基于P2P的匿名通信技术研究

张国印 李 璐 李 影 姚爱红
(哈尔滨工程大学计算机科学与技术学院 哈尔滨 150001)

摘 要 随着P2P网络的发展和广泛使用,用户隐私安全的重视程度不断提高。现有的加密技术虽然可以保护通信中的数据内容,却不能很好地保护用户身份,因此在利用匿名技术的同时也要防止不法分子趁机散布非法信息。为了在P2P系统中实现匿名通信,提出了一种基于无环分组路由选择机制,即通过将网络地址切割、成员分组保护、组管理员统一管理来实现通信隐私和涉密通信。仿真实验表明,采用此机制的路由策略得到明显改善,且在保证通信效率的同时提高了用户的匿名性,从而使P2P网络得到更有效的实时保护。

关键词 匿名通信,无环分组,P2P,网络安全
中图法分类号 TP302.1 **文献标识码** A

Research on P2P-based Anonymous Communication Technology

ZHANG Guo-yin LI Lu LI Ying YAO Ai-hong
(College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China)

Abstract With the development and wide use of P2P networks, user privacy and security come to be valued continuously. Although the content of correspondence data can be protected by existing encryption which can not protect the user's identity very well, we can't give outlaw the opportunity to spread the illegal information when the anonymous technology is used. Therefore, this paper put forward a method based on acyclic packet routing mechanism to achieve the anonymous communication in P2P system, communicate privacy and confidential communications by cutting network address, protecting members group, and uniformly managing group administrator. Simulation results show that routing strategy has been improved obviously by the mechanism. The mechanism can guarantee communication efficiency while improving the anonymity of user, and the P2P network can be protected in real-time effectively.

Keywords Anonymous communications, Acyclic packet, P2P, Network security

P2P(Peer-to-Peer)网络作为对等网络模式,克服了普通C/S模式的一些弊端,节点的平等性、高分布性等特点给网络领域注入了新的活力和发展空间。但由于P2P网络框架的特殊性,节点加入与离开的管理、数据交换方法、网络拓扑机制、数据的调度策略等都是影响P2P网络安全性及用户匿名性的重要因素[1]。

匿名性作为信息安全的重要研究方向,与网络及国家的信息安全密切相关。匿名通信是在通信的过程中增加节点转发及对数据的处理来隐藏信源、信宿的身份,实现用户的隐私保护。

在P2P对等网络中匿名通信技术主要面临两个问题:

(1)如果匿名通信技术存在漏洞,将会导致窃听者通过盗取的信息包来推断出通信双方的位置及私秘信息,网络就会遭受进一步的攻击,而导致更加严重的破坏。

(2)匿名通信技术若被不法分子利用,进而散布非法信息及泄露涉密信息,将严重危害社会的稳定及国家的安全。

针对以上问题,国内外各大院所展开研究,并提出一些较为成熟的方案,但多存在为提高通信效率而忽略网络中环路的问题,即同一路径上的节点多于(或者等于)2次的访问将会造成系统通信匿名度下降的危险。

因此,为在P2P网络中实现匿名通信技术,解决P2P已有路由选择机制存在的环路问题及匿名滥用问题,本文提出了一种无环分组的路由选择机制,并对网络框架进行适当改善,利用地址切割、分处存放的思想实现了对匿名滥用节点的IP追踪机制,以提高网络的安全性,保证网络的健壮性。

1 相关技术

在匿名通信技术方面,国外起步较早。David Chaum首次指出电子邮件的追踪问题并提出MIX解决办法,利用MIX的转接作用及对传送数据重新排序、延时、填充流量的方法隔断网络通信双方直接通信,这样接受方所获得的仅为发送方身份的数据及MIX身份[2,3]。我国匿名技术起步较晚,但是

到稿日期:2011-05-09 返修日期:2011-06-27 本文受国家自然科学基金(61073042)移动P2P网络数据分发机制研究项目,中央高校基本科研业务费专项资金(HEUCF 100606)资助。

张国印(1962-),男,博士,教授,CCF会员,主要研究方向为网络信息安全,E-mail:zhangguoyin@hrbeu.edu.cn;李 璐(1985-),女,博士生,主要研究方向为网络信息安全;李 影(1985-),女,硕士生,主要研究方向为网络信息安全;姚爱红(1972-),女,博士,副教授,主要研究方向为可重构计算。

各大院所的积极跟踪及研究所取得的成果已达到一定水平[4]。孙黎等在基于机构化 P2P 的匿名系统中通过为每个新加入成员分配初始信誉度值并调整动态，使得具有不同信誉度值的成员获得的匿名通信服务级别不同，从而通过在数据包中内嵌由发送源所在的管理员加密的发送源地址信息，实现了对发送源的匿名撤销[5]。

针对匿名通信，现有基于 P2P 的路由选择机制主要有以下几种：

(1)重路由模型

这种方法可以通过重路由传输途径将包头地址多次重新写入，以达到攻击者无法从截获的包头中分析出通信双方的地址，起到匿名通信的目的。

(2)路径选择策略

随机选择、倾向性选择、区域性选择、偏前缀选择均是 MIX 级联结构的固定匿名路径，发送者只要在发送时选择其中一种即可[6]。

(3)源路由选择

在源路由选择方式中，发送者确定重路由成员的选择。发送者在匿名通信网络中选取部分成员，通过密钥机制确定会话密钥，进而建立起重路由路径[7]。

(4)下一跳路由选择

下一跳路由方式与源路由方式是截然相反的，是指在匿名通信系统中，发送者 S 从其邻居节点中选取某个节点来作为下一跳的中间节点，这里设为 I_1，然后将包含接收者地址的连接请求发送给 I_1。I_1 接到请求后有两种选择：一种是直接发送给接收者，另一种是继续转发给下一个中间节点。为了使转发路径无限长，这里设定了一个转发概率 P_f 并且已有人对此进行研究，形成了转发概率递减的策略，从而控制了重路由路径长度[8]。

2 无环分组路由选择机制设计

前述的下一跳路由选择机制，虽然具有通信效率高的优点，但是却存在问题：某节点在同一条路径中被大于或者等于 2 次使用，即路径中可能存在环路，这样就会降低系统通信的匿名度[9]。为了解决环路问题，本文将系统成员进行分组管理，每一分组分配一个管理员节点。在这里设定的组管理员可以向节点认证中心申请下载其他管理员节点的地址信息，而普通的成员节点不允许知道其他的组管理员的地址。图 1 对这种拓扑结构的路径节点选择策略的举例进行了详细说明。

图 1 为无环分组路径选择策略图。S 为发送者，当其要建立一条重路由路径时，它先向本组的管理员 AD_m 发送建立路径的请求。管理员接到本组成员 S 建立路径的请求后，随机选择 K 组(不包括本组，但是可能包括接收者的分组管理员节点)，向该 K 组管理员节点发送索要节点请求。图 1 中另外选择了管理员节点 AD_n、AD_k 和 AD_j，这 3 个管理员节点从本组中选择 N_A 个成员节点作为 S 要建立的路径的中间节点，在这里每个管理员在本组内选择一个节点的概率为 P_t，不选择节点的概率为 $1-P_t$。P_t 选取不同的概率值时，每个组内选择的节点数的期望值有所不同。

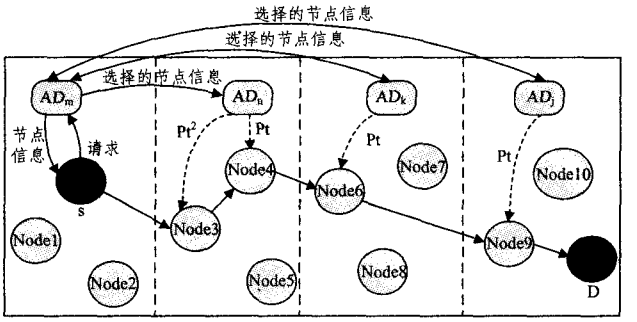

图 1 无环分组路由选择策略

管理员节点 AD_n、AD_k 和 AD_j 将已经选择的本组内作为路径节点的信息发送给管理员 AD_m，然后 AD_m 再将其发送给发送者 S。S 通过得到的中间节点的地址信息以及公钥对其进行反向加密。此时，虽然 S 所收到的节点信息都经过本组的管理员发送，但路径节点的通信顺序是由 S 决定的，且这个顺序只有 S 本身知道。在图 1 中，管理员 AD_n 在本组内选择了两个节点作为 S 服务的节点，管理员 AD_k 在本组选择一个节点作为中间节点。管理员 AD_j 是接收者 D 所在的组，将其选作中间节点的组，该管理员在本组选择一个节点作为中间节点。S 在收到了从本组管理员 AD_m 发送过来的路径中间节点信息后，对节点进行排序，然后反向加密进行封装，再经过管理员 AD_m 发送出去。

在无环分组路径选择策略中，S 对外的信息都是经过管理员发送出去的，返回的信息也是经由管理员，因此对管理员的安全性能要求极高。对组外的观测者来说，他们只能观测到本组向外发送或者接收数据，并不能确定本组内哪个节点在动作，这是利用了成员的平等性来隐匿自身的信息。此外，虽然管理员可能在将路径节点信息发送给 S 之前自己会保留一份，但是他不会确定 S 将节点发送出去的具体排列顺序，所以相对抵消了一些不安全隐患。

无环分组路由选择的优点是避免了同一条重路由路径会经过相同的中间节点而产生环的可能性[10]。由于节点是从不同组的不同节点群中选取的，即使在组内选择的节点具有随机性，但是选定之后中间节点就已经确定，将不会再有变动。至于中间节点的前后，则由 S 来确定。如果 S 在确定顺序的策略中，同一节点被人为地多次使用(这种可能性是存在的)，S 为了增强匿名安全性，不会制定一个产生死锁的策略。

在中间节点选择策略的问题上，虽然组管理员在从组内选择节点时，每个节点被选择的几率都是相同的，但是还会发生某些节点被多次使用、而某些节点几乎未被使用的情况，这样就会导致多次使用的节点的负载过重[11,12]。

为了解决这个问题，本文采取一种负载均衡的手段。通过数据结构中最小堆的思想来设定节点选择的优先级，从而实现系统成员的负载均衡：构造优先级队列，且此队列由每个分组的组管理员保持和使用。将组内 R 个成员节点的负载、当前是否可用等信息存储到构造好的优先级队列中。具体实现方法如下：

(1)首先构造结构体，用来存储节点的负载信息(已使用该节点的重路由数)，当前节点是否可用(0：不可用；1：可用)。

(2)然后初始化每个节点的结构体信息，以 N 个节点的

负载信息作为关键字(key),利用可用的节点来构造优先级队列。

在每次路径转发选取节点时,从优先级队列中选取堆头的可用节点作为转发节点,因为在此时刻,该节点具有最小的负载。同时,将该节点的负载加 1,并调整优先级队列。这样,每次所选取的转发节点,必然是当前可用的负载最小的可用节点,因而可以实现网络节点的负载均衡。

3 匿名度分析

3.1 下一跳路由选择机制匿名分析

本文将系统中攻击者或者能被攻击者控制的节点称为泄密节点,把不受攻击者控制的安全节点称为非泄密节点。非泄密节点不会对泄密节点提供任何信息。假设包括 N 个成员的系统中含有 R 个泄密节点,匿名性能的主要衡量参数为 $P(I|H_{1+})$,即当路径上包括泄密节点时攻击者能够准确判断数据发送者身份的概率。规定 $P(I|H_{1+})<0.5$ 时系统的匿名等级意义达到 probable innocence,满足匿名系统的基本要求。可以看到 $P(I|H_{1+})$ 越大,系统匿名性能越差;$P(I|H_{1+})$ 越小,系统匿名性能越好。在这里定义:

I:表示路径上第一个泄密节点的上一节点是数据发送者;

H_k:表示路径上第一个泄密节点在该路径的第 k 个位置上($k\geqslant 1$,发送者位置设为 0);

H_{k+}:表示路径上第一个泄密节点在该路径的第 k 个位置之后。

在 Crowds 匿名通信系统中,n 为中转代理总数,$P=N-R/N$ 为非泄密者比例,P_s 为转发概率,根据文献[13],在路径中没有环存在的情况下攻击者判断发送者身份成功的概率为

$$P(I|H_{1+})=\frac{1-PP_s}{1-(PP_s)^n} \tag{1}$$

当路径中存在环时

$$P(I|H_{1+})=1-PP_s+\frac{P_s}{n} \tag{2}$$

且

$$\frac{1-PP_s}{1-(PP_s)^n}\leqslant=1-PP_s+\frac{P_s}{n} \tag{3}$$

即在下一跳路由选择机制中无环情况下要比有环情况下的匿名性能好。

3.2 无环分组路由选择机制匿名分析

定理 1 本文设路径的节点数为 n,同上面的中转代理总数相对应,设节点被选为路径中间节点的概率为 P_z,则

$$P(I|H_{1+})=\frac{1-P}{1-P^n} \tag{4}$$

证明:首先第一个泄密节点出现在第 i 个位置的概率为

$$P(H_i)=P_zP^{i-1}(1-P) \tag{5}$$

由式(5)可得出第一个泄密节点在通信路径的第 k 个位置以及在其之后的概率为

$$P(H_{k+})=\sum_{i=k}^{n}P_zP^{i-1}(1-P) \tag{6}$$

当路径的第一个节点就是泄密节点时,它的前驱的确是数据发送者,猜中的概率为 1,即 $P(I|H_1)=1$。这种路由方式中不存在环,所以第一个泄密节点如果在第二个节点之后,那么它的前驱节点一定不是数据发送者,所以猜中的概率一定为 0,即 $P(I|H_{2+})=0$。所以可得

$$\begin{aligned}P(I)&=P(H_1)\cdot P(I|H_1)+P(I|H_{2+})\cdot P(I|H_{2+})\\&=P_z\cdot(1-P)+0=P_z\cdot(1-P)\end{aligned} \tag{7}$$

继而得到

$$P(I|H_{1+})=\frac{P(I)}{P(I|H_{1+})}=\frac{P_z(1-P)}{\sum_{i=1}^{n}P_zP^{i-1}(1-P)}=\frac{1-P}{1-P^n} \tag{8}$$

从公式中可以看出,当 n 相同,即路径长度一定时,无环分组路由选择机制的匿名性能要比下一跳路由选择机制的匿名性能好,即

$$\frac{1-P}{1-P^n}\leqslant\frac{1-PP_s}{1-(PP_s)^n}$$

这里的转发概率 $P_s\leqslant 1$。

4 仿真实验及分析

选择非泄密者比例为 $P=0.9$,对于下一跳路由选择机制分别选择 $n=5$ 和 50,转发概率 P_s 分别选取 0.2,0.4,0.6,0.8;而对于无环分组路由选择机制分别选择 $n=5$、30 和 50。对两种路由选择机制进行建模测试,每一轮测试 1000 个请求,得出的数据如表 1 和 2 所列。

表 1 下一跳路由机制的理论数据与实际测试数据比较

P_s	理论值 n=5	实际值 n=5	理论值 n=50	实际值 n=50
0.2	0.82	0.828	0.82	0.823
0.4	0.644	0.667	0.64	0.642
0.6	0.482	0.502	0.46	0.471
0.8	0.347	0.397	0.28	0.293

表 2 无环路由机制的理论数据与实际测试数据

P	理论值 n=5	实际值 n=5	理论值 n=30	实际值 n=30	理论值 n=50	实际值 n=50
0.9	0.244	0.267	0.104	0.113	0.1	0.109

由此可以看到实际测试值与理论计算值基本吻合,从而验证了结论。

下一跳路由机制与无环分组路由机制对比曲线如图 2 所示。

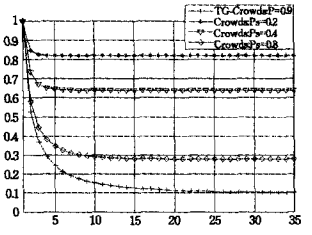

图 2 下一跳路由机制和无环分组路由机制匿名度对比曲线

从图 2 中可以得到,无环路由机制中匿名度小于所有下一跳路由机制中转发概率取值的情况,也即无论重路由路径的长度是多少,无环分组路由选择机制的匿名性能都比下一跳路由选择机制的匿名性能好。可以看到在下一跳路由选择机制中,当路径长度 $n>10$ 时,曲线趋于平滑,系统的匿名度不会再降低,而在无环分组路由选择机制中,当 $n>30$ 时,曲线才趋向于平滑,所以匿名性能一直处于上升的趋势,这说明当通信路径变长时,下一跳路由选择的匿名性的可塑性低于无环路由选择机制。

结束语 本文为实现 P2P 网络匿名通信提出了一种新的无环分组的路由选择机制。将网络中的成员节点进行分组

(下转第 109 页)

表 2　基于 FRS-FCM 的集成入侵检测模型与传统 FCM、GABP 检测模型的检测率、误报率与低频攻击检测率对比

实验结果	连接总数	正常连接数	异常连接数	FRS-FCM			FCM			GABP		
				检测率(%)	误检率(%)	低检率(%)	检测率(%)	误检率(%)	低检率(%)	检测率(%)	误检率(%)	低检率(%)
TS1	2000	1900	100	85	5.1	89.7	82	5.3	79.9	83	7.8	65.3
TS2	2000	1900	100	90	4.8	90.1	85	7.1	81.3	90	8.2	72.6
TS3	2000	1900	100	86	6.7	79.5	88	8.8	80.2	76	8.8	78.3
TS4	2000	1900	100	85	5.4	83.9	83	9.1	75.8	83	6.7	69.4
TS5	2000	1900	100	86	7.4	86.8	85	7.6	78.4	81	8.1	70.1

结束语　本文针对当前入侵检测方法存在的不足，提出基于模糊粗糙集与 ReliefF 技术的 FRS-FCM 算法，并将此算法运用到集成入侵检测中。最后，利用 KDD Cup 1999 数据集进行实验，验证了基于 FRS-FCM 算法的集成入侵检测方法能够在降低误检率的同时提高检测率与泛化能力，并对低频攻击的检测率有较大的提高。但本方法在确定聚类中心个数以及子分类器检测结果的集成权重时还存在主观因素，而且对检测报警信息并没有进行相关处理，因此下一步的研究工作主要包括以下两个方面：(1)采用机器学习算法确定初始聚类中心的个数以及子分类器检测结果的集成权重，并对集成权重进行动态学习更新；(2)对报警信息进行相关性和聚类处理，增强报警针对性，以减少网络管理员的工作负担。

参 考 文 献

[1] Tsai C-F, Hsu Y-F, Lin C-Y. Intrusion detection by machine learning: A review [J]. Expert Systems with Applications, 2009, 36: 11994-12000

[2] Wang Gang, Hao Jin-xing, Ma Jian, et al. A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering [J]. Expert System with Applications, 2010, 37: 6525-6232

[3] 袁妍，洪晓光. 基于模糊-粗糙集的移动对象 k 近邻预测[J]. 计算机科学，2008，35(2)：140-143

[4] 赵越，张为群. 一种基于 CFCM 的集群入侵检测方法的研究[J]. 计算机科学，2010，37(6)：176-178

[5] 李洁，高新波. 基于特征加权的模糊聚类新算法[J]. 电子学报，2006，34(1)：89-95

[6] 王骏，王士同. 基于混合距离学习的双指数模糊 C 均值算法[J]. 软件学报，2010，21(8)：1878-1888

[7] 徐冲，王汝传. 基于集成学习的入侵检测方法[J]. 计算机科学，2010，37(7)：217-219

[8] 吴春琼. 基于神经网络与遗传算法的入侵检测研究[J]. 计算机安全，2010，11：25-27

[9] 张义荣. 一种基于粗糙集属性约简的支持向量异常入侵检测方法[J]. 计算机科学，2006，33(6)：64-71

[10] 杨德刚. 基于模糊 C 均值聚类的网络入侵检测算法[J]. 计算机科学，2005，32(1)：86-91

[11] 肖敏，柴蓉，杨富平，等. 基于可拓集的入侵检测模型[J]. 重庆邮电大学学报：自然科学版，2010，22(3)：345-349

(上接第 100 页)

管理，为每组设定组管理员。由组管理员选择成员节点，由发送者编排路径节点通信顺序。本文证明了无环分组路由选择机制的匿名度与系统中非泄密节点成员呈比例的关系，并分析了相同路径长度下的下一跳路由选择机制的匿名度，将两者进行了对比和测试分析。实验数据表明，在通信路径长度相同的情况下，无环分组的路由选择策略比下一跳路由选择策略的匿名性能要好，并且在路径长度 $n>10$ 时，下一跳路由选择策略的匿名性能不会再提高，而无环分组的路由选择机制的匿名性能却一直增加。直到 $n>30$ 时，匿名性能才趋于定值，进一步完善了 P2P 网络的安全性。

参 考 文 献

[1] 欧中洪，宋美娜，战晓苏，等. 移动对等网络关键技术[J]. 软件学报，2008，19(2)：404-418

[2] Chaum D. Untraceable electronic mail, return addresses, and digital pseudonyms [J]. Communications of the ACM, 1981, 4(2): 84-88

[3] Jakobsson M. A practical mix[C]// EUROCRYPT'98. 1998: 448-461

[4] 高蕾，李大兴. 基于 P2P 网络的匿名通信系统[D]. 青岛：山东大学，2008

[5] 孙黎，王小刚. 基于结构化 P2P 的可控匿名通信系统的研究[J]. 科学技术与工程，2010，5(1)：306-310

[6] 陆天波，时金桥，程学旗. 基于互联网的匿名技术研究[J]. 计算机科学与探索，2009，3(1)：35-42

[7] M'Raihi D, Pointcheval D. Distributed Trustees and Revocability: A Framework for Internet Payment[C]// Lecture Notes in Computer Science. 1998, 1645: 28-50

[8] Zhang F, Zhang F T, Wang Y. Fair electronic cash systems with multiple banks based on ACJT group blind signature[J]. Journal of Wuhan University of Technology, 2000, 32(5): 849-852

[9] 邓琳，谢鲲，李仁发. P2P 匿名通信系统的匿名度量及协议研究[D]. 长沙：湖南大学，2009

[10] Serjantov A, Danezis G. Towards an information theoretic metric for anonymity[C]// Proceedings of Privacy Enhancing Technologies Workshop. 2003: 41-53

[11] Sander T, Ta-Shma A. Flow control: A new approach for anonymity control in electronic cash systems [J]. Conference on Computational Intelligence and Security, 1999, 1(1): 354-379P

[12] 江丽，徐红云. 基于组群的匿名通信协议研究与探讨[J]. 计算机工程与应用，2008，44(9)：125-128

[13] Camenisch J, Maurer U, Stadler M. Digital payment systems with passive anonymity-revoking trustees[J]. Journal of Computer Security, 1997, 5(1): 69-89