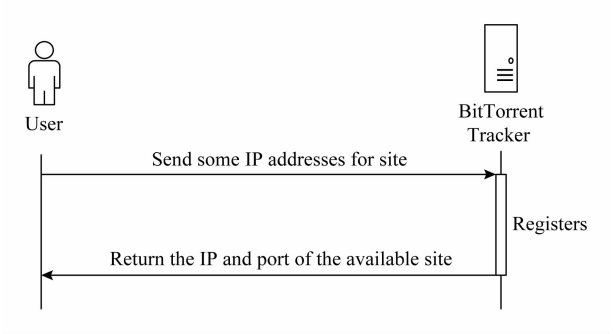

Fig. 15 Workflow of user looking up the owner of resource

图 15 用户查找网站资源的拥有者流程

② 用户下载网站资源. 该操作在用户和其他访问者之间进行,其他访问者是拥有用户请求资源的可访问网站,如图 16 所示. 在步骤 1 之后,用户获得了可访问网站的 IP 及端口号,首先会下载该网站的签名信息集文件,该文件包含所有文件名、Hash 值和网站所有者的加密签名. 之后用户使用文件中网站地址和网站所有者的签名来验证下载的签名信息集文件. 验证完毕后用户可以使用签名信息集文件中的 SHA512 Hash 值下载并检验网站中的其他文件,包括 HTML,CSS,JS 等.

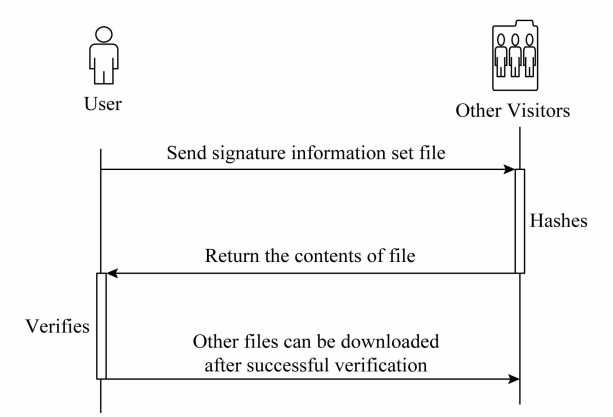

Fig. 16 Workflow of user downloading site resources

图 16 用户下载网站资源流程

用户在访问网站后即可为其他访问者提供该网站的下载服务,提升整个站点文件的下载速度. 用户可使用 Tor 网络与 BitTorrent Tracker 服务器通信以隐藏其真实 IP 地址.

ZeroNet 网站由其所有者更新,为该网站提供文件下载服务的所有用户节点仅接收对网站内容所做的增量更新. 该操作是在网站所有者和网站访问者之间进行的,如图 17 所示. 如果网站所有者修改了站点内容,他将签署一个新的签名信息集文件,该文件会自动映射到内置的本地数据库,同时将其发送给访问者. 访问者使用签名验证文件的完整性,将修改后的文件内容保存在本地,以供其他访问者访问下载,浏览器也会实时更新修改后的网站内容.

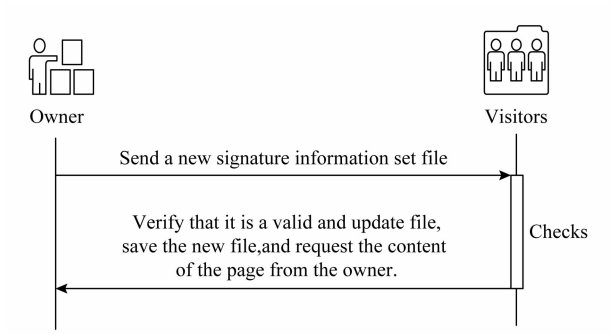

Fig. 17 Workflow of ZeroNet site update

图 17 ZeroNet 网站更新流程

ZeroNet 站点可同时由多个用户进行维护,此类网站需要其他维护站点的用户向网站所有者发送维护申请许可,如图 18 所示. 用户需要发送其地址(即公钥)给网站所有者,网站所有者在签名信息集文件中加入该地址,并生成基于 BIP32 的 Bitcoin 地址给每个用户,表明该用户为有效签名者,并向所有的访问者发送新的签名信息集文件,以更新网站的发布权限.

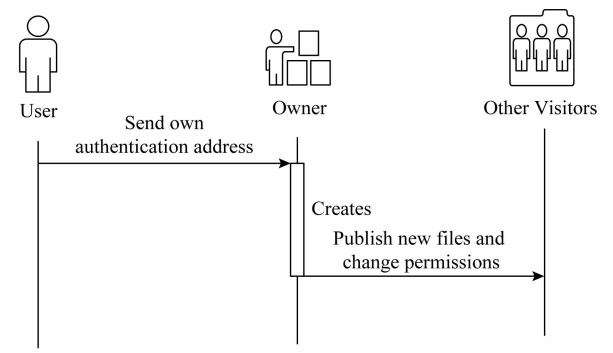

Fig. 18 Workflow of multi-user site registration

图 18 多用户网站注册流程

综上所述,ZeroNet 是一个由访问者提供服务的 P2P 暗网. 虽然该系统没有采用任何的匿名技术保护站点和用户之间的通信隐私,但它利用 BitTorrent 网络架构并结合 Bitcoin 加密技术将站点文件组织在一个封闭的 P2P 环境中,使得搜索引擎无法检索到站点内容,形成了 ZeroNet 暗网.

3 匿名通信攻防研究

围绕匿名通信系统的隐私保护效果,研究人员开展了相关的攻防研究.这些工作探讨了匿名通信系统的设计缺陷,并寻求相应的防御方法.

3.1 匿名通信攻击技术

匿名通信攻击可以分为针对网络流量的攻击和针对节点的攻击:针对流量的攻击通常基于流量分析技术,即通过嗅探并分析网络通信流量的各种模式,以获取有价值信息;而针对节点的攻击则主要是匿名通信节点系统和协议存在漏洞的利用.

本节首先针对现有的基于网络流量分析的去匿名化攻击技术,从2个维度[54]介绍网络流量攻击:

1) 被动和主动攻击.根据攻击者对通信行为的干涉程度进行划分,攻击者可以被动监听或主动操纵目标流量.

2) 单端和端到端攻击.根据威胁模型的不同进行划分,攻击者通过监控或控制发送方或/和接收方的相关设备.

被动攻击是指攻击者通过被动网络窃听分析抽取流量特征,在这个过程中并不会影响数据的正常传输,其优势在于隐蔽性强.主动攻击则是攻击者对数据通信过程本身施加干扰,例如对数据包进行修改、重放、丢弃或延迟等操作,从而达到更高效地进行流量特征分析和抽取的目的.

单端攻击是指仅占据发送端或接收端实施的攻击技术.执行单端攻击的攻击者需要监控通过发送端或接收端设备的流量,从而破坏用户的通信保密性和匿名性.单端被动攻击提取流量模式作为指纹,并基于此推断应用层的流量内容,例如用户访问的网站.单端主动攻击则可以主动将内容注入到流量中,或控制客户端的入口节点以迫使客户端直接向攻击者发送信号并暴露其真实IP地址.

端到端攻击指同时占据通信入口和出口实施的攻击,可以采用被动或主动攻击方式来关联客户端和服务器之间的通信关系.为了实施攻击,攻击者通过在发送器和接收器两侧控制或监听设备(例如两端路由器,或Tor入口和出口节点),然后基于流量特征比对嫌疑发送端发出的流量和嫌疑接收端收到的流量,以确认通信关系.

本节将分别从端到端被动攻击、端到端主动攻击、单端被动攻击和单端主动攻击这4个方面阐述现有相关工作.此外,还将介绍针对节点系统和协议漏洞的相关攻击技术.

3.1.1 被动端到端流量分析

端到端被动攻击的目的是被动记录流量,并用统计方法来评估发送者的出站流量和接收者的入站流量之间的相似性.图19说明了端到端被动攻击的基本工作流程.此类技术可以利用的流量特征包括数据包计数、流量模式相关性、时序相关性等.例如,攻击者可以简单地在发送器的输出链路上对几个时间间隔中输出的数据包数目进行计数,然后在接收器的输入链路上对相同时间间隔中到达的数据包数目进行计数,最后用距离函数来计算这2个链路在相关流量特征上的距离.

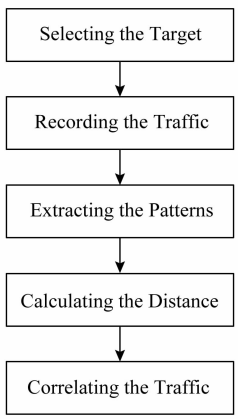

Fig. 19 Workflow of end-to-end passive attacks

图19 端到端被动攻击的工作流程

由于端到端被动攻击仅需要监听流量,因此它的主要优点是具有较高的隐蔽性.然而,这种方案的准确率低而误报率高.因此,攻击者需要足够的时间来观察流量并发现发送方和接收方之间的流量模式相似性,以便降低误报率.

根据攻击者能力的不同,现有被动端到端攻击可以分为2种:一种是攻击者可实施全局监听,例如可以监听整个自治域系统(autonomous systems, AS)或互联网交换中心(Internet exchange, IX)内的信道;另一种是攻击者可实施局部监听,例如向匿名通信网络注入节点,通过一定策略增加自身被选中的概率,从而在某个节点上实施攻击.

1) 全局监听攻击.全局监听攻击是指攻击者可以在独立的AS或IX甚至更大范围地对匿名链路进行监控分析,从而实现通信关系的确认.由于Tor现有的路由选择策略会尽可能使一条链路穿越不同的国家,这就意味着即使是AS/IX级别攻击者

也并不一定能同时监控出口和入口节点. 而 Edman 等人[55]的研究证实了单个自治系统能够通过流量观测和关联来威胁 Tor 暗网安全性,在此基础上该工作评估了一些简单的 AS 感知路径选择算法的有效性,并提出了一种路径感知路由选择算法,可以改善路由选择中的位置多样性. Akhoondi 等人[56]设计了一种高效算法,用于判断 Tor 链路是否能够被 AS 进行流量关联,并基于该算法设计了一种新的 Tor 客户端 LASTor. LASTor 能够避免建立可被 AS 流量关联的 Tor 链路,从而提高了 Tor 暗网的安全性. IX 具有比 AS 更大的监控范围,Murdoch 等人[57]验证了 IX 级别攻击,通过对真实的流量进行采样,提取报文发送率、报文长度等统计特征,实现了实体通信关系的确认. Johnson 等人[58]在仿真环境下进行实验,在同时控制具有高带宽的 Tor 节点以及 IX 的条件下,可以对 80% 的随机链路实现通信关系的确认. 在 AS/IX 级别攻击中,攻击者需要对相当大范围的网络流量进行监控和分析,这就对其所能掌控的资源有很高要求. Nasr 等人[59]设计了一个流量关联系统 DeepCorr,利用深度学习架构学习适用于 Tor 复杂网络的流量关联函数,可将 Tor 连接关联起来,准确度明显强于现有算法.

2) 局部监听攻击. 局部监听攻击是指通过向网络中提供满足带宽、在线时间要求的恶意节点,使其成为匿名系统的一部分来实施攻击. 而随着匿名网络规模的不断扩大,受限于攻击成本,攻击者无法提供足够的高带宽节点来获得更多链路的控制权,因此需要采取一些措施提高注入节点被选中的概率. Bauer 等人[60]提出了通过上传虚假高带宽信息提高恶意节点被选中概率的策略. Pappas 等人[61]利用恶意用户节点构建环形链路,恶意消耗链路中合法中继节点的资源,最终使其资源耗尽并拒绝服务,从而间接提高恶意节点被选中的概率. 随着一些新提出的链路构建方法可自动选取相对可靠中继节点,局部监听攻击的有效性面临着更多的挑战.

3.1.2 主动端到端流量分析

端到端主动攻击的基本思想是攻击者可以通过在发送端或接收端的目标流量中嵌入特殊信号,然后检测接收端或发送端侧的流量,以便识别信号并确认发送端和接收端之间的通信关系. 这种类型的攻击也被称为基于流水印的攻击,工作流程如图 20 所示.

对于流水印嵌入,攻击者可以利用不同网络层次的多种特征. 本节分别从网络层、协议层和应用层

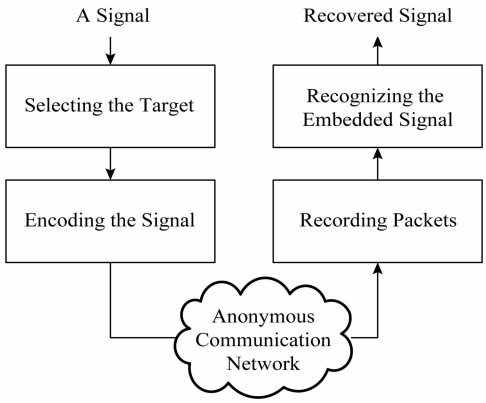

Fig. 20 Workflow of end-to-end active attacks[62]

图 20 端到端主动攻击的工作流程[62]

3 个层面阐述现有的基于流水印的攻击方法.

1) 在网络层. 攻击者可以利用诸如流量速率[63]、包延迟间隔[64]和包大小等特征来将信号嵌入到目标流量中. 例如 Yu 等人[63]提出一种水印嵌入方法,将不可见的直序扩频信号(direct sequence spread spectrum, DSSS)嵌入在流量中,以调制其流速率模式. 然后,嵌入信号和数据内容通过匿名通信网络一起传输到接收方. 最后,攻击者可以识别信号并破坏发送方和接收方之间的匿名性. 为了减少时延抖动的影响,Wang 等人[65]提出一种攻击技术,可追踪互联网上的 P2P 匿名 VoIP 电话. 该技术的思想是在选定流的 n 个独立同分布的数据包中随机选择 $2r$ 个数据包,分为 r 对数据包,通过调整 r 对数据包的 IPD(inter-packet delay)差值的平均值来嵌入追踪信息. 实验研究表明追踪 P2P 匿名 VoIP 电话是可行的,该技术可有效地追踪任何长于 90 s 的 P2P 匿名 VoIP 电话. Wang 等人[64]研究了基于质心的包延迟间隔的水印技术,通过调整时间间隔质心,可以将一系列二进制信号位嵌入流量中. Houmansadr 等人[66]利用非盲水印的思想调制报文间隔,并且使用扩频技术降低了先前技术中所用的时延,用少量的数据包就能保证流水印的鲁棒性和隐蔽性. 为了抵御多流攻击,Houmansadr 等人[67]还设计了 SWIRL 流水印机制,该方案是第 1 个可用于大规模流量分析的盲水印,可抗多流攻击和 Tor 网络拥塞,并以不同模式标记每个流. SWIRL 对于包丢失和网络抖动具有鲁棒性,且引入的延迟较小,因此隐蔽性较高. Rezaei 等人[68]提出了一种盲流水印识别系统 TagIt,将水印信号调制成网络流的定时模式,即将数据包稍微延迟到只有水印识别方知道的隐蔽时间间隔,从而实现对第三方不可见.

此外,可利用随机化来抵抗各种检测攻击,如多流攻击.攻击者还可以改变包大小来将信号嵌入到目标流量中.例如,Ling 等人[69]提出了一种攻击方法,可以控制 Web 服务器并操纵响应 HTTP 数据包的大小,用特定的包长度表示单个十六进制位.通过改变几个包的长度,可以将消息编码到流量中.尽管包长度在单跳代理处被部分填充,但是攻击者仍然可以推断包长度,以便在接收端侧恢复原始信号来确认发送端和接收端之间的通信关系.另外,为了保证这种攻击的隐蔽性,攻击者需要保持原始包大小的分布和自相似性.为此,攻击者需要选择适当的包并改变其大小.

2) 在协议层.水印攻击可以采用匿名通信系统的不同协议特征.例如,Ling 等人[70]在可控 OR 节点上控制 Tor 信元发送规律,在出口节点处连发 3 个 Tor 信元代表信号 1,1 个 Tor 信元代表信号 0,并通过分析信元在网络逐跳传输中可能出现的变化,设计信号恢复算法;在入口节点处对信号进行识别,可以在较短的时间内实现对 Tor 匿名流量通信关系的快速确认.Ling 等人[71]深入探讨了 Tor 的通信协议,发现 Tor 使用高级加密标准(AES-CTR)的计数器模式来加密和解密 Tor 信元.因此,每个 Tor 节点,包括链路中的 Tor 客户端,维持本地计数器以使计数器值彼此同步,以便正确地加密或解密 Tor 信元.该攻击利用多跳路径中的计数器同步机制的特征,并且干扰沿着该路径的某节点处的计数器值,会导致 Tor 信元的加密/解密失败.通过这种方式,攻击者可快速确认链路的源地址和目的地址之间的通信关系.

3) 在应用层.攻击者可以在服务器端将特殊内容注入目标网络响应流量中,以强制客户端生成特殊流量模式作为信号.然后,攻击者可以在客户端观察到该信号并确认发送方和接收方之间的通信关系.Ramsbrock 等人[72]在应用层改变数据包长度,通过在数据包中填充字符来嵌入信号.Wang 等人[73]提出一种攻击方案,一旦攻击者发现通过其出口节点的目标页面响应流量,则可以注入空图像的恶意网络链接,使得客户端侧的浏览器下载这些链接产生特定的流量模式.若攻击者在入口节点处检测到期望的流量模式,便可确认发送者和接收者之间的通信关系.Chakravarty 等人[74]在 Web 服务器端加入代码让用户下载一个较大且不易被察觉的文件,然后根据统计相关性在收集到的众多入口 NetFlow 流量记录中找到符合此流量特征的入口节点,从而确认通信关系.此类攻击还可在用户的返回流量中注入 JavaScript 代码,以触发用户端浏览器产生特定的信号流量.

3.1.3 被动单端流量分析

单端被动攻击的思路是监控目标和匿名代理之间的流量,并通过将预期流量模式与预先收集的网站指纹进行比较来识别实际访问的网站.这种类型的攻击也被称为网站指纹(website fingerprinting,WF)攻击.

图 21 说明了单端被动攻击的基本工作流程,包含离线训练和在线分类 2 个阶段.在离线训练阶段,攻击者首先需要选择多个感兴趣的网站,并设置被攻击者的环境以模仿其浏览活动的过程.然后,攻击者将逐一浏览网站并收集相应的流量.此外,需对收集的数据进行预处理以便去除噪声.例如,因为网页上的广告链接是动态的,网站指纹攻击的准确性将受到影响.因此,需要通过使用一些预处理策略,对此类噪声进行过滤.此外,攻击者从预处理的流量中提取合适的特征来表示通常隐藏在流量中真实有效的模式.目前已被验证的有效特征包括包长度分布、流量、总时间、流量方向、包长度顺序、上行/下行字节、突发流量字节长度[75]等.最后,攻击者选择合适的分类器,利用采集到的样本数据来生成分类规则.该类攻击中常用的分类器包括多项式朴素贝叶斯分类器、支持向量机、决策树等.在第 2 阶段,攻击者利用监控工具捕获目标用户和匿名代理之间的流量,并利用相同的预处理方法去除噪声,然后根据训练好的分类模型对流量进行特征匹配和识别.

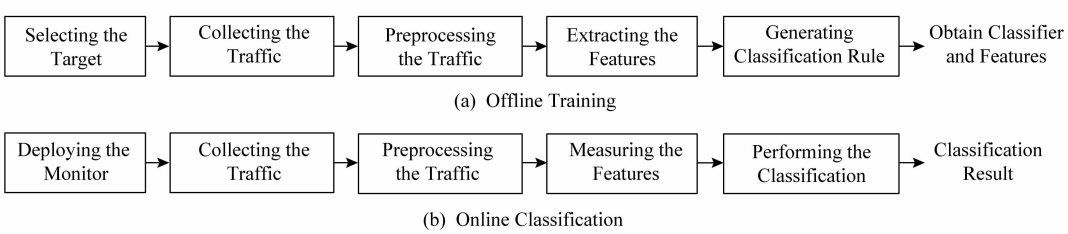

Fig. 21 Workflow of single-end passive attacks

图 21 单端被动攻击的工作流程

Hintz[76]最先提出了 Web 站点指纹攻击的概念,并在理论上证明了指纹攻击的可行性与有效性,但该攻击方案仅适用于 HTTP1.0 协议,对之后的 HTTP 协议不再有效. Liberatore 等人[77]在前人工作的基础上,仅使用报文长度分布为特征,并首次将机器学习领域的朴素贝叶斯分类器应用于指纹识别,大大提高了单跳匿名代理上指纹攻击的成功率. 但该方案依赖于上下行流量中报文的长度,并不适用于对报文进行了定长封装处理的多跳匿名通信系统(例如 Tor),具有较大的局限性. Lu 等人[78]指出数据包长度顺序信息,可用于增强网站指纹识别. Panchenko 等人[79]综合多种流量特征,包括特定长度报文出现次数、总传输量、上下行报文数据量及所占比例等,并使用支持向量机对指纹进行分类,将多跳匿名代理上的识别率从 3%提升至 55%. Cai 等人[80]使用最佳字符串编辑距离(optimal string alignment distance, OSAD)替换支持向量机的核函数衡量指纹相似性,在 Tor 上取得较好攻击效果. 在此基础上,Wang 等人[81]将特征集扩大,调整不同特征所占权重,使用 K 近邻作为分类器,降低了计算成本,攻击效果进一步提升. Hayes 等人[82]提出了一种基于随机决策森林的网站指纹识别技术,称为 k-指纹识别技术,该技术比以前的工作更大规模地评估了在标准网页以及 Tor 隐藏服务上的攻击性能,结果表明性能得到很大提升. Panchenko 等人[83]提出了一种新颖的网站指纹攻击,利用具有代表性的数据集探索了互联网规模的网站指纹识别的实际限制,并研究了攻击者可能遵循的不同策略的实际成功概率. Nasr 等人[84]将压缩感知技术应用于网站指纹识别,实验结果表明该技术能够保留最优的可压缩网络流量特征,在获得较好识别结果的同时降低了攻击开销. Rimmer 等人[85]提出了一种自动化的特征提取方法,并应用深度学习技术自动地对 Tor 流量进行去匿名化,达到较高的攻击准确率. Zhuo 等人[86]提出了一种基于隐马尔可夫模型的网站建模方法,实验结果表明该方法可以在 closed-world 环境下高效识别网页和网站. 然而,Web 站点指纹攻击的实施依赖于一系列假设,如用户浏览器关闭了缓存功能、用户浏览网页过程中较少出现背景流量等,消除这些假设对于方法的实用性具有重要的意义.

除此之外,还有部分针对 Freenet 的攻击技术,通过观测邻居节点流量特征判断其是否为文件请求的发起者,从而实现去匿名化攻击. 例如,Levine 等人[87]利用请求发起者的 HTL 默认初始值为 18 的特点以及相应的混淆机制,通过主动部署节点记录从其邻居节点接收到的 HTL 值等于 18 或 17 的请求,并利用朴素贝叶斯的方法判断邻居节点是否为该请求的发起者. 实验结果表明该方法的误报率仅为 2%.

3.1.4 主动单端流量分析

目前的单端主动攻击主要在 2 处位置实施:1)在出口节点与应用服务器间的未加密链路上注入恶意代码;2)控制用户的入口节点或链路并执行主动 Web 站点指纹攻击.

恶意代码注入是指在服务器端主动地将恶意代码插入用户流量中,以便代码到达被攻击主机并执行. 如果流量端到端进行加密,攻击者可实施中间人攻击[88-89],绕过匿名通信系统的客户端并直接建立直达恶意服务器的连接. 为此,攻击者需控制代理和远程服务器之间的未加密连接. 例如在 Tor 网络中,攻击者可以控制 Tor 出口节点并任意地注入或修改非加密流量的内容. 在控制了非加密连接之后,攻击者可以将各种软件实例注入到连接中,包括 Flash, Javascript, ActiveX 控件和 Java. 一旦这些软件在浏览器中执行,它们将绕过浏览器中的本地代理,直接创建到特定远程服务器的连接,这将会暴露客户端的真实 IP 地址. 为了防止遇到主动的单端攻击,用户应禁用插件,以避免恶意代码在浏览器中的执行. 或者用户可以在客户端部署透明代理,以确保所有流量都被引导到匿名通信系统中.

在被动的 Web 指纹攻击中,攻击者仅监听链路,并不会对流量进行主动调制. 由于 Tor 等匿名通信系统的定长封装机制和 HTTP 持久连接、流水线等技术的影响,Web 页面不同对象的数据在返回流量中出现重叠难以区分,导致指纹攻击的正确率无法进一步提升. 如果可以采取某些方法使不同对象的返回流量区分开来,则可以更好地为不同 Web 站点建立指纹. He 等人[90]首次提出并针对 Tor 系统进行了主动 Web 站点指纹攻击,通过对 Tor 流量进行观察确定用户开始发送请求报文的位置,然后主动延迟用户发出的请求报文,使前一个请求对象的响应数据有足够时间完成传输,从而达到分离不同对象流量的目的. 然而该攻击会造成报文重传,隐蔽性较差,并且延迟操作的粒度较粗,未对上行流量中存在的大量匿名协议控制报文进行识别. 针对上述问题,Yang 等人[91]进而设计了 Tor 流量控制报文和 Web 请求报文识别算法,通过主动延迟 HTTP

请求报文实现不同 Web 对象响应数据的分离,达到了较好的 Web 站点识别效果.

针对 Freenet,Baumeister 等人[92]利用 Freenet 中的邻居节点更新机制实施了一种路由表插入(routing table insertion, RTI)攻击,将恶意节点插入目标节点的路由表中,并通过模拟实验验证了该攻击的有效性. Tian 等人[93]基于 RTI 攻击提出了一种追踪文件下载节点的攻击,根据多个监测节点接收到的请求信息推断该请求的发起节点.

3.1.5 其他攻击技术

节点攻击是对匿名网络中目标主机发起的针对性攻击,通过利用协议设计缺陷,使目标节点系统服务失效或暴露目标主机的隐私信息.

在 Tor 网络中,攻击者可以利用 Tor 设计方面的不足来瘫痪任何一台中继节点,发现网络中的隐藏服务以及对用户进行去匿名化攻击. 当前对 Tor 网络影响最大的节点攻击方法是 Sniper 攻击[94],该攻击方法由 Jansen 等人在 2014 年提出. Sniper 攻击是一种针对 Tor 中继节点的拒绝服务攻击,可以用来瘫痪任意 Tor 中继节点,攻击成本非常低,但破坏性极强. Tor 设计的拥塞控制算法并不能避免网络拥塞,而是实现了端到端的流量控制. 利用该缺陷,攻击者可以杀死目标节点中的 Tor 进程,从而使其服务失效.

除此之外,I2P 网络中也存在类似的协议漏洞. Egger 等人[95]提出一种 I2P 节点上传并查询自身节点信息的漏洞,可以对 I2P 用户进行去匿名化攻击. I2P 节点使用 DatabaseStore 消息将自己的 RouterInfo 等直接上传到与其逻辑距离接近的恶意 Floodfill 节点上,并通过其探测隧道发送 DatabaseLookup 消息查询自身的 RouterInfo 等信息. 因此,在隧道有效时间内,恶意的 Floodfill 节点可以关联节点信息和探测隧道信息. 当节点使用探测隧道向共谋的 Floodfill 节点查询隐藏服务 LeaseSet 信息时,便可以关联节点和其访问的隐藏服务,从而完成对 I2P 节点的去匿名化攻击.

3.2 匿名通信增强技术

为了缓解各种攻击技术带来的威胁,目前已存在一些用来防御各种攻击的研究工作. 广义地讲,可以从网络层、协议层和应用层 3 个角度进行部署.

由于可以利用网络流量特性来对用户之间的通信进行去匿名化,所以阻止网络层攻击的基本思路是去除与用户相关联流量的特征,包括数据包大小分布、数据包顺序、流速率、流量时间等. 具体而言,包填充技术可以基于一定策略填充数据包大小,以便去除数据包长度相关特征[96]. Wright 等人[97]提出了流量整形技术,使数据包长度与目标页面流量数据包长度分布相似,可以有效防御 Liberatore 等人[77]提出的攻击方法. Chan-Tin 等人[98]提出了一种基于聚类的流量整形技术,使簇中的不同网站流量特征看起来基本相同,从而使分类精度从 70%下降到不足 1%.

流量填充技术可以将伪数据包注入到用户的原始流量中,以便模糊流量大小. 例如,为了阻止网站指纹攻击,网络服务器可以首先选择目标页面,然后模仿该目标网页的数据包大小分布. 为了应对 Panchenko 等人[79]的网站指纹攻击,Tor 项目[99-100]提出通过随机化 HTTP 流水线管道中的最大请求数来进行防御. 除此之外,Nithyanand 等人[101]指出,在 Tor 暗网中 40%的链路易受 AS 级别的流量关联攻击,为此他们提出了一种新的 Tor 客户端 Astoria,通过网络测量和路径预测来进行智能路由选择,降低来自 AS 级别的威胁. Juen 等人[102]通过路径预测技术来抵御来自 AS 和网络交换机的流量分析攻击,但该方法仍有提升空间.

在协议层,报文填充和流量填充技术可以隐藏与用户相关联的流量特征. secure shell(SSH),TLS 和 IPSec 应用这样的报文填充技术可以将明文与块密码边界对齐,从而在一定程度上模糊了包大小. 为了进一步提高安全性,可以选择随机数量的数据包进行填充[75]. 此外还可以使用协议级流量填充技术. 例如,Tor 通常不将填充数据单元的功能用于链路级填充,因为它会显著降低链路的性能,因此可以设计协议级报文填充和流量填充技术,以在一定程度上减少开销.

在应用层,可以利用 HTTP 特征和背景流量(即诱饵网页)从用户流中移除流量特征. 例如,HTTP 流水线技术和 HTTP range 字段可以用于调整传入和传出的数据包大小[103]. 此外,在客户端改变 HTTP 请求的顺序可以在一定程度上改变流量模式. 为了在应用层使用背景流量技术,当用户浏览目标网页时,可以在背景流量中加载诱饵网页. 但这种类型的防御技术只能用于一些特定的协议(例如 HTTP),不能广泛应用于所有协议. Wang 等人[104]提出了一种高效的网站指纹防御技术 Walkie-Talkie,通过修改浏览器以半双工模式进行通信,可以模拟突发序列,使敏感和非敏感页面的流量模式类似. 实验结果表明 Walkie-Talkie 能够以较低的带宽和时间开销防御网站指纹攻击.

Table 1 Research Works on Anonymous Communication Attack and Defense

表 1 匿名通信攻击和防御的研究

Research		Attack		Defense		
		Active	Passive	Application Layer	Protocol Layer	Network Layer
Network	End-to-end	Network Layer Watermarking[62-69] Protocol Layer Watermarking[70-71] Application Layer Watermarking[72-74]	AS/IX-level Attack[55-59] Local Monitoring Attack[60-61]	Background Traffic Techniques [103] WF Defense [104]	Dummy Traffic Techniques [75]	Packet Padding[96] Traffic Morphing [97-98] Intelligent Relay Selection[101-102]
	Single-ended	Inject Malicious Code Attack[88-89] Active WF Attack[90-91] RTI Attack[92-93]	WF Attack[75-86] Statistical Detection[87]			
Node	System	DoS Attack[94]				
	Protocol	Vulnerability Exploit Attack[95]				

总体而言,匿名通信攻击和匿名增强技术的研究是目前学术界关注的一个重要领域,表 1 总结了目前主要技术的分类、特征和相关文献. 这些工作或者从流量分析的角度,或者从匿名节点的角度指出相关系统存在的缺陷和面临的问题,并进而从应用层、协议层和网络层不同层次给出相应的解决方案,旨在提升这些系统的匿名性. 但实际上,这些攻击技术,尤其是去匿名化技术的提出,本质上也为暗网治理指明了可能的技术途径.

4 暗网治理技术研究

围绕暗网治理目标,目前已从网络本身和网络用户 2 个方面开展了较多的研究工作.

针对暗网服务的滥用问题,提出了对暗网 bridge 节点的发现方法[105-107]. 通过分析典型匿名通信系统 Tor 的节点选择算法,提出以节点注入方式枚举在线 bridge 节点的方法,并利用协议特征以区分客户端、正常匿名入口节点和 bridge 节点. 通过理论分析推算枚举所有节点的时间,并根据理论结果指导实际部署,以达到最优的资源配置和最快的 bridge 节点捕获速率. 此外,在获取部分 bridge 节点的基础上,通过对受控网络的监控,分析出、入流量中存在的 bridge 节点的网络连接特征,从而关联分析获取更多 bridge 节点.

针对匿名通信系统 Tor 的暗网隐藏服务节点,提出了隐藏服务定位技术[73,108],为暗网中的非法内容监管提供支持. 在分析 Tor 隐藏服务节点与客户端之间的匿名通信协议的基础上,部署部分 Tor 入口节点,通过控制客户端不断向目标隐藏服务节点发起连接,以产生具有可检测特征的隐蔽、快速流量,并在受控节点观察流量特征以定位和确认隐藏服务节点. 通过对暗网通信系统选路算法的理论分析,推算其隐藏服务节点选择受控节点的概率,以此指导节点部署的规模,以最小资源代价实现暗网隐藏服务节点的捕获.

针对滥用暗网的用户,提出暗网用户的上网行为分析技术,为暗网用户的监管提供支撑. 暗网用户流量经过加密、混淆等操作后,其识别难度大大增加,为此提出了暗网通信流量识别技术[109]. 针对典型暗网系统,分别从协议设计和实现机制入手,研究匿名通信系统中通信数据的缓存、封装和调度机制. 通过对数据包分布、流统计属性等不同层面特性的综合分析,筛选匿名通信流量的可区分特征,并从理论上分析证明特征选择的合理性. 在此基础上,研究并选取合适的机器识别方法,实现对匿名通信流量的快速、准确在线识别.

在识别暗网用户的基础上,为进一步推测暗网用户加密流量中的应用类型,提出了暗网通信应用分类技术[25]. 在对各种交互式和非交互式暗网通信应用流量进行深入分析的基础上,建立包括上下行流量比、并发连接数等指标的流统计模型,并将其作为先验知识. 针对采集的流量样本使用特征选择算法,获取不同匿名协议相应的对噪声不敏感、区分度高的特征. 在此基础上,选择抗干扰分类性能稳定的分类模型,对匿名通信流量上层应用进行分类. 建立分类评价模型,根据分类误差率、计算时间复杂度等指标,对所选特征和分类模型进行评价.

针对暗网用户的网站访问流量,文献[91,110]提出了暗网加密流量内容分析技术,以推测用户访问的站点.分别针对单跳和多跳匿名通信系统,研究其所采用的安全传输和匿名协议,并深入分析上层应用对暗网通信包长、时间间隔以及并发匿名链路数量等流量特征的影响.在此基础上,重点针对HTTP等典型暗网通信流量,通过被动和主动方式采集并预处理流量数据,提取特征生成指纹,从而建立所关注目标站点的指纹库.使用高效分类模型将未知暗网流量与指纹库中的数据进行匹配,分析可能的暗网通信目的端.

对于访问非法站点的用户,提出了不同层次的通信追踪技术[62,69-71,73,111-116],能够快速、准确且隐蔽地确认匿名流量之间的通信关系.通过分析匿名通信机制以及各种网络干扰对网络流量时间特征的影响,利用随机过程理论建立干扰环境下匿名通信流量时间特征变化的数学模型.在通用流水印追踪架构的基础上,根据水印追踪的隐蔽性和健壮性要求,利用信息论方法分析评估基于现有水印机制构建的隐蔽信道的容量.在此基础上,引入最优停止理论,研究并设计了自适应的水印嵌入与检测机制.

上述工作从网络本身和网络用户2个角度研究暗网治理相关技术,涵盖了隐藏节点发现和隐藏服务定位、暗网用户网络行为分析,以及暗网流量追踪和通信关系确认等各个方面,可为暗网的监管和治理提供有效的技术支撑.

5 总结与展望

作为网络安全领域的一项关键技术,多年来匿名通信一直是学术界研究的热点,而近几年频频曝光的涉及暗网的一系列安全事件,更是使匿名通信与暗网受到了社会各个层面极大的关注.为了更好地剖析暗网及作为其核心支撑的匿名通信技术,本文首先给出了匿名通信系统的基本分类,阐明匿名通信系统与暗网的关系,并介绍了典型匿名通信系统的基本工作原理,涵盖了Tor,I2P,Freenet和ZeroNet这4种主流暗网形态.其次,从匿名接入、匿名路由和暗网服务3个方面,阐述了匿名通信系统的关键技术及针对安全性和性能优化的相关研究工作.然后,从针对网络流量和针对节点攻击2个角度,总结了当前在匿名通信攻击和防御2个方面研究工作的现状,包括主动/被动-端到端/单端的流量分析、系统/协议漏洞利用等攻击手段,以及相应的防御措施.最后,本文还介绍了暗网治理方面的现有研究成果,包括暗网隐藏节点发现和隐藏服务定位、暗网用户的网络行为分析,以及暗网流量追踪和通信关系确认等.

随着人们对隐私保护和系统性能等方面需求的不断提高,需要对匿名通信技术进行进一步改进和优化,并研究综合应用mix技术和洋葱路由技术的新型匿名通信系统.在匿名接入方面,为规避流量监管,隐蔽、鲁棒的流量伪装技术仍然是当前研究的热点.此外,由于机器学习、深度学习等人工智能技术在匿名通信系统攻击中的广泛应用和效果提升,各种适用于低延迟系统的智能流量混淆、填充等匿名增强技术将成为关注的焦点,这些研究均以对抗人工智能技术为目标.

在另一方面,匿名通信系统和技术的不断完善,又给暗网治理带来了更大的挑战.本文认为暗网治理工作需要从暗网站点侧和用户侧2个方面开展:

1)在暗网站点侧.首先针对不同暗网架构进行分析,研究暗网站点发现及内容获取的相关技术.在此基础上,通过分析暗网站点内容的分布规律,可从公网和暗网2个方面大规模、自动化地收集网络内容,进而借助于大数据处理、知识图谱、人工智能等技术分析暗网数据,挖掘数据中的新物品、新术语、新事件等,为整个暗网空间的安全态势感知提供技术支撑.此外,针对不同匿名通信系统的暗网服务,需要深入分析其暗网服务技术的原理和匿名路由协议,设计有效的暗网服务定位方法,从而在发现非法暗网站点时,可有效、快速地开展对其服务站点的定位和治理工作.

2)在暗网用户侧.亟需研发相应的流量识别、用户节点发现、内容分析、追踪定位等技术,为暗网用户监管提供技术支持:①在流量识别方面,用户侧流量的监管与匿名接入技术的不断发展形成了“军备竞赛”,各种高级、智能的流量伪装等增强技术的应用,增加了对暗网侧用户流量的分析难度,亟待研发适用于大规模高速网络环境中高效、准确、易部署的暗网流量监管方案.②在用户节点发现方面,暗网通信节点发现技术可丰富暗网用户侧监管的手段,通过分析暗网中节点的种类和分布规律,设计有效、快速的暗网拓扑和节点发现方法,以快速发现访问暗网节点的用户.③在内容分析方面,加密暗网通信流量的内容分析仍然是热点的研究问题,可引入机器学习技术并通过对模型参数的优化,不断地提高对用户侧流量内容推测的准确率.在这过程中,实际

网络中存在的干扰和噪声数据以及防御技术的不断提升,对该技术的实际应用将带来诸多挑战.④在追踪定位方面,现有技术虽可实现对暗网用户与目标站点的通信关系关联,但依赖关键网络节点或暗网节点资源的部署.因此,如何降低对大规模节点资源的需求和依赖,研发易于大规模部署且适用于各类暗网应用的通信关系关联技术仍需要进一步的探索.

除了加大暗网技术的研发实现源头的治理,还需要借助于网络安全法、社会教育、多种侦查手段并举等措施.为此,可通过完善暗网行为在法律中具体规定、正确引导公众对暗网的认识、加强传统侦查方法和暗网分析技术的融合等方式,从多方面降低甚至消除暗网对社会和人们的负面影响.

致谢 本文在撰写过程中得到了丁翔、陈阳、许丹妮、徐永欢、党一菲等硕士研究生和王春棉、王姗博士研究生的帮助.本人的工作获得了江苏省网络与信息安全重点实验室、计算机网络和信息集成教育部重点实验室的资助,在此特别表示感谢!

参 考 文 献

[1] Tor Metrics Team. Onion services-Tor metrics [EB/OL]. [2018-10-01]. http://metrics.torproject.org/

[2] Chaum D L. Untraceable electronic mail, return addresses, and digital pseudonyms [J]. Communications of the ACM, 1981, 24(2): 84-88

[3] Danezis G. Mix-networks with restricted routes [C] //Proc of the 3rd Int Workshop on Privacy Enhancing Technologies. Berlin: Springer, 2003: 1-17

[4] Dingledine R, Freedman M J, Hopwood D, et al. A reputation system to increase MIX-Net reliability [C] //Proc of the 4th Int Workshop on Information Hiding. Berlin: Springer, 2001: 126-141

[5] Dingledine R, Syverson P. Reliable MIX cascade networks through reputation [C] //Proc of the 6th Int Financial Cryptography Conf. Berlin: Springer, 2002: 253-268

[6] Gulcu C, Tsudik G. Mixing E-mail with Babel [C] //Proc of Int Society Symp on Network and Distributed Systems Security. Rosten: The Internet Society, 1996: 2-16

[7] Moeller U, Cottrell L, Palfrader P, et al. Mixmaster protocol-version 2 [EB/OL]. [2018-09-01]. https://tools.ietf.org/pdf/draft-sassaman-mixmaster-03.pdf

[8] Danezis G, Dingledine R, Mathewson N. Mixminion: Design of a type Ⅲ anonymous remailer protocol [C] //Proc of the 24th IEEE Symp on Security and Privacy. Piscataway, NJ: IEEE, 2003: 2-15

[9] Goldberg I, Shostack A. Freedom network 1.0 architecture and protocols [EB/OL]. [2018-09-01]. https://adam.shostack.org/zeroknowledgewhitepapers/arch-tech.pdf

[10] McLachlan J, Tran A, Hopper N, et al. Scalable onion routing with Torsk [C] //Proc of the 16th ACM Conf on Computer and Communications Security. New York: ACM, 2009: 590-599

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

[12] Chen C, Asoni D E, Perrig A, et al. TARANET: Traffic-analysis resistant anonymity at the network layer [EB/OL]. [2018-09-01]. https://arxiv.org/pdf/1802.08415.pdf

[13] Kwon A, Lazar D, Devadas S, et al. Riffle: An efficient communication system with strong anonymity [J]. Proceedings on Privacy Enhancing Technologies, 2016, 2016(2): 115-134

[14] Piotrowska A M, Hayes J, Elahi T, et al. The loopix anonymity system [C] //Proc of the 26th USENIX Security Symp. Berkeley, CA: USENIX Association, 2017: 16-18

[15] Dingledine R, Mathewson N, Syverson P. Tor: The second-generation onion router [C] //Proc of the 13th USENIX Security Symp. Berkeley, CA: USENIX Association, 2004: 1-18

[16] The I2P Development Team. The Invisible Internet Project [EB/OL]. [2018-09-01]. https://geti2p.net/zh/

[17] The I2P Development Team. The Invisible Internet Project: Introduction to I2P [EB/OL]. [2018-09-01]. https://geti2p.net/zh/blog/page/

[18] The I2P Development Team. The Invisible Internet Project: The network database [EB/OL]. [2018-09-01]. http://www.i2pproject.net/zh/docs/how/network-database

[19] Maymounkov P, Mazieres D. Kademlia: A peer-to-peer information system based on the xor metric [C] //Proc of the 1st Int Workshop on Peer-to-Peer Systems. Berlin: Springer, 2002: 53-65

[20] Douceur J R. The sybil attack [C] //Proc of the 1st Int Workshop on Peer-to-Peer Systems. Berlin: Springer, 2002: 251-260

[21] Clarke I, Sandberg O, Wiley B, et al. Freenet: A distributed anonymous information storage and retrieval system [C] //Proc of Int Workshop on Designing Privacy Enhancing Technologies: Design Issues in Anonymity and Unobservability. Berlin: Springer, 2001: 46-66

[22] Clarke I, Miller S G, Hong T W, et al. Protecting free expression online with Freenet [J]. IEEE Internet Computing, 2002, 6(1): 40-49

[23] Clarke I, Sandberg O, Toseland M, et al. Private communication through a network of trusted connections: The dark Freenet [EB/OL]. [2018-09-01]. https://freenetproject.org/assets/papers/freenet-0.7.5-paper.pdf

[24] Smits R, Jain D, Pidcock S, et al. BridgeSPA: Improving Tor bridges with single packet authorization [C] //Proc of the 10th ACM Workshop on Privacy in the Electronic Society. New York: ACM, 2011: 93-102

[25] He G, Yang M, Luo J, et al. A novel application classification attack against Tor [J]. Concurrency & Computation Practice & Experience, 2016, 27(18): 5640-5661

[26] Kadianakis G. pluggable-transports/obfsproxy obfs2 [OL]. [2018-10-25]. https://gitweb. torproject. org/pluggable-transports/obfsproxy. git/tree/doc/obfs2/obfs2-protocol-spec. txt

[27] Kadianakis G. pluggable-transports/obfsproxy obfs3 [OL]. [2018-10-25] https://gitweb. torproject. org/pluggable-transports/obfsproxy. git/tree/doc/obfs3/obfs3-protocol-spec. txt

[28] Angel Y. obfs4/blob/master/doc/obfs4-spec. txt obfs4 [OL]. [2018-10-25]. https://github. com/Yawning/obfs4/blob/master/doc/obfs4-spec. txt

[29] Winter P, Pulls T, Fuss J. ScrambleSuit: A polymorphic network protocol to circumvent censorship [C] //Proc of the 12th ACM Workshop on Privacy in the Electronic Society. New York: ACM, 2013: 213-224

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

[31] Dyer K P, Coull S E, Ristenpart T, et al. Protocol misidentification made easy with format-transforming encryption [C] //Proc of ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2013: 61-72

[32] Moshchuk A, Gribble S D, Levy H M. Flashproxy: Transparently enabling rich Web content via remote execution [C] //Proc of the 6th Int Conf on Mobile Systems, Applications, and Services. New York: ACM, 2008: 81-93

[33] Snader R, Borisov N. A tune-up for Tor: Improving security and performance in the Tor network [C] //Proc of the 15th Annual Network and Distributed System Security Symp. Rosten: The Internet Society, 2008: 1-10

[34] Panchenko A, Lanze F, Engel T. Improving performance and anonymity in the Tor network [C] //Proc of the 31st Int Performance Computing and Communications Conf. Piscataway, NJ: IEEE, 2012: 1-10

[35] Rochet F, Pereira O. Waterfiling: Balancing the Tor network with maximum diversity [J]. Proceedings on Privacy Enhancing Technologies, 2017, 2017(2): 4-22

[36] Wang T, Bauer K, Forero C, et al. Congestion-aware path selection for Tor [C] //Proc of the 16th Int Conf on Financial Cryptography and Data Security. Berlin: Springer, 2012: 98-113

[37] Liu Z, Liu Y, Winter P, et al. TorPolice: Towards enforcing service-defined access policies for anonymous communication in the Tor network [C] //Proc of the 25th Int Conf on Network Protocols. Piscataway, NJ: IEEE, 2017: 1-10

[38] Tang C, Goldberg I. An improved algorithm for Tor circuit scheduling [C] //Proc of the 17th ACM Conf on Computer and Communications Security. New York: ACM, 2010: 329-339

[39] Alsabah M, Bauer K, Goldberg I. Enhancing Tor's performance using real-time traffic classification [C] //Proc of the 19th ACM Conf on Computer and Communications Security. New York: ACM, 2012: 73-84

[40] Alsabah M, Bauer K, Elahi T, et al. The path less travelled: Overcoming Tor's bottlenecks with traffic splitting [C] //Proc of the 13th Int Symp on Privacy Enhancing Technologies Symp. Berlin: Springer, 2013: 143-163

[41] Wacek C, Tan H, Bauer K S, et al. An empirical evaluation of relay selection in Tor [C] //Proc of the 20th Network and Distributed System Security Symp. Rosten: The Internet Society, 2013: 24-41

[42] Backes M, Kate A, Meiser S, et al. (Nothing else) MATor (s): Monitoring the anonymity of Tor's path selection [C] //Proc of ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2014: 513-524

[43] Backes M, Meiser S, Slowik M. Your choice MATor(s): Large-scale quantitative anonymity assessment of Tor path selection algorithms against structural attacks [J]. Proceedings on Privacy Enhancing Technologies, 2016, 2016(2): 40-60

[44] Matic S, Troncoso C, Caballero J. Dissecting Tor bridges: A security evaluation of their private and public infrastructures [C] //Proc of the 24th Network and Distributed Systems Security Symp. Rosten: The Internet Society, 2017: 1-15

[45] Alsabah M, Goldberg I. Performance and security improvements for Tor: A survey [J]. ACM Computing Surveys, 2016, 49(2): No. 32

[46] Schimmer L, et al. Peer profiling and selection in the I2P anonymous network [EB/OL]. [2009-01-10]. https://gnunet. org/sites/default/files/I2P-PET-CON-2009. 1. pdf

[47] Liu Peipeng, Wang Lihong, Shi Jinqiao, et al. Towards analysis of security in I2P's path selection [J]. Journal of Computer Research and Development, 2014, 51(7): 1555-1564 (in Chinese)
(刘培朋,王丽宏,时金桥,等. 匿名网络I2P路径选择的安全性分析[J]. 计算机研究与发展, 2014, 51(7): 1555-1564)

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

[49] Baumeister T, Dong Y F, Duan Z H, et al. Routing Table Insertion (RTI) attack on Freenet: Technical report [EB/OL]. [2018-09-01]. http://www.ee.hawaii.edu/~dong/traceback/RTI.pdf

[50] Øverlier L, Syverson P. Valet services: Improving hidden servers with a personal touch [C] //Proc of the 6th Int Conf on Privacy Enhancing Technologies. Berlin: Springer, 2006: 223-244

[51] Loesing K, Sandmann W, Wilms C, et al. Performance measurements and statistics of Tor hidden services [C] //Proc of Int Symp on Applications and the Internet. Piscataway, NJ: IEEE, 2008: 1-7

[52] Lenhard J, Loesing K, Wirtz G. Performance measurements of Tor hidden services in low-bandwidth access networks [C] //Proc of the 7th Int Conf on Applied Cryptography and Network Security. Berlin: Springer, 2009: 324-341

[53] Ian C, Matthew T, Oskar S, et al. Freenet Project [OL]. [2018-09-01]. https://freenetproject.org/pages/documentation.html

[54] Yang M, Luo J, Ling Z, et al. De-anonymizing and countermeasures in anonymous communication networks [J]. IEEE Communications Magazine, 2015, 53(4): 60-66

[55] Edman M, Syverson P. As-awareness in Tor path selection [C] //Proc of the 16th ACM Conf on Computer and Communications Security. New York: ACM, 2009: 380-389

[56] Akhoondi M, Yu C, Madhyastha H V. LASTor: A low-latency As-Aware Tor client [C] //Proc of the 33rd IEEE Symp on Security and Privacy. Piscataway, NJ: IEEE, 2012: 476-490

[57] Murdoch S J, Zieliński P. Sampled traffic analysis by Internet-exchange-level adversaries [C] //Proc of the 7th Int Workshop on Privacy Enhancing Technologies. Berlin: Springer, 2007: 167-183

[58] Johnson A, Wacek C, Jansen R, et al. Users get routed: Traffic correlation on Tor by realistic adversaries [C] //Proc of the 20th ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2013: 337-348

[59] Nasr M, Bahramali A, Houmansadr A. DeepCorr: Strong flow correlation attacks on Tor using deep learning [C] //Proc of the 25th ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2018: 1962-1976

[60] Bauer K, Mccoy D, Grunwald D, et al. Low-resource routing attacks against Tor [C] //Proc of ACM Workshop on Privacy in the Electronic Society. New York: ACM, 2007: 11-20

[61] Pappas V, Athanasopoulos E, Ioannidis S, et al. Compromising anonymity using packet spinning [C] //Proc of the 11th Int Conf on Information Security. Berlin: Springer, 2008: 161-174

[62] Ling Z, Luo J Z, Yu W, et al. A new cell-counting-based attack against Tor [J]. IEEE/ACM Transactions on Networking, 2012, 20(4): 1245-1261

[63] Yu W, Fu X, Graham S, et al. DSSS-Based flow marking technique for invisible traceback [C] //Proc of the 28th IEEE Symp on Security and Privacy. Piscataway, NJ: IEEE, 2007: 18-32

[64] Wang X, Chen S, Jajodia S. Network flow watermarking attack on low-latency anonymous communication systems [C] //Proc of the 28th IEEE Symp on Security and Privacy. Piscataway, NJ: IEEE, 2007: 116-130

[65] Wang X, Chen S, Jajodia S. Tracking anonymous peer-to-peer VoIP calls on the Internet [C] //Proc of the 12th ACM Conf on Computer and Communications Security. New York: ACM, 2005: 81-91

[66] Houmansadr A, Kiyavash N, Borisov N. RAINBOW: A robust and invisible non-blind watermark for network flows [C] //Proc of the 16th Network and Distributed Systems Security Symp. Rosten: The Internet Society, 2009: 222-228

[67] Houmansadr A, Borisov N. SWIRL: A scalable watermark to detect correlated network flows [C] //Proc of the 18th Network and Distributed Systems Security Symp. Rosten: The Internet Society, 2011: 17-24

[68] Rezaei F, Houmansadr A. TagIt: Tagging network flows using blind fingerprints [J]. Proceedings on Privacy Enhancing Technologies, 2017, 2017(4): 290-307

[69] Ling Z, Fu X, Jia W, et al. Novel packet size-based covert channel attacks against anonymizer [J]. IEEE Transactions on Computers, 2013, 62(12): 2411-2426

[70] Ling Z, Luo J, Yu W, et al. A new cell counter based attack against Tor [C] //Proc of the 16th ACM Conf on Computer and Communications Security. New York: ACM, 2009: 578-589

[71] Ling Z, Luo J, Yu W, et al. Protocol-level attacks against Tor [J]. Computer Networks the International Journal of Computer & Telecommunications Networking, 2013, 57(4): 869-886

[72] Ramsbrock D, Wang X, Jiang X. A first step towards live botmaster traceback [C] //Proc of the 11th Int Workshop on Recent Advances in Intrusion Detection. Berlin: Springer, 2008: 59-77

[73] Wang X, Luo J, Yang M, et al. A potential HTTP-based application-level attack against Tor [J]. Future Generation Computer Systems, 2011, 27(1): 67-77

[74] Chakravarty S, Barbera M V, Portokalidis G, et al. On the effectiveness of traffic analysis against anonymity networks using flow records [C] //Proc of the 15th Passive and Active Measurement Conf. Los Angeles: PAM, 2014: 247-257

[75] Dyer K P, Coull S E, Ristenpart T, et al. Peek-a-Boo, I still see you: Why efficient traffic analysis countermeasures fail [C] //Proc of the 33rd IEEE Symp on Security and Privacy. Piscataway, NJ: IEEE, 2012: 332-346

[76] Hintz A. Fingerprinting websites using traffic analysis [C] // Proc of the 2nd Int Workshop on Privacy Enhancing Technologies. Berlin: Springer, 2002: 171-178

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

[78] Lu L, Chang E C, Chan M C. Website fingerprinting and identification using ordered feature sequences [C] //Proc of the 15th European Conf on Research in Computer Security. Berlin: Springer, 2010: 199-214

[79] Panchenko A, Niessen L, Zinnen A, et al. Website fingerprinting in onion routing based anonymization networks [C] //Proc of the 10th ACM Workshop on Privacy in the Electronic Society. New York: ACM, 2011: 103-114

[80] Cai X, Zhang X C, Joshi B, et al. Touching from a distance: Website fingerprinting attacks and defenses [C] //Proc of the 19th ACM Conf on Computer and Communications Security. New York: ACM, 2012: 605-616

[81] Wang T, Cai X, Nithyanand R, et al. Effective attacks and provable defenses for website fingerprinting [C] //Proc of the 23rd USENIX Security Symp. Berkeley, CA: USENIX Association, 2014: 143-157

[82] Hayes J, Danezis G. *k*-fingerprinting: A robust scalable website fingerprinting technique [C] //Proc of the 25th USENIX Security Symp. Berkeley, CA: USENIX Association, 2016: 1187-1203

[83] Panchenko A, Lanze F, Zinnen A, et al. Website fingerprinting at Internet scale [C] //Proc of the 23rd Network and Distributed System Security Symp. Rosten: The Internet Society, 2016: 1-15

[84] Nasr M, Houmansadr A, Mazumdar A. Compressive traffic analysis: A new paradigm for scalable traffic analysis [C] // Proc of the 24th ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2017: 2053-2069

[85] Rimmer V, Preuveneers D, Juarez M, et al. Automated website fingerprinting through deep learning [C] //Proc of the 25th Network and Distributed Systems Security Symp. Rosten: The Internet Society, 2018: 1-16

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

[87] Levine B N, Liberatore M, Lynn B, et al. Statistical detection of downloaders in Freenet [C] //Proc of the 3rd IEEE Int Workshop on Privacy Engineering. Piscataway, NJ: IEEE, 2017: 25-32

[88] Winter P, Ensafi R, Loesing K, et al. Identifying and characterizing Sybils in the Tor network [C] //Proc of the 25th USENIX Security Symp. Berkeley, CA: USENIX Association, 2016: 1169-1185

[89] Winter P, Köwer R, Mulazzani M, et al. Spoiled onions: Exposing malicious Tor exit relays [C] //Proc of the 14th Int Symp on Privacy Enhancing Technologies Symp. Berlin: Springer, 2014: 304-331

[90] He G, Yang M, Gu X, et al. A novel active website fingerprinting attack against Tor anonymous system [C] // Proc of the 18th IEEE Int Conf on Computer Supported Cooperative Work in Design. Piscataway, NJ: IEEE, 2014: 112-117

[91] Yang M, Gu X, Ling Z, et al. An active de-anonymizing attack against Tor Web traffic [J]. Tsinghua Science and Technology, 2017, 22(6): 702-713

[92] Baumeister T, Dong Y, Duan Z, et al. A routing table insertion (RTI) attack on Freenet [C] //Proc of Int Conf on Cyber Security. Piscataway, NJ: IEEE, 2012: 8-15

[93] Tian G, Duan Z, Baumeister T, et al. A traceback attack on Freenet [C] //Proc of the 32nd IEEE Int Conf on Computer Communications. Piscataway, NJ: IEEE, 2013: 1797-1805

[94] Jansen R, Tschorsch F, Johnson A, et al. The sniper attack: Anonymously deanonymizing and disabling the Tor network [R]. Arlington, VA: Office of Naval Research, 2014

[95] Egger C, Schlumberger J, Kruegel C, et al. Practical attacks against the I2P network [C] //Proc of the 16th Int Workshop on Recent Advances in Intrusion Detection. Berlin: Springer, 2013: 432-451

[96] Cai X, Nithyanand R, Wang T, et al. A systematic approach to developing and evaluating website fingerprinting defenses [C] //Proc of the 21st ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2014: 227-238

[97] Wright C V, Coull S E, Monrose F. Traffic Morphing: An efficient defense against statistical traffic analysis [C] //Proc of the 16th Network and Distributed System Security Symp. Rosten: The Internet Society, 2009: 237-250

[98] Chan-Tin E, Kim T, Kim J. Website fingerprinting attack mitigation using traffic morphing [C] //Proc of the 38th IEEE Int Conf on Distributed Computing Systems. Piscataway, NJ: IEEE Computer Society, 2018: 1575-1578

[99] Perry M. Experimental defense for website traffic fingerprinting [OL]. [2018-12-19]. https://blog.torproject.org/blog/experimentaldefense-website-traffic-fingerprinting

[100] Perry M. A critique of website traffic fingerprinting attacks [OL]. [2018-12-19]. https://blog.torproject.org/critique-website-traffic-fingerprinting-attacks

[101] Nithyanand R, Starov O, Zair A, et al. Measuring and mitigating AS-level adversaries against Tor [J]. Computer Science, 2015: 1-13

[102] Juen J, Das A, Johnson A, et al. Defending Tor from network adversaries: A case study of network path prediction [J]. Proceedings on Privacy Enhancing Technologies, 2015, 2015(2): 171-187

[103] Luo X, Zhou P, Chan E W, et al. HTTPOS: Sealing information leaks with browser-side obfuscation of encrypted flows [C] //Proc of the 18th Network and Distributed System Security Symp. Rosten: The Internet Society, 2011: 1-20

[104] Wang T, Goldberg I. Walkie-Talkie: An efficient defense against passive website fingerprinting attacks [C] //Proc of the 26th USENIX Security Symp. Berkeley, CA: USENIX Association, 2017: 1375-1390

[105] Ling Z, Luo J Z, Yu W, et al. Tor bridge discovery: Extensive analysis and large-scale empirical evaluation [J]. IEEE Transactions on Parallel and Distributed Systems, 2015, 26(7): 1887-1899

[106] Yang M, Luo J, Zhang L, et al. How to block Tor's hidden bridges: Detecting methods and countermeasures [J]. Journal of Supercomputing, 2013, 66(3): 1285-1305

[107] Ling Z, Luo J, Yu W, et al. Extensive analysis and large-scale empirical evaluation of Tor bridge discovery [C] //Proc of the 31st IEEE Int Conf on Computer Communications. Piscataway, NJ: IEEE, 2012: 2381-2389

[108] Ling Z, Luo J, Wu K, et al. Protocol-level hidden server discovery [C] //Proc of IEEE Int Conf on Computer Communications. Piscataway, NJ: IEEE, 2013: 1043-1051

[109] He Gaofeng, Yang Ming, Luo Junzhou, et al. Online identification of Tor anonymous communication traffic [J]. 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[J]. 软件学报, 2013, 24(3): 540-556)

[110] Ling Z, Luo J, Zhang Y, et al. A novel network delay based side-channel attack: Modeling and defense [C] //Proc of the 33rd IEEE Int Conf on Computer Communications. Piscataway, NJ: IEEE, 2013: 2390-2398

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

[112] Wang X, Yang M, Luo J. A novel sequential watermark detection model for efficient traceback of secret network attack flows [J]. Journal of Network and Computer Applications, 2013, 36(6): 1660-1670

[113] Wang X, Luo J, Yang M. An interval centroid based spread spectrum watermark for tracing multiple network flows [C] //Proc of IEEE Int Conf on Systems, Man and Cybernetics. Piscataway, NJ: IEEE, 2009: 4000-4006

[114] Zhang Lu, Luo Junzhou, Yang Ming, et al. Interval centroid based flow watermarking technique for anonymous communication traceback [J]. 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[J]. 软件学报, 2011, 22(10): 2358-2371)

[115] Zhang Lu, Luo Junzhou, Yang Ming, et al. Synchronization in inter-packet delay based flow correlation techniques [J]. Journal of Computer Research and Development, 2011, 48(9): 1643-1651 (in Chinese)
(张璐, 罗军舟, 杨明, 等. 包间隔流关联技术中的同步问题研究[J]. 计算机研究与发展, 2011, 48(9): 1643-1651)

[116] Ling Z, Fu X, Jia W, et al. A novel packet size based covert channel attack against anonymizer [C] //Proc of the 30th IEEE Int Conf on Computer Communications. Piscataway, NJ: IEEE, 2011: 186-190

Luo Junzhou, born in 1960. PhD, professor, PhD supervisor. Member of IEEE and ACM, as well as co-chair of IEEE SMC Technical Committee on Computer Supported Cooperative Work in Design, and chair of ACM SIGCOMM China. His research interests include next generation network, protocol engineering, network security, cloud computing, and wireless LAN.

Yang Ming, born in 1979. PhD, professor, PhD supervisor. Member of CCF and ACM, as well as Deputy Director of Key Laboratory of Computer Network and Information Integration, Ministry of Education. His main research interests include network security and privacy.

Ling Zhen, born in 1982. PhD, associate professor. Member of IEEE, ACM and CCF. His main research interests include smartphone security, network security, privacy and forensics.

Wu Wenjia, born in 1983. PhD, associate professor. Member of CCF and ACM. His main research interests include computer networks.

Gu Xiaodan, born in 1987. PhD, lecturer. Her main research interests include network security and privacy.