匿名通信综述*

陈　欢[1],苏马婧[1],王学宾[2],宋　栋[1]
(1.华北计算机系统工程研究所,北京 100083;2.中国科学院信息工程研究所,北京 100083)

摘　要:匿名通信系统是指建立在网络应用层之上,结合数据转发、内容加密、流量混淆等一系列技术,实现通信实体之间关联关系和通信内容对第三方隐藏的网络。在商业秘密传输、电子投票等使用场景中,保证用户的个人身份、行为不被网络窃听者所识别是重要的评价标准,匿名通信系统正是为了解决上述问题。匿名通信系统具有消息不可溯源、通信无法被第三方监测等特点。一方面,从数据隐蔽传输的角度,匿名通信系统能够增强系统的安全性,另一方面,由于匿名通信技术的滥用,非法构建的私有隐蔽服务花样繁多、层出不穷,给网络安全和社会安定带来了巨大危害,因此,对目前匿名通信系统和技术进行梳理,有助于了解匿名通信技术发展现状,进而更为合理地利用匿名通信技术,构建匿名通信系统,实现信息安全可靠传输。

关键词:匿名通信系统;发展沿革;原理;威胁;评价
中图分类号:TN918　　**文献标识码**:A　　DOI:10.16157/j.issn.0258-7998.200995

中文引用格式:陈欢,苏马婧,王学宾,等. 匿名通信综述[J].电子技术应用,2021,47(4):46-53,58.
英文引用格式:Chen Huan,Su Majing,Wang Xuebin,et al. Survey of anonymous communication system[J]. Application of Electronic Technique,2021,47(4):46-53,58.

Survey of anonymous communication system

Chen Huan[1],Su Majing[1],Wang Xuebin[2],Song Dong[1]
(1.National Computer System Engineering Research Institute of China,Beijing 100083,China;
2.Institute of Information Engineering,Chinese Academy of Sciences,China)

Abstract:Anonymous communication system refers to a network that is built on the network application layer and combines data forwarding, content encryption, traffic obfuscation and other technologies to realize the association between communication entities and the hidden content of the communication content from third parties. In the usage scenarios of business secret transmission and electronic voting, it is an important evaluation standard to ensure that the user's personal identity and behavior are not recognized by eavesdroppers. Anonymous communication system is to solve the above problems. Anonymous communication system has the characteristics of untraceable message and communication cannot be monitored by a third party. On the one hand, from the perspective of hidden data transmission, anonymous communication systems can enhance the security of the system. On the other hand, due to the abuse of anonymous communication technology, illegally constructed private hidden services are numerous and endless, bringing network security and social stability.Sorting out the current anonymous communication systems and technologies is helpful to understand the development status of anonymous communication technologies, and then use anonymous communication technologies more reasonably to build anonymous communication systems to achieve safe and reliable transmission of information.

Key words:anonymous communication system;evolution;principle;threat;evaluation

0 引言

匿名即通过一些方法实现隐蔽个人身份或者是个人特征信息。匿名作为网络中一种特性,在现实生活中具有广泛而实际的需求。利用现代加密方法,如公私钥加密、电子签名和密钥协商等能够很好地保障传统意义上的信息安全四要素:机密性、完整性、可用性和真实性。但是,面对用户日益增加的对个人隐私保护的需求,传统的密码技术并不能有效解决用户诸如网络通信地址的匿名性问题,匿名通信系统的出现很好地解决了上述问题。互联网中的匿名是指,将通信双方的身份、网络地址等敏感信息通过一定的技术手段处理,使得网络中的恶意第三方无法有效识别通信双方的身份和网络特征信息。匿名通信系统,是指在现有的网络上采取诸如网络编码、通信加密、路由转发和链路混淆等多种技术,实现隐藏通信实体之间关系和通信内容的通信系统。

在匿名通信系统快速发展的同时,针对匿名通信的

*基金项目:国防基础科研计划项目(JCKY2019211B001)

威胁也层出不穷，根据攻击原理可具体分为两个大类：流量攻击和基于匿名通信协议本身弱点的攻击技术。前者主要是主动和被动攻击等，后者有网桥发现和重放攻击等。

此外，由于匿名通信可以实现自身信息隐藏，防止安全部门的追踪管理，从而逃避犯罪，因此如何确保匿名通信系统的可信性，防止其被恶意使用者滥用，也是当下互联网监管部门的重点关注方向。

目前已经有一些研究人员对匿名通信系统进行了大量研究[1-3]。鉴于此，为了深入理解匿名通信技术的功能结构、技术原理、部署机制、安全威胁等一系列问题，对其研究方向进行总体把握，对匿名通信系统进行综述具有重要意义。

本文首先阐述了匿名相关概念，给出了匿名通信系统定义及其含义，按照时间发展顺序系统地梳理了匿名通信系统发展至今的历程，并按照匿名通信系统在网络中工作的不同层次对其分类介绍，并进一步分析了匿名通信系统的工作原理、面临的威胁挑战以及应对方法；最后对匿名通信系统未来的发展进行了展望。

1 匿名通信系统定义及演化过程

1.1 匿名通信系统定义

文献[1]中定义匿名通信为：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。进一步地，将进行数据转发的由多跳加密代理节点组成的系统称之为匿名通信系统。

文献[2]中定义匿名通信系统为：一种建立在应用层之上，结合利用数据转发、内容加密、流量混淆等多种隐私保护技术来隐藏通信实体关系和内容的覆盖网络。

本文对匿名通信系统定义为：一种运行在应用层或者是网络层，综合运用密码、路径加密、流量混淆和链路控制等一系列技术，实现链路中通信双方身份和数据等隐私信息不被第三方恶意攻击者所获悉的隐蔽通信系统。匿名通信系统应当有效地保障通信实体双方的身份等特征信息在有限的实体集中不可识别；在多次的通信过程中实体没有固定的身份信息；执行具体操作的实体和执行动作之间不具备关联性，达到抵御网络中恶意攻击攻击者基于观测到的有限信息从发送和接收两个事件集合中分辨特定事件企图的作用。

1.2 匿名通信系统发展历史

匿名通信系统起源于20世纪80年代，历经40年演化，图1展示了匿名通信系统发展历程中的重要事件节点，期间出现了以Mix、Tor、LAP和Hornet等为代表的多种类型的系统。

1981年CHAUM D L提出Mix[4]，Mix是一种由用户节点和提供转发服务的多个Mix节点组成的链路混合网络，实现了最初的简单的匿名。匿名通信系统在此之后也得以开始被广泛研究发展并进一步投入到互联网隐私通信使用之中。

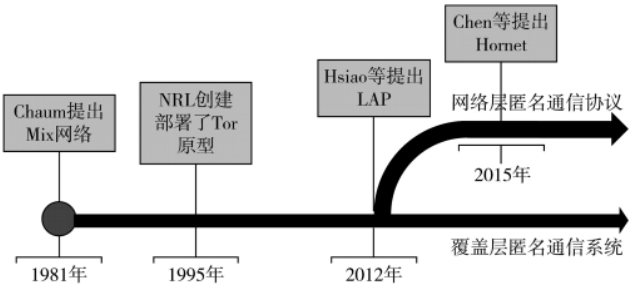

图1 匿名通信系统发展历程

1995年，美国海军实验室(NRL)的David Goldschlag、Mike Reed和Paul Syverson创造和部署了Tor(洋葱路由)的原型，用于防止在网络建立时泄露通信双方的信息。之后在1997年交由美国国防高等研究计划署进行进一步开发。

1996年，CULCU C等提出的Babel[5]引入了Mix路径，通过延迟批量消息来抵御流量分析攻击。

1998年由AT&T实验室提出的一种面向Web访问的一种匿名通信系——Crowds[6]。Crowds采用重路由机制，即通信数据通过多个节点组成的重路由路径到达接收方，以达到匿名的效果。

2001年DINGLEDINE R等提出基于信誉系统的Mix节点选择技术[7]，提高了Mix网络的可靠性和效率。

2002年DINGLEDINE R在之前的基础上又提出了具有分布式信任的Mix级联协议[8]，主要是通过信誉度重新排列Mix级联，进一步提高协议在具体使用过程中的可靠性。同年，Roger Dingledine开始参与Paul Syverson在NRL的洋葱路由项目工作。为了区分NRL的这个原始的工作和在其他地方出现的洋葱路由工作，Roger把这个项目叫做Tor，Tor之后经历了三代的发展已成为目前网络上使用最为广泛的匿名通信系统。2002年FREEDMAN M J等首次提到一种旨在消除流量分析攻击的对等匿名通信系统——Tarzan[9]。Tarzan融合了Mix思想和P2P技术，使用Chord查找算法选择中间节点并建立IP隧道。

2003年DANEZIS G提出结合“Mix级联模型”和“自由路由”的受限路由Mix网络[10]，以保证消息的匿名性。COTTRELL L等人提出Mixmaster协议[11]，在每个数据包末尾添加随机数据，将消息转化加密为统一大小，保证所有发送者的路由信息相同。DANEZIS G等人提出Mixminion协议[12]，部署了一组冗余和同步的目录服务器系统，为电子邮件消息提供发送人和接收人匿名。

之后的十余年，这种基于覆盖层的匿名通信系统的发展一直处于在原有系统的基础上进行改良和增补。例如通过加入可信计算[13]、网络编码[14]、SDN网络[15]等方式来抵御花样繁杂的攻击和监管。但是此类系统无一例外地面临着拓展性较弱和性能限制方面的问题。随着互

联网架构的进一步发展，研究人员得以利用下一代互联网架构来设计基于网络层的匿名通信协议，这种协议可以成为网络中默认路由架构的一部分，利用源选择的路由体系结构实现高度拓展且高效的匿名路由。网络层的匿名通信系统主要是借助新的互联网架构中的新型路由技术(pathlet routing/segment routing 等)，使得路由器等网络基础设施参与建立匿名通信系统，并协助转发匿名流量。与前述的覆盖层匿名通信系统间接隐藏分组头的匿名通信方法不同，网络层匿名通信系统在网络层直接隐藏信息的分组头，在理论上认为有更快的传输速度和更高的拓展性。

2012 年由 HSIAO H T 提出了 LAP[16]，在网络层实现的轻量级匿名通信协议，可以双向匿名通信，具有低延伸性和宽松的攻击者模型两个特点。之后的 Dovetail[17]是在 LAP 的基础上进行了改进，使用中间节点进行隐藏目的地。

2015 年 CHEN C 等提出的 Horne[18]是一种处于网络层的匿名通信协议。

2017 年 CHEN C 等提出的 PHI[19]解决了 LAP 和 Dovetail 中存在的问题，同时保持了两者的高效性。这种协议提出了一种隐藏路径信息的有效包头格式和一种新的可与当前及未来互联网架构兼容的后退路径建立方法。

2018 年 CHEN C 提出了 TARANET[20]，一种可拓展的、高速率且抗流量分析的匿名通信协议。TARANET 可以直接内置于现有的网络基础架构之中，可以实现短路经和高吞吐量。

综上，研究人员不断提出改进匿名通信体系结构，迭代系统版本，提升系统性能和匿名性。同时，针对匿名网络的攻击和对抗攻击的方法也在不断提升。

2 匿名通信系统工作原理及评价

近年来发展的新型匿名通信协议运行于网络层，以 LAP 为代表的轻量级匿名通信协议主要是通过隐藏通信过程中数据包包头内的转发信息来有效地实现防止恶意攻击者，这种轻型的匿名通信协议避免了对数据包有效载荷进行加密，有效地降低了通信对资源的消耗。以 Hornet 为代表的网络层洋葱路由协议在轻量级匿名通信协议的基础上提供了对数据包载荷的保护，其原理是一种由数据包来携带加密状态的洋葱路由协议的数据包格式，即，将通信的会话状态卸载到终端主机，并将它们自己的状态嵌入数据包以便中间节点在数据包转发过程中可以提取自己的状态进行转发操作。

2.1 典型匿名通信系统

按照通信系统构建所依赖的环境和运行机制，可以将匿名通信系统分为覆盖层匿名通信系统和网络层匿名通信协议两大类，见图 2。其中覆盖层的匿名通信系统以 Mix 和 Tor 为主要代表，基于网络层的匿名通信协议以 LAP 和 Hronet 为代表。覆盖层匿名通信系统和网络层匿名通信协议在可拓展性、时延、吞吐量、安全和匿名性以及部署规模等方面的对比如表 1 所示。下面分别对这几种典型的匿名通信系统进行介绍。

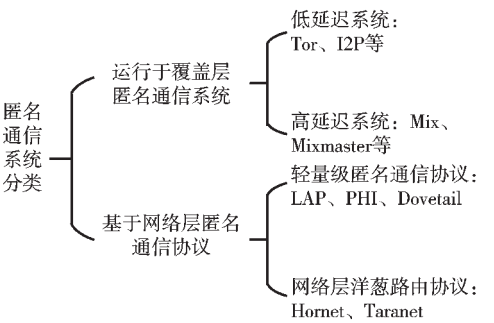

图 2 匿名通信系统分类图

表 1 两种不同类型匿名通信系统比较

类别	可拓展性	吞吐量	延迟	安全性	部署情况
覆盖层匿名通信系统	强	低	高	强	全球大规模部署
网络层匿名通信协议	弱	高	地	弱	未大规模部署

2.1.1 Mix

Mix 是一种利用多个节点进行多级路径转发消息的通信网络，该网络可以对通信双方身份和通信消息实现一定程度的保护。在实际使用过程中，发送者可以选择 N 个连续的目标进行数据传输，其中只有一个是真正的接收者，如图 3 所示。网络窃听者在一段链路中获取真正接收者的概率为 $1/N$，并且在实际传输过程中中间节点可以采用重新排序、延迟或者是填充手段使得窃听者成功的概率更低。

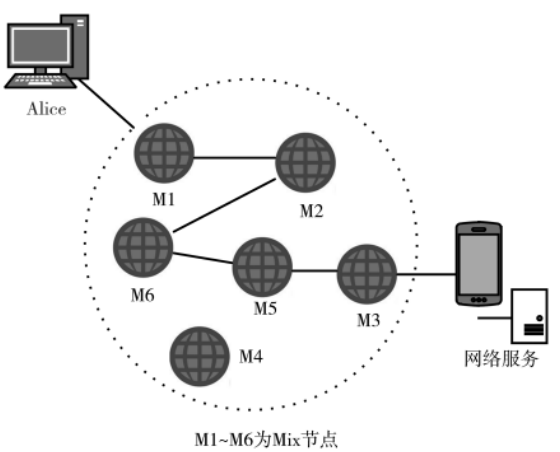

图 3 Mix 网络结构图

实现上述的匿名通信需要满足以下两个条件：(1)网络中各目标可以可靠地完成工作并且彼此之间有安全通道；(2)中间节点知晓所有的路径。

Mix 利用多个转发节点的方法来实现对信息准确传输路径的模糊，从而达到使得网络窃听者无法轻易获知

通信双发的关联信息的目的，实现一定程度的匿名性。

Mix 网络在实际的测试和使用过程中存在一些问题，基于 RSA 的混合消息格式不但会占用大量资源，同时也被证明该机制在标记攻击中存在脆弱性[21]。

2.1.2 洋葱路由 Tor

(1)第一代洋葱路由

将 NRL 最早部署的 Tor 称之为第一代洋葱路由。该系统通过多次混淆消息、层层加解密达到隐蔽网络路径结构，对抗路径跟踪和流量分析等攻击行为。其工作原理如图 4 所示。

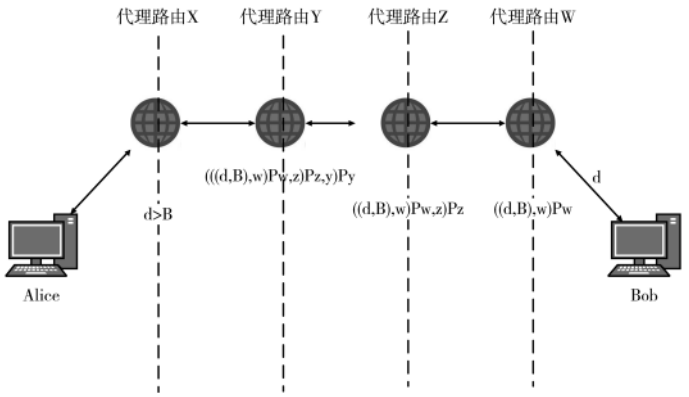

图 4 洋葱路由示意图

该系统也存在着一些技术上的缺陷：①代理路由器的潜在威胁。在代理路由器被攻陷的情况下，传输层的 IP 和端口信息将暴露。②加解密带来时延和额外开销。通信链路路由之间数据采用公钥密码机制加解密，当通信路径过长时，会导致中间节点路由加解密资源消耗过大，造成整个路径通信时延过长。③节点数据库信息冗余。所有的节点必须存储公私钥、认证码和可信标识等信息在数据库中，当节点数量过大时，数据库的维护和管理将消耗更多资源，导致系统拓展性较差。

(2)第二代洋葱路由

第二代洋葱路由使用了实时混合技术，即通过通信链路上的路由器对一定时间内的通信数据消息进行重新排序，以此使系统抵抗被动流量攻击。同时第二代洋葱路由系统还通过允许客户端加入且仅服务于选定的客户端来提高系统的灵活性。

(3)第三代洋葱路由

第三代洋葱路由，即 Tor，是目前使用最为广泛的匿名通信系统，在全球范围内大规模部署。相比于前代洋葱路由，Tor 做了以下几点改进：

①Tor 放弃了实时混合和重新对分组数据进行排序的操作。这些机制在后来的试验和实际使用中证明并不能有效地抵御流量分析攻击，并且会带来高昂的延迟和带宽开销。

②Tor 采用了目录服务器来存储链路相关信息，维护网络拓扑和路由证书。使得通信发起者可以根据自己需要独立从目录服务器中选择通信用到的节点。

③Tor 客户端以迭代方式设置通信链路，该机制使用 Diffe-Hellman 密钥交换机制与链路上的洋葱路由进行对称密钥的协商。

Tor 将数据打包成 512 B 大小的单元，每个单元在节点传输之前都要按照之前协商的密钥进行加密，依次是：出口节点、中间节点和入口节点。在消息传输时，各个中间节点使用自己的密钥将单元解密，得到下一跳信息后传给下一个节点，直到最终节点进行消息还原并传给接收者。

在数据的传输过程中，每个路由仅知道与其通信的前驱和后继路由器信息，通过网络传输链路上的有限信息实现匿名。在网络链路中的每一跳，根据信息流的方向逐层对信息进行加密或者是解密。中间节点不能够读取到单元的具体内容，单元的外观总是在经过中继节点时不断变化，使得不能通过网络信息流观察到通信双方之间的联系。

2.1.3 LAP 及其改进

LAP 是最早出现的网络层轻量级匿名通信协议，具有较低的延伸匿名和较为宽松的攻击模型。

LAP 实现了对匿名消息的发起者信息匿名和地理位置的保护，对于消息接收者，协议使用约会节点的方式使其匿名性得以保护。LAP 与目前的网络相互适应，同时也可与未来网络架构兼容。

LAP 网络拓扑模型如图 5 所示。

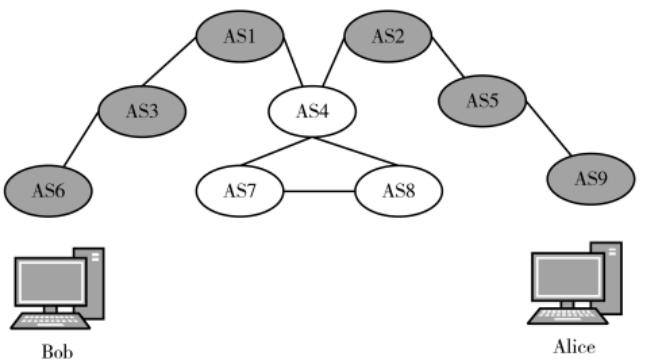

图 5 LAP 网络拓扑模型

LAP 协议的机制是通过模糊终端主机在网络中的拓扑位置来增强系统的匿名性。该协议中每个数据包都包含了加密的数据包包头，包头中包含了路由信息。

假设通信双方分别是 Bob 和 Alice。通信连接的建立始于 Bob 向 Alice 发送的空数据包。空数据包中包含了 Alice 的地址信息，网络中的每个 AS(Autonomous System)都知道目的地址，从而每个 AS 都可以独立自主的决定转发数据包的路径。

在 LAP 协议中，每个 AS 都持有本地密钥。在图 5 所示的传输过程中，位于第一跳的 AS6 首先使用自己的本地密钥 K6，对转发消息(Bob 建立链接使用的空数据包，即 AS 对的路由决策)进行加密，之后 AS6 会将加密的路径信息添加到数据包包头中，路径上的其他 AS 也会执

行相同的操作，具体过程见图 6。AS9 完成操作后，所有的加密信息都包含在数据包包头之中。在正式通信的过程中，每个 AS 只需用自己的密钥进行解密即可得到转发信息来转发数据包。

Dovetail 是在 LAP 基础上改进的一种协议。LAP 存在单一节点同时知晓源和目的地址，会产生匿名通信信息泄露的风险。Dovetail 采用间接节点隐藏目的地的机制，具体是 Dovetail 会随机选择三个第三方节点作为通信网络的辅助节点。

PHI 在前述的 LAP 和 Dovetail 的基础上进一步增强了协议的匿名性。其使用了三种新的技术：

(1)将节点状态以伪随机的顺序放置在数据包头中，实现节点位置信息隐藏。具体见图 7。

(2)使用 back-off 路径构造方法。

(3)加密绑定有效负载以防止会话劫持攻击。

PHI 在同等级别的资源开销下具有比 Dovetail 更强的匿名性。并且 PHI 可以与传统网络体系结构兼容。同时，PHI 克服了现有已识别的轻量级匿名通信系统攻击。

2.1.4 Hornet

Hornet 是一种利用下一代 Internet 体系结构设计的高效可拓展匿名通信系统。其设计初衷是以尽可能少的资源开销完成可靠的、高速率的匿名通信，并且能够向下兼容底层网络，具有很强的拓展性。相比于轻量级匿名通信协议，以 Hornet 为代表的匿名通信协议在之前的基础上增加了对数据包载荷的保护。其原理是一种由数据包来携带加密状态的洋葱路由协议的数据包格式，即，将通信的会话状态卸载到终端主机，并将它们自己的状态嵌入数据包以便中间节点在数据包转发过程中可以提取自己的状态进行转发操作。

2.2 匿名通信系统评价

匿名通信系统繁杂多样，对于各种匿名通信系统或协议在同一尺度标准进行分析评价具有重要意义，不仅利于使用者增强通信隐私保护，还利于匿名通信系统更加规范化的发展。研究人员提出了一系列针对匿名通信系统的评价指标和评价方法。

PFITZMANN A 等在文献[22]中提出了不可观测性概念。此后在各种相关文献中常以不可观测性作为匿名通信系统评价的重要指标。

REITER M K 等在文献[23]中提出采用 $1-p$ 来度量匿名通信系统匿名度。其中 p 表示攻击者可以从匿名集合中识别当前通信用户的概率。因此 $1-p$ 可以作为衡量同一匿名集合中不同对象之间可被攻击者识别的程度。

DIAZ C 等在文献[24]中提出了一种匿名通信系统匿名性的评价方法。该方法基于香农熵，比较全面地考虑了匿名集合大小以及攻击者对匿名集中不同元素之间的可识别状态的概率分布，因此在攻击者获取匿名集合中相关信息时能够有效地评估新状态下的匿名性。

HAMEL A 等人在文献[25]中从一种全新的角度衡量匿名系统性能。通过对攻击者能力进行分析，将之与系统带宽进行等价代换。通过攻击程度与带宽的转换，

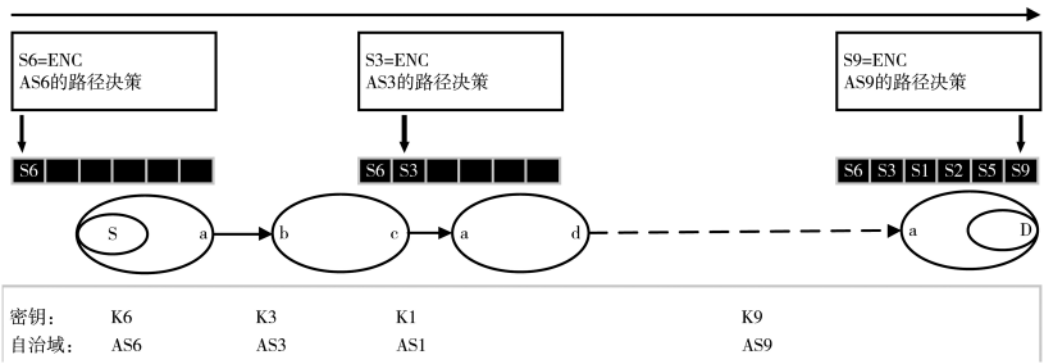

图 6 包头中加密路径形成过程

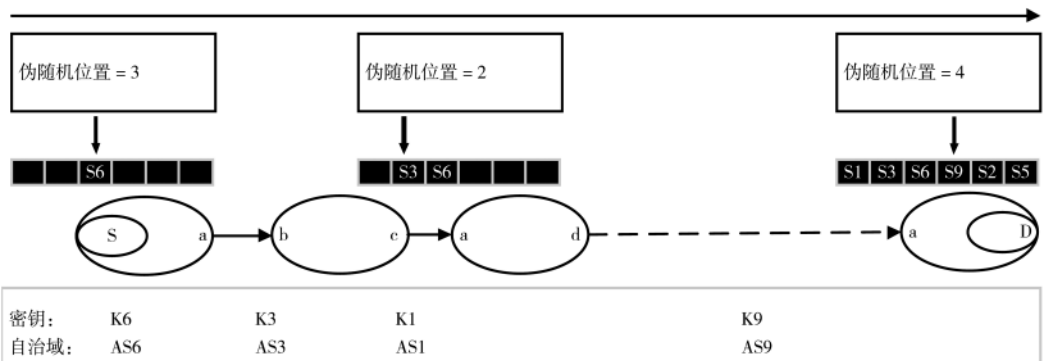

图 7 PHI 中段位置随机化过程

可以在一定带宽资源的统一前提下对攻击给匿名通信系统产生的影响进行评估。该方法侧重于关心匿名通信系统的性能。

谭庆丰等在文献[26]中提出针对匿名通信系统全方位不可观测性评价,并提出基于相对熵的方法对不可观测性进行度量。文中给出不可观测性定义:

(1)匿名性。即通信系统主体在匿名集合中的不可识别的状态。

(2)不可检测性。即从网络窃听者或者是攻击者的角度不能有效识别其感兴趣的通信主体。

(3)不可观测性。即指网络攻击者感兴趣的通信客体在任何其他相同类型的通信客体集合中不可区分的状态。其主要有两层含义:通信主体的匿名性和通信客体的不可检测性。

以 d 表示系统的不可观测性,d 的取值范围在[0,1],取值越小表明不可观测性属性越好。

本文认为需要从两个方面对匿名通信系统进行评价。一方面是系统的性能特性,这包括系统的兼容性、可拓展性、通信时延和通信效率等,这些特性是影响系统能够大规模部署和广泛使用的一个重要因素;另一方面是系统的安全特性,主要包括通信双方的不可观测性和通信内容的不可还原性,这些特性是系统匿名性的重要保障因素。以上两方面紧密结合才能形成一个性能优良、安全可靠的匿名通信系统。因使用场景的不同,对两者的实际要求会存在一定程度的偏向,因此对匿名通信系统的评价应该在实际使用场景和两者之间进行合理平衡。

3 匿名通信系统威胁与挑战

对匿名通信系统的攻击主要是通过一系列技术手段,达到发现通信流量、关联通信链路、关联通信双方具体身份进而还原通信内容的目的。总结起来,对匿名通信系统典型的攻击手段主要有流量分析攻击和监听,下面将分别对两类攻击方法进行介绍。

3.1 流量分析攻击

流量分析攻击是指攻击者控制匿名通信系统(网络)的入口和出口节点,对通信双方进行流量识别,运用统计及相关性方法关联通信实体、识别匿名通信用户的身份等信息。流量分析攻击可分为两种:被动式及主动式。被动流量分析攻击的攻击者先会从一端网络找出一段流量的特征,然后在另一端网络查找该特征,主动流量分析攻击的攻击者会在一端网络依据特定模式修改数据包,然后在另一端查找符合该模式的数据包。

攻击者可以籍此把两端的流量联系起来,使其去匿名化。即使在数据包上加入定时噪声,也有攻击手段能够抗衡[27]。

3.1.1 被动流量分析攻击

(1)流量动态匹配

攻击者在两个观测点监测流量(包括监测同一节点的进出口流量),来尝试通过搜索匹配观测点之间的流量相似性来判断观测点观测到的数据是否属于同一流量[28-29]。

(2)模板攻击

攻击者通过匿名通信系统访问已知 Web 站点或是其他 Web 服务端,构建出访问这些 Web 的流量模式(模板)数据库。在窃听通信流量时,攻击者将观测到的流量和数据库中的模式进行对比,若匹配成功则可以以较大概率猜测出客户端访问的 Web 站点或者是 Web 服务端[30-31]。

(3)网络统计关联

这种攻击方式是攻击者监测网络不同部分的特征并将其与目标匿名流的特征进行比较。例如将目标双向流的 RTT 与测得的到大量网络地址的 RTT 进行比较,在目标流的 RTT 与被监测网络之一的 RTT 有强关联的情况下可以判断目标主机的可能网络位置。同样的,通过记录单向流(随时间变化)的吞吐量,之后再与各个网络位置的吞吐量进行比较,敌手可以对目标主机的位置进行最终定位。

3.1.2 主动流量攻击

主动流量攻击使用和被动流量分析攻击相似的技术,但主动流量攻击还涉及攻击者对目标的流量操纵。

(1)动态流量修改

修改动态流量,在数据包中添加水印(或标签),攻击者可以监测到流量流向,这种攻击称之为流水印。另一种相似的攻击称之为流指纹,攻击者将多位信息编码插入动态流量中,之后再将编码的信息在同一网络的另一个节点解码用以识别具体网络。

(2)拥塞攻击

动态流量修改的前提是攻击者需要控制尽可能多的节点,拥塞攻击与其类似。但是攻击者仅需操纵单个可控节点流量,造成网络中其他节点拥塞或是网络波动。之后观察这些网络异常是否影响到了目标,如是,则说明被监测目标的流量遍历了发生网络流量波动的节点。

(3)指纹识别

通过对网站指纹进行研究,可以发起网站指纹攻击,破解用户所访问的隐藏服务[32]。政府级别的强制监管可以和企业合作,要求 ISP 在传输过程中复制用户的流量动态并且通过安全通道将其转发,也可以将监控设备架设在主干网等位置,保证拥有更快的应对速度,再利用数据分析工具筛选流量,识别 Tor 用户。

3.2 监听

3.2.1 出口节点攻击

可以通过运行和监听 Tor 出口节点,截获电子邮箱账号的用户名和密码[33]。Tor 不能加密出口节点到目标服务器之间流量,所以导致任一的出口节点皆有能力截获通过该节点且没有经过 TLS 或 SSL 进行端到端加密的流量。通过对截获的流量进行分析,攻击者可以在实际数据和协议数据中找到源端的相关信息[34]。文献[35-36]提出一种威胁 Tor 通信网络的方法,作者宣称可以达到

解密通信的效果。

3.2.2 自治域系统监听

若客户至入口中继和出口中继至目标地址这两段网络路径在同一个自治域之中，则该自治域的管理者可以经过统计，获知入口路段和出口路段之间的关系，并且能推断出数据包的具体流向。

3.3 应对威胁的方法

3.3.1 流量伪装

流量伪装技术(例如协议混淆、流量变种等)是应对流量分析的常用对抗手段。主要思想是将一种流量的特征伪装成为另一种流量，以此降低基于流特征分析的准确性。在匿名通信系统中，通过多次转发和改变报文的样式消除报文之间的对应关系，为通信的发送者和接收者提供可靠地隐私保护。文献[37]中 WRIGHT C V 提出一种可以实时改变数据分组的凸优化方法，该方法可以将一种流量的分组大小分布伪装成另外一种流量的分组大小分布，经变换之后的流量可以有效地规避流量分器的识别。此外比较典型的是文献[38-40]提出的SkypeMorph、StepTorus 和 CensorSpoofer，分别将 Tor 流量伪装成 Skype 视频流量、HTTP 流量和基于 SIP 的 VOIP 协议。

3.3.2 Tor 传输层插件

Tor 为了应对安全威胁，先后部署了 obfs、Meek 和 FTE 等基于传输层的插件来支持协议混淆和协议伪装。

(1)obfs 混淆代理

为了有效抵抗深度包监测技术，Tor 的 obfs 由此而生。obfs 先后经历了四个版本，分别是 obfs、obfs2[41]、obfs3[42]和 obfs4[43]。obfs2 采用分组密码加密方式对通信数据进行加密，擦除 Tor 流量相关标识，有效实现混淆。obfs3 使用 Diffie-Hellman 协议交换确认通信双方的密钥，但是此种方法在双方密钥交换阶段缺乏对网桥身份的验证，存在中间人攻击风险。obfs4 利用 BridgeDB 实现基于网桥身份验证的密钥交换，客户端通过 BridgeDB 查询可用的节点，并获取其相关信息(IP 地址、节点 ID 和公钥信息)，三个同时验证成功，才能通过 obfs4 建立通信链接。此种方法，既能有效地混淆 Tor 流量，又可防止中间人攻击。

(2)Meek

Meek[44]是一种前置域匿名通道构建技术。Meek-client 把真正传输的 Tor 数据封装在 HTTP POST 载荷之中，将目标网桥地址写入 HTTP HOST 载荷。加密后的 HTTP HOST 相关内容无法被监管者发现。前置域服务器接收到数据之后，根据实际载荷字段信息将数据转发到网桥节点，在节点运行的 Meek-server 对 HTTP 报头进行处理后将封装好的 Tor 流量发送至下一个中继节点。Meek 现依赖于 Google、Amazon 和 Azure 等大型服务提供商的前置域名服务器。通过这种方法，造成 Tor 客户端在访问正常往网站的假象，从而规避针对 Tor 的流量监控。

(3)FTE

Format-Transforming Encryption[45]在 2013 年被提出。通过拓展传统的对称加密，将密文转换为指定的传输格式。依据用户输入的正则表达式，输出具有一定协议格式的数据流量。其中用户输入的正则表达式，可以从DPI 系统源码中提取亦或是通过应用层流量自动学习得到。依此，使得基于正则表达式的 DPI 技术会将处理后的流量误识别为用户选定的协议流量，实现规避审查。特别的，在 Tor 中，常使用 HTTP 正则表达式将 Tor 流量转换为 HTTP 协议，实现流量伪装。

4 结论

匿名通信技术作为互联网发展过程中对匿名性保证的关键技术，自其提出以来便一直是相关领域的研究热点，经历了从最简单的 Mix 网络到实用性较强的 Tor，从最初的覆盖层匿名通信系统逐渐发展出适应未来互联网架构的网络层匿名通信协议。这一方面得益于在目前互联网大环境下网络中的匿名性作为公民的一项基本权利愈加受到各方面的挑战，公众对个人隐私的保护越来越重视。另一方面得益于匿名通信在商业领域的广泛使用。

未来匿名通信的发展主要还是沿着两条主线发展，一种是公开化网络，即采用无中心化的 P2P 方式，是 Tor 思想的延续；另一种则是私有化部署和自建的匿名通信系统，主要目标是为了实现隐私保护或受保护目标的防护(例如用于采集濒危动物信息的数据传输等)。

参考文献

[1] 罗军舟，杨明，凌振，等.匿名通信与暗网研究综述[J].计算机研究与发展，2019，56(1)：103-130.

[2] 王良民，倪晓铃，赵蕙.网络层匿名通信协议综述[J].网络与信息安全学报，2020，6(1)：11-26.

[3] 吕博，廖勇，谢海永.Tor 匿名网络攻击技术综述[J].中国电子科学研究院学报，2017，12(1)：14-19.

[4] CHAUM D L.Untraceabl electronic mail，return addreses，and dital Pseudonyms.Communications of the ACM，1981，24(2)：84-88

[5] GULCU C，TSUDIK G.Mixing E-mail with Babel[C].Proceedings of Internet Society Symposium on Network and Distributed Syetems Security，1996.

[6] REITER M K，RUBIN A D.Crowds：anonymity for web transactions[J].ACM Transactions on Information and System Security，1998，1(1)：62-92.

[7] DINGLEDINE R，FREEDMAN M J，HOPWOOD D，et al.Areputation system to increase MIX-Net reliability[C].Proceedings of the 4th Int Workshop on Information Hiding，2001.

[8] DINGLEDINE R，SYVERSON P.Relibale MIX cascade networks throuhh reputation[C].Proceedings of the 6th Internet Financial Cryptography Conference，2002.

[9] FREEDMAN M J，MORRIS R.Tarzan：a peer-to-peer anonymizing network layer[C].Proceedings of the ACM

Conference on Computer and Communications Security(CCS), 2002.

[10] DANEZIS G.Mix-networks with restrictes routes[C]. Proceedings of the 3rd Int Workshop on Privacy Enhancing Technologies, 2003.

[11] MOELLER U, COTTRELL L, PALFRADER P, et al. Mixmaster prococol-version 2[EB/OL].[2020-10-12].https://tools.ietf.org/pdf/draft-sassaman-mixmaster-03.pdf.

[12] DANEZIS G, DINGLEDINE R, MATHEWSON N.Mixminion: design of a type III anonymous remailer protocol[C].2003 IEEE Symposium on Security and Privacy, 2003.

[13] 吴振强,周彦伟,乔子芮.一种可控可信的匿名通信方案[J].计算机学报,2010,33(9):1686-1702.

[14] 周彦伟,杨波,吴振强,等.基于网络编码的匿名通信模型[J].中国科学:信息科学,2014,44(12):1560-1579.

[15] 赵蕙,王良民.基于SDN节点淆乱机制的接收方不可追踪的混合匿名通道[J].通信学报,2019,40(10):55-66

[16] HSIAO H C, KIM T J, PERRIG A, et al.LAP: lightweight anonymity and privacy[J].IEEE Security & Privacy, 2012, 19:506-520.

[17] SANKEY J, WRIGHT M.Dovetail: stronger anonymity in next-generation internet routing[M].Lecture Notes in Computer Science, 2014.

[18] CHEN C, ASONI D E, BARRERA D, et al.HORNET: high- speed onion routing at the network layer[C].Proceedings of the ACM Conference on Computer and Communications Security(CCS), 2015.

[19] CHEN C, PERRIG A.PHI: path-hidden lightweight anonymity protocol at network layer[J].Nephron Clinical Practice 2017, 2017(1):100-117.

[20] CHEN C, ASONI D E, PERRIG A, et al.TARANET: traffic-analysis resistant anonymity at the network layer[J]. EuroS&P, 2018:137-152.

[21] PFITZMANN B, PFITZMANN A.How to break the direct RSA-implementation of mixes[C].Proceedings of the International Conference on the Theory and Applications of Cryptographic Techniques(EUROCRYPT), 1989.

[22] PFITZMANN A, HANSEN M.A terminology for talking about privacy by data minimization: anonymity, unlinkability, undetectability, unobservability, pesudonymity, and identity management[DB/OL].[2020-10-12].http://dud.inf.tu-dresden.de/literatur/Anon_Terminolog_v0.34.pdf.

[23] REITER M K, RUBIN A D.Crowds: anonymity for Wwb transactions[J].ACM Transactions on Information and System Security(TISSEC), 1998, 1(1):66-92.

[24] DIAZ C, SEYS S, CLAESSENS J, et al.Towards measuring anonymity[C].Proceedings of the 2ed International Conference on privacy Enhancing Technologies, 2003.

[25] HAMEL A, GREGOIRE J, GOLDBERG I.The misentropists: new approaches to measures in TOR.CACR 2015-18[R]. Waterloo.Ontario, Canada: University of Waterloo.2011.

[26] 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[J].计算机研究与发展,2015,52(10):2373-2381.

[27] RAMIN S, DENNIS G, DON T, et al.Towards provably inrisible network flow fingerprints[C].2017 51st Asilomar Conference on Signals, Systems, and Computers, 2017-11-27:258-262.arXiv:1711.10079.

[28] MITTAL P, KHURSHID A, JUEN J, et al.Stealthy traffic analysis of low-latency anonymous communication using throughput fingerprinting[C].Proceedings of the ACM Conference on Computer and Communications Security(CCS), 2011.

[29] MURDOCH S J, DANEZIS G.Low-cost traffic analysis of Tor[C].Proceedings of the IEEE Symposium on Security and Privacy(Oakland), 2005.

[30] Gong Xun, BORISOV N, KIYAVASH N, et al.Website detection using remote traffic analysis[C].Proceedings of the Privacy Enhancing Technologies Symposium(PETS), 2012.

[31] Wang Tao, Cai Xiang, NITHYANAND R, et al.Effective attacks and provable defenses for website fingerprinting[C]. Proceedings of the USENIX Security Symposium, 2014.

[32] KWON A, ALSABAH M, LAZAR D.Circuit fingerprinting attacks: passive deanonymization of tor hidden services[C]. USENIX Conference on Security Symposium.USENIX Association, 2015:287-302.

[33] ZETTER K.Rogue nodes turn Tor anonymizer into eavesdropper's paradise[M].Wired, 2007.

[34] ROBERT L.Tor hack proposed to catch criminals[DB/OL]. (2007-03-08)[2020-10-12]. https://www.securityfocus.com/news/11447/1.

[35] The Hacker News.Tor anonymizing network compromised by French researchers[DB/OL].(2011-10-24)[2020-10-12].https://thehackernews.com/2011/10/tor-anonymizing-network-compromised-by.html.

[36] Des chercheurs Francais cassent le reseau d'anonymisation Tor[DB/OL].(2011-10-17)[2020-10-12].https://bfm-business.bfmtv.com/01-business-forum/des-chercheurs-francais-cassent-le-reseau-danonymisation-tor-544024.html.

[37] WRIGHT C V, COULL S E, MONROSE F.Traffic morphing: an efficient defense against statistical traffic analysis[C]. NDSS, 2009.

[38] MOGHADDAM H, LI B, DERAHKSHANI M, et al.Skype-Morph; a acmouflage proxy for the Tor anonymity system[C]. Proceedings of the 19th ACM Conference on Computer and Communications Security.New York: ACM, 2012:97-108.

[39] WEINBERG Z, WANG J, YEGNESWARAN V, et al. StegoTorus: a camouflage proxy for the Tor anonymity

(下转第58页)

On the effectiveness of traffic analysis against anonymity networks using flow records[C].International Conference on Passive and Active Network Measurement,2014:247-257.
[6] JANSEN R,JUAREZ M,GALVEZ R,et al.Inside job:applying traffic analysis to measure Tor from within[C].NDSS,2018.
[7] HERRMANN D,WENDOLSKY R,FEDERRATH H,et al.Website fingerprinting:attacking popular privacy enhancing technologies with the multinomial naïve-bayes classifier[C].IEEE International Conference on Cloud Computing Technology and Science,2009:31-42.
[8] HAYES J,DANESIZ G.K-fingerprinting:a robust scalable website fingerprinting technique[C].Usenix Security Symposium,2016:1187-1203.
[9] PANCHENKO A,LANZE F,ZINNEN A,et al.Website fingerprinting at internet scale[C].Network and Distributed System Security Symposium,2016.
[10] NEPAL S,DAHAL S,SHIN S,et al.Deanonymizing schemes of hidden services in tor network:a survey[C].International Conference on Information Networking.IEEE,2015.
[11] BIRYUKOV A,PUSTOGAROV I,WEINMANN R P.Trawling for Tor hidden services:detection,measurement,deanonymization[C].2013 IEEE Symposium on Security and Privacy,2013:80-94.
[12] BIRYUKOV A,PUSTOGAROV I,THILL F,et al.Content and popularity analysis of Tor hidden services[C].2014 IEEE 34th International Conference on Distributed Computing Systems Workshops,2014:188-193.
[13] ÖZÇOMAK M S,KARTAL M,SENGER O,et al.Comparison of the powers of the kolmogorov-smirnov two-sample test and the mann-whitney test for different kurtosis and skewness coefficients using the monte carlo simulation method[J].Journal of Statal & Econometric Methods,2013,2(4):81-98.
[14] PEDREGOSA F,VAROQUAUX G,GRAMFORT A,et al.Scikit-learn:machine learning in python[J].Journal of Machine Learning Research,2011:2825-2830.
[15] CUNNINGHAM P,DELANY S J.K-nearest neighbour classifiers[Z].2007.
[16] PENG J,LEE K,INGERSOLL G.An introduction to logistic regression analysis and reporting[J].Journal of Educational Research,2012:3-14.
[17] KAVIANI P,DHOTRE S.Short survey on naive bayes algorithm[J].International Journal of Advance Research in Computer Science and Management,2017.
[18] Tian Yingjie,Shi Yong,Liu Xiaohui.Recent advances on support vector machines research[J].Technological and Economic Development of Economy,2012,18(1):5-33.
[19] ALI J,KHAN R,AHMAD N,et al.Random forests and decision trees[J].International Journal of Computer Science Issues,2012.
[20] CHICCO D,JURMAN G.The advantages of the Matthews correlation coefficient(MCC) over F1 score and accuracy in binary classification evaluation[J].BMC Genomics,2020.

(收稿日期:2020-07-16)

作者简介:
张玲(1976-),女,博士研究生,高级工程师,主要研究方向:网络安全、数据分析。
卫传征(1984-),男,硕士研究生,工程师,主要研究方向:网络安全。
段琳琳(1974-),女,博士研究生,讲师,主要研究方向:数据分析与信号处理。

(上接第 53 页)

system[C].Proceedings of the 19th ACM Conference on Computer and Communications Security,2012.
[40] WANG Q,GONG X,NGUYEN C T K,et al.CensorSpoofer:asymmetric communication using IP spoofling for the censorship-resistant Web browsing[C].Proceedings of the 19th ACM Conference on Computer and Communications Security,2012.
[41] KADIANAKIS G.Pluggable-transports/obfsproxy obfs2[DB/OL].[2020-10-12].https://gitweb.Torprojet.Org/pluggable-transports/obfsproxy.Git/tree/doc/obfs2/obfs2-protocol-spec.txt.
[42] KADIANAKIS G.Pluggable-transports/obfsproxy obfs2[DB/OL].[2020-10-12].https://gitweb.Torprojet.Org/pluggable-transports/obfsproxy.Git/tree/doc/obfs3/obfs3-protocol-spec.txt.
[43] ANGEL Y.Obfs4/blog/master/doc/obfs4-spec.txt obfs4[DB/OL].[2020-10-12].https://github.com/Yawning/obfs4/blog/master/doc/obfs4-spec.txt.
[44] FIFILED D,LAN C,HYNES R,et al.Block-resistant communication through domain fronting[C].Proceedings on Privacy Enhancing Technologies,2015.
[45] DYER K P,COULL S E,RISTENPART T,et al.Protocol misidentification made easy with format-transforming encryption[C].Proceedings of ACM SIGSAC Conference on Computer and Communications Security,2013.

(收稿日期:2020-10-12)

作者简介:
陈欢(1995-),男,硕士研究生,主要研究方向:匿名通信、暗网探测。
苏马婧(1985-),女,博士,正高级工程师,主要研究方向:网络空间测绘、网络安全。
王学宾(1986-),男,博士,工程师,主要研究方向:计算机网络、信息安全等。

版权声明