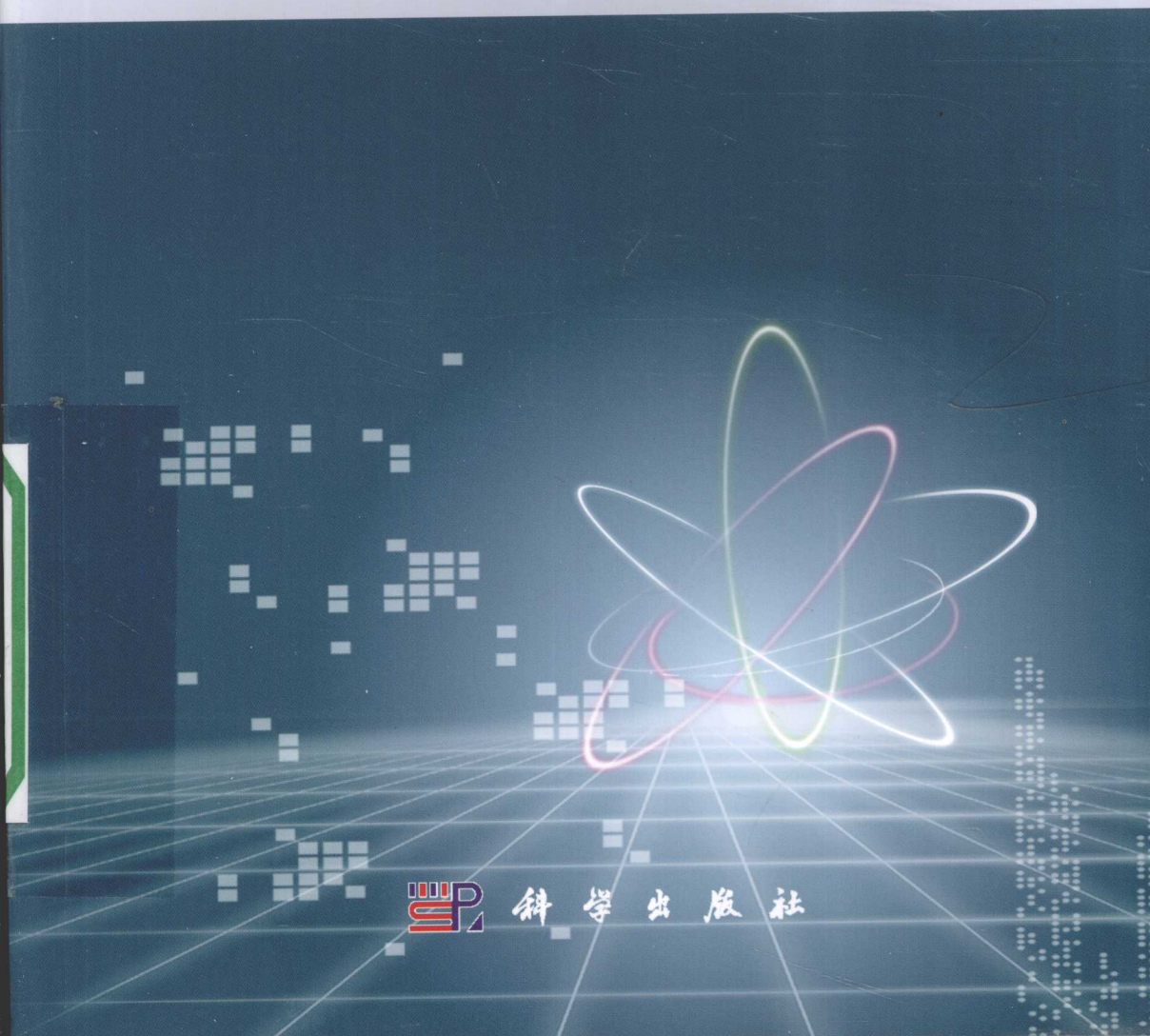
匿名通信理论与技术

吴振强　周彦伟　霍成义　著



科学出版社

匿名通信理论与技术

吴振强　周彦伟　霍成义　著

阜阳师范学院图书馆
藏书

科学出版社
北京

内 容 简 介

随着移动互联网的不断普及和物联网的逐步推广，未来人们在数字空间下的活动将成为生活的一部分。尤其是在我国信息消费产业的持续推动下，增强人们在数字空间下的自我保护意识，指导人们在网络空间下用匿名通信的方法对自己的隐私进行保护将是一项非常有价值的工作。本书以网络空间下用户隐私保护的匿名通信技术为研究对象，对匿名通信的理论、技术和应用等进行系统性的归纳和总结，对引导人们在网络空间下的隐私保护具有重要的理论和应用价值。

本书以理论为指导，以应用为目标，包含大量的模型与实例，可作为计算机、通信和信息安全等专业本科生、研究生的教材或自学参考书，也可供教师、科研人员及工程技术人员参考。

图书在版编目(CIP)数据

匿名通信理论与技术/吴振强，周彦伟，霍成义著．—北京：科学出版社，2015.9

ISBN 978-7-03-045743-1

Ⅰ.①匿… Ⅱ.①吴… ②周… ③霍… Ⅲ.①通信保密-研究 Ⅳ.①TN918

中国版本图书馆 CIP 数据核字(2015)第 225245 号

责任编辑：宋无汗　杨向萍　纪四稳/责任校对：韩　杨
责任印制：赵　博/封面设计：红叶图文

科 学 出 版 社 出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com

三河市骏杰印刷有限公司印刷
科学出版社发行　各地新华书店经销

*

2015 年 9 月第　一　版　开本：720×1000　1/16
2015 年 9 月第一次印刷　印张：19
字数：378 000

定价：100.00 元

（如有印装质量问题，我社负责调换）

前　言

匿名技术是网络安全领域中一个非常重要的研究领域。现有的隐私保护主要是通过密码技术实现的，这对保护网络中的通信消息内容比较有效，但密码技术难以防范攻击者在通信链路上的流量分析攻击。于是，基于密码学的匿名通信技术针对有线互联网提出了广播、代理、洋葱路由和包混淆等匿名通信模型，它们通过一定的变换方法将信息流中用户的身份、通信双方的关系、通信特征、信源或信宿等加以隐藏，使攻击者无法获知通信双方的一些隐私信息。然而，无线网络中无线链路的开放性、网络拓扑结构的动态变化性和资源的有限性决定了无线网络的匿名性要求更高，现有匿名系统中采用的伪名（pseudonym）机制只能解决用户身份的匿名需求，攻击者仍可以通过嗅探方式对匿名系统进行攻击，建立节点流入与流出消息之间的映射关系。为了阻止嗅探攻击，广播模型有广播隐式地址（implicit addresses）法和 DC 网络（dining cryptographers network）法，这两种方法要求系统的参与方在一个封闭的匿名集合中，无法适应开放式的无线网络；代理模型有 Anonymizer 式的单代理和 Crowds 式的多代理，攻击者仍能根据节点的出入消息建立相应的映射关系而威胁系统的匿名性，且这种模型要求网络节点是封闭的静态集合，也无法适应开放式的无线网络；洋葱路由（onion routing）模型通过消息的多层封装与填充，以阻止嗅探攻击，要求通信过程中预先建立虚电路，这导致系统时延和计算复杂度的增加，同样无法适应拓扑结构动态变化的无线网络；混淆模型是采取重新排序、延迟和填充等方法来加大攻击者流量分析的难度，尽管这种模型不要求节点的集合在一个封闭网络中，但填充法会增加无线链路的负载，增加时延会严重地影响无线网络的实时性应用，因此传统的混淆模型也无法完全满足无线网络对匿名性的要求。

本书在介绍现有匿名通信模型的基础上，针对目前用户对移动互联网服务的位置、身份和通信特征等隐私保护，以及物联网和 CPS（cyber physical system）隐私保护的需求，在国家自然科学基金面上项目（项目编号：61173190）和陕西师范大学优秀著作出版基金资助的基础上，重点研究基于网络编码的编码混淆方法与技术。本书对这些工作进行系统性的归纳与整理，从匿名概念、匿名通信理论、匿名通信技术和匿名通信的应用等多个层次进行系统化的介绍，整理出国内第一本系统化的匿名通信理论与技术专著，为用户在网络空间下的隐私保护提供基本指导。

目　录

此为试读，需要完整PDF请访问：www.ertongbook.com

1 概　　论

网络空间下的匿名通信技术研究始于 Chaum 在 1981 年发表的开创性研究成果[1]。该论文提出消息混淆（mix）的思想，并将其应用到匿名电子邮件系统中。此后，很多研究工作致力于构建、分析和攻击匿名通信系统。1996 年 6 月在英国剑桥召开的第一届信息隐藏年会（Workshop on Information Hiding，IH），以及 2000 年 7 月在美国加利福尼亚召开的以隐私保护为主题的第一届学术会议（Workshop on Privacy Enhancing Technologies，PET）进一步促进了匿名通信技术的发展。

从目前的研究看，匿名性没有一个统一的定义和表达方式，也没有一个衡量匿名度的标准，更没有一个统一的衡量匿名的数学模型，因此目前还没有完整的匿名性理论。比利时研究机构在“电子服务中的匿名与隐私”（Anonymity and Privacy in Electronic Service，APES）[2]的项目报告表明，目前匿名性研究仍处于应用分析阶段。报告把目前的匿名应用分为匿名通信、匿名 E-mail、匿名发布、匿名浏览、匿名支付、匿名投票、匿名拍卖和法律等。

针对目前国际上匿名研究现状，本章主要对匿名的基本概念进行介绍，并对目前研究中普遍采用的名词和术语进行定义。

1.1　匿名的相关概念

网络中通信的实体可以分为发送者（sender）和接收者（receiver），通信的内容称为消息（message）。传统的网络安全更多地研究通信内容安全和保密问题，即消息的机密性、完整性、可用性、可控性和不可否认性。匿名问题考虑的是如何保护发送者和接收者的身份信息。一个匿名系统的攻击者（attacker）希望得到的是“谁和谁”在通信，甚至要控制或破坏通信过程。

下面结合 Pfitzmann 等[3]给出的匿名性、不可关联性、不可观察性的定义，给出匿名相关概念的描述。

1.1.1　匿名性

就概念而言，匿名性（anonymity）是多种学科相互交叉，并且涉及技术和法律的概念。根据《韦氏新大学词典》的定义，匿名性是不知道名字或身份的一种状态，或者是不知道作品的作者或作品的来源，或者是缺乏个体的特征、差异和可识别性。

此为试读，需要完整PDF请访问：www.ertongbook.com

匿名性通常也定义为缺乏身份标识，根据文献[4]，考虑到假冒或伪造的身份，本书倾向于定义匿名性是缺乏真正或实际身份的信息，实际上匿名通信中所见到的身份一般都是伪名。

匿名（anonymity）是指一个对象在一组对象的集合（即匿名集合，anonymityset）中不可识别的状态。匿名集合是指发生某种行为（如发送一封电子邮件或者访问某个网站）的可能实体（如用户）的集合，匿名集合的概念是研究匿名技术的基础。匿名是借助于其他实体的行为来隐藏自己的行为。例如，当使用某种匿名服务时，一个实体只有混入其他使用该匿名服务的实体中才可能取得匿名。通常，匿名集合越大，分布越均匀，匿名性就越强。在开放环境中，一个实体只要其行为的概率不为零，其就是匿名集中的成员。例如，在匿名通信系统中，某一消息的发送者可能是 s_1，s_2，…，s_n，那么 $S=\{s_1, s_2, \cdots, s_n\}$ 就是发送该消息的匿名集合。

匿名与身份有直接联系，一个匿名用户一般不可能揭示出其真实身份。实际情况是每个人都希望控制自己的个人信息并保护好自己的秘密，因此，匿名可以作为保密的一种手段。

1.1.2 不可关联性

不可关联性（unlinkability）可以适用于系统中任何对象（实体、信息、事件、行为）。两个或多个对象是不可关联的，意味着在系统中，相对于其先验知识的关联性（linkability），其关联性没有发生变化。这意味着攻击发生前（攻击者的先验知识）和发生后（攻击者的后验知识）对象可被关联的概率保持不变。例如，攻击者在攻击前和攻击后认为某两个消息是由同一个发送者发送的概率保持不变，则可以说这两个消息是不可关联的。可关联性的破坏不一定会使匿名性受到破坏，甚至有些情况下匿名程度都不会受影响。例如，在一些情况下可以跟踪到谁发送了什么信息，同时可以跟踪到谁接收了什么信息，但可以保证发送者和接收者之间的通信关系是无法推知的，即可以保证通信关系的匿名性。

1.1.3 不可观察性

不可观察性（unobserability）是指不能从发送或接收事件集合中分辨出某个发送或接收事件。这意味着不能从“随机噪声”中分辨出消息来。不可观察性是匿名性的必要条件，但不是充分条件，即匿名性可以满足不可观察性。例如，发送者匿名意味着发送者肯定是不可观察的，因为该发送者是否发送了消息都没办法观测到，也就谈不上与其他任何消息进行关联。但是不可观察性的破坏不一定导致匿名关系的破坏，如可以观察到谁在发送信息，谁在接收信息，但无法确定发送者与接收者之间的关系。

1.1.4 匿名通信

匿名通信（anonymous communication）是指通过一定的方法将业务流中通信实体的网络地址、实体间的通信关系等隐私信息加以隐藏，使攻击者无从直接获知或推知双方的通信关系或通信的一方。在发送者 Alice 与接收者 Bob 的通信中，如果 Bob 不知道 Alice 的身份，并且第三方观察者 Eve 不能把 Alice 和 Bob 关联起来，则称 Alice 匿名地与 Bob 通信。有时允许 Bob 知道 Alice 的身份，但 Alice 和 Bob 都会对 Eve 隐藏他们之间的通信。

1.2 匿名通信术语

为了便于对后续章节内容的阅读和理解，准确和一致性的术语是非常必要的，本节将给出目前有关匿名通信的术语的清晰定义和说明。

1.2.1 保密

保密（privacy）是一个很宽泛的概念，总体来讲，对不同的个体和不同的社会形态，其表述是不相同的，并且是围绕着利益进行竞争。《韦氏新大学词典》（*The Webster's New Collegiate Dictionary*）将保密定义为：从群体或观测资料中获取事物性质或状态；《剑桥国际英语词典》（*The Cambridge International Dictionary of English*）将保密直观地定义为：未经授权进入而获得的特权或保持个体间秘密关系的活动。根据应用中的保密定义，保密是指个体有持续的个人空间，从别人或组织中应享有的自由。保密表现在以下四个方面。

（1）个人的保密，是指个人身体的整体性。

（2）个人行为的保密，是指在私下或公开场合下的个人行为，尤其是指对社会文化的态度。

（3）个人通信的保密，是指对个体间相互通信的保护，或对其他人或组织的通信进行常规控制。

（4）个人数据的保密，是指为保护个人的数据而限制他人或组织进行访问操作，至少应有一套完善的数据存取访问控制方案。

信息保密是通信保密和数据保密的结合，在电子领域中优先考虑的是内容保密。由于包含大量个人信息的计算机数据库的普遍使用，对信息保密产生了一系列挑战，这一方面是缺乏足够的安全设施导致无法有效地保护个人信息，另一方面是鼓励采用保密技术，一个有效的应用案例就是匿名通信技术。

1.2.2 身份

身份（identity）在真实世界中可以表示为个体的个性或有差异的特征，身

份不简单是一个名字，而是具有一系列特征的集合。在真实世界中这些特征包括个体的一些不变特性、职位以及他人的理解等。多年以来，大多数人的身份是由两个相关联但又从未区分清楚的两个部分组成，一种是用外表来表示身份，这主要体现于现实生活的交际之中；另一种是根据记录的信息来识别一个独立个体，如在 Internet 中就引入了一个新的术语“数字身份”。

数字身份是在个体的可识别信息与数字信息之间建立映射，数字身份是一种凭借与数字角色相关的数据，通过收集、存储和分析与个体相关的数字表示形式。在电子交易中，数字身份代表个体属性的集合，如头衔、角色、地位、权力和公司资源的特权、人员的薪水、利益等。

数字身份可以通过电子邮件、信用卡号、IP 地址等进行跟踪。然而身份验证或授权访问要用特殊的方法进行处理，如基于身份的数字签名。

1.2.3 伪名

伪名（pseudonymity）是通过伪造有区别的标记而使用的名字，目的是保持匿名性。考虑到匿名表示，识别连接到某个合法用户对象是否使用伪名是非常重要的。

在电子领域中使用的伪名就是通过所谓的“数字伪名”，在公钥系统中，密钥的匿名持有者进行签名验证，授权机构建立“花名册”或伪名表，并决定何种应用可应用伪名，并且是不可跟踪的。如文献[5]将邮件发往匿名的授权机构，他们可以提供不可跟踪的匿名 E-mail。

伪名和匿名的差别是：伪名是指对象使用伪名作为其身份进行标识，伪名与对象之间具有不可关联性，伪名是提供匿名的一种方法；在匿名中，伪名是一个伪造的某人身份，用于数据的传输和通信，必须是通信各方不能揭示出匿名用户的真实身份。在真正或完全匿名中，通过伪名是不可能得出任何真实身份信息的。

通过伪名可以验证对象发出的消息，如使用公钥作为伪名。与匿名相比，使用不变的伪名可以建立问责制和声誉机制，防止匿名系统的滥用。

1.2.4 实体

如今大多数应用都是基于客户机/服务器或浏览器/服务器模式，客户机是请求者，服务器是响应者，二者在功能上扮演不同的角色，至少是行为不同，故把这二者称为实体（entity）。

为了分析应用的匿名特性，人们必须分辨出不同的实体或在应用执行模型中的参与各方，在应用系统风险分析上重要的是分析破坏匿名性。例如，在分布式客户机/服务器系统中，这些角色由客户机、服务器和一些可能的潜在服务组成。

在实际情况下，实体的确定要依赖于特定的应用，如 web 浏览中的实体有浏览器（客户机）和 web 服务器（服务器）。

考虑到不同角色的差异，它们之间有可能进行合作，以获取敏感信息，故将这种行为称为串通或合谋攻击。

1.2.5 角色

在匿名模型中，不同的角色（role）对敏感信息的存取权限通常是不同的。例如，在一个匿名浏览系统中，客户肯定知道自己的身份，但服务器或窃听者是不可能知道此信息的。此类信息与实体在应用中的特定任务和职责有关。后续章节将用角色来区分不同的职责。

在一个应用系统中，不同角色相互协作以达到一个共同的目标。尽管应用系统的功能不尽相同，但从一个抽象层次上看，它们的角色往往是相似的。为了进一步讨论角色，将从以下三个方面着手。

(1) 角色与应用。从功能角度看，应用系统提供的服务有以下角色。

请求者：是应用系统的请求方主动地向服务器发出一个请求。

接收方或响应者：是请求的地址方，一个接收方是应用系统的受动方，响应方的反应是向请求者发送一个响应包。

发送者：在应用系统中，任何一个用于进行通信的实体，通常是请求者常常扮演成发送者来传送一个请求到接收方或响应者。

接收者：用于接收通信的任何实体，一般处于接收方或响应者一边。

(2) 角色与匿名性。为了讨论匿名的级别，在一个系统中需要以下角色。

匿名请求者：为了建立一个匿名系统，请求者的作用是修改或扩充合并增值服务，因此这里再一次包括了请求者角色。

匿名接收方或匿名响应者：同角色与应用中的讨论相似。

匿名提供者：是匿名服务中的任何一个实体，其重要特征是对用户信息的了解情况。

可推知者：提供者作为匿名服务的一部分，他所获取的信息能危及用户身份安全。

不可推知者：提供者作为匿名服务的一部分，但不能获取该类信息。

托管者：是系统中的一个共知的身份，直接完成一个授权任务。对一个定义好的任务子集，被执行任务的所有实体信任，这种信任为了防止误用而需提前定义。在许多系统中，这个托管角色就是授权机构。

(3) 其他角色。

攻击者：对系统中试图破坏匿名特性的实体，例如，窃听者的位置可能是局部的，也可能是全球性的，因此窃听者有可能听到部分或全部匿名应用的通信。

软件/硬件的开发商或供应商：这一角色的任务是编写软件或维护硬件。这类角色对匿名系统构成了严重的威胁，遗憾的是这一角色通常被忽略。

在一个应用系统中，一个实体可以有不同的角色，例如，一个角色与 E-mail 系统的通信有关，同时也与匿名系统有联系。在一个集中式应用系统中，必须分辨出系统中实体所扮演的不同角色。但是，一个系统明显由不同的阶段组成，如电子商务中明显有撤销和支付阶段，在每一个阶段中实体也可能有不同的角色。因此，实体和角色也需要在应用中的特定阶段来分析。

1.3 匿名性分类

在前面讨论的实体和角色基础上，本节利用分类学方法对匿名类型进行分类。由于匿名特征明显依赖于应用系统中的特定角色，所以任何对匿名性的声明总与特定角色有关。故用表达式可以精确地表示匿名特性：role X is $<$type, degree$>$ towards role Y。

为了准确理解不同的匿名特征，下面引入了 5 个重要的衡量参数。

识别性（identifiable）：是匿名系统中通过数据交换识别出通信方（发送方或接收方）真实身份的可能性。若 p_j 表示节点 v_j 是发送方的概率，则有

$$\sum_{j=1}^{n} p_j = 1 \tag{1-1}$$

式中，n 是系统的节点数。

连接性（linkable）：是用户在使用资源或享受服务时被第三方将这些资源或服务与用户角色联系在一起的可能性。$X_{E,F}$ 表示活动 E 和活动 F 之间存在对应关系，对两个活动 E 和 F，如果攻击者 A 在观测到每个信息 B 后仍满足 $0 \leqslant P(X_{E,F} \mid B) \leqslant 1$，则称 E 和 F 是无连接的。如果满足 $P(X_{E,F} \mid B) = P(X_{E,F})$，则称活动 E 和活动 F 是完全无连接的。因此 $P(X_{E,F} \mid B)$ 可以用来度量节点的连接性。

跟踪性（traceable）：是指攻击者在应用组件之间跟踪通信并获取保密信息的可能性。识别性是通过真实数据交换获取身份信息，而跟踪性是通过获取的通信内容来推测出通信用户身份信息。跟踪性除了关心身份，还关心所有其他秘密信息。用 p_j 表示节点 v_j 的可跟踪性概率，满足式（1-1）。

条件性（conditional）：指匿名的可控性，是当可信方（如密钥管理中心）在必要时（如法官审理案件）要求揭示用户身份信息的可能性。用 p_j 表示节点 v_j 所需具有的条件性概率。

持久性（durable）：是指匿名通信保持时间的一种量化表示，用 t 表示匿名通信的持续时间。

这5个特征参数中，前3个参数的作用是识别匿名类型，基于这3个参数，给出了匿名类型的组合分类表，如表1-1所示。

表1-1 匿名类型参数的组合分类表

序号	识别性	连接性	跟踪性	评 价
1	√	√	√	
2	√	√	×	
3	√	×	√	相互矛盾，不可能组合
4	√	×	×	相互矛盾，不可能组合
5	×	√	√	
6	×	√	×	
7	×	×	√	相互矛盾，不可能组合
8	×	×	×	

从表1-1中的可识别性、可连接性和可跟踪性3个参数的8种组合中，有3种组合是相互矛盾的，从而是不可能出现的组合。因为能识别身份就意味着是连接的，故组合3、4无效，同样能跟踪的也隐含着是连接的，组合7也无效。其他组合的含义如下。

组合1根本不具备匿名性，与系统有关的身份是可知道的且通信是可跟踪的，因此可以理解为最弱的匿名性。

组合2也不具备匿名性，然而这种组合与组合1相比是不可跟踪的，这种组合的匿名性通常只用于系统中一个或多个特定角色。例如，在匿名连接情况下，交换数据可能会包含识别信息。在这种情况下，可能会识别出身份的地址。根据定义假设窃听不到任何其他信息，也不可能跟踪通信。

组合5这种弱的匿名形式属于半匿名，在这种情况下，应用系统的数据交换并不能揭示出任何身份信息（如通过移动地址信息或实行加密）。然而通信还是可跟踪的，但是身份仍有可能知道，只是难度更大了。这种类型匿名可以是有条件的，也可以是无条件的。

组合6是利用伪名进行通信的持久匿名，它是通过一个安全匿名身份（通常是伪名）来隐藏真正的身份，由于是连接性的，同一个伪名在通信过程中重复使用。伪名的生成和管理可以由托管中心（有条件的）或客户端自身（无条件的）进行。由于利用伪名的通信要求不可跟踪且不能揭示出信源，所以由客户端生成和管理伪名是比较好的选择。这种情况可以通过匿名连接来实现。

组合8属于全匿名，通过伪名来隐藏实体的真实身份，但这种情况是每次传输用的伪名都不相同，这是一种最强的匿名形式。这种情况可以在有条件或无条件下实现。

此为试读，需要完整PDF请访问：www.ertongbook.com