

文章编号:1001-9081(2010)06-1483-03

分布式代理记忆机制的 P2P 网络研究

秋兴国,王博辉,龚尚福

(西安科技大学 计算机科学与技术学院,西安 710054)

(gongsf126@126.com)

摘要:为了解决节点频繁离线、信任机制缺乏和带宽有限等问题对 P2P 网络服务质量的影响,提出了一种分布式代理记忆机制的 P2P 网络模型。在该模型中,数据被分成若干个数据块,节点访问结束后对数据块的服务进行评价,数据块内容和服务评价更新存储于邻域节点及后继节点;节点访问信息时,根据本地策略优化搜索代理记忆;对服务评价较低的代理记忆进行定期更新与清除。该模型通过数据块分布式代理记忆和数据动态更新的方法,有效地提高了数据可用性,阻止了病毒文件的传播,减轻了带宽压力,提高了搜索效率,增强了系统的安全性和网络性能。

关键词:P2P 网络;分布式代理;记忆更新;服务评价;本地策略

中图分类号:TP393.08 **文献标志码:**A

Study of P2P network model based on distributed Agent memory mechanism

QIU Xing-guo, WANG Bo-hui, GONG Shang-fu

(College of Computer Science and Technology, Xi'an University of Science and Technology, Xi'an Shaanxi 710054, China)

Abstract: Since such problems of P2P network as node frequent offline, lack of trust mechanism and bandwidth limited influence the service quality, a distributed Agent memory mechanism of P2P network model was presented. In this model, the data were divided into several blocks of data, the block service was evaluated after the node visit, and data block content and service evaluation update were stored in the neighborhood node and subsequent node; when the node visits information, according to local strategy, optimize memory, and regularly update and delete for a low service evaluation memory. The model of distributed data blocks through memory and dynamic data updated Agent improves the data availability effectively, stops the spread of the virus document, reduces bandwidth pressure, improves the searching efficiency, and increases the system security and network performance.

Key words: P2P network; distributed Agent; memory update; service evaluation; local strategy

0 引言

P2P(Peer-to-Peer)网络是一种对等化网络,它通过系统间的直接变换达成计算机资源与信息的共享,其各个终端彼此保持对等独立自治关系,共同享有信息资源,共同承担服务成本。它的研究使充分利用边缘化网络资源成为可能,但同时也面临着管理机制缺乏、数字版权、网络病毒传播、信用机制和带宽与流量流向等问题^[1]。文献[2]提出把数据分成若干个数据块,采用动态分配算法存储到节点中,同时在节点邻居中保存数据块的副本以提高数据块的有效性,实验表明,该算法能够获得较高的数据可用性。但是此方案对存在欺诈或在使用过程中被修改的数据,以及病毒等危险性数据或用户频繁离线的数据的复制算法缺乏研究。文献[3]对数据的放置策略进行了研究,提出了数据块的分布式存储转发随机放置策略,在一定程度上避免了服务端节点故障造成的数据丢失,提高了系统的容错能力,保证了数据的可靠性。但是远距离节点存储操作不具协同性,此外对服务评价低的数据块没有有效的处理,副本冗余持久,致使带宽压力增大。文献[4]对 P2P 网络数据共享中动态解密授权的实现问题提出了一种动态解密授权实现方案,此方案虽然解决了被访问端离线时无法提供即时访问的缺陷,但是动态解密授权工作平台的严格要求(必须找到安全的信任代理平台)会限制移动代理的灵活性,无界移动代理也会使数据安全处于未知状态。文

献[5]提出一种基于模糊理论构建服务行为评价的 P2P 网络信任管理模型,此模型通过将历史行为分组,并赋予不同的权值,提供了一种简捷有效的行为相关性算法。同时使用分布式存储的方法来处理推荐信息,降低了搜集推荐信息的消耗。但是对于共享度高的信息,服务评价趋于可靠,长期良好的评价会导致评价欺诈和伙同欺骗;再次,该模型提高搜索效率有限。

为了解决以上问题,本文提出了一种分布式代理记忆机制(Distributed Agent Memory Mechanism, DAMM)的 P2P 网络模型。

1 DAMM 概述

在 DAMM 模型中,节点对数据是否访问由数据块服务评价 ω 和本地策略 Π 共同决定,只有当服务评价 ω 满足本地策略 Π 时,节点才对数据进行访问,进而才开始分布式代理记忆。

当源节点 A 发布信息 E 时,数据会被分成若干个数据块 e_i ,邻域节点根据本地历史行为分组,结合本地安全防护软件提供的建议对数据块 e_i 进行分布式代理记忆,此时源节点 A 的离开不影响数据块的可用性。若后继节点 B 在线,则对数据块 e_i 搜索,重构数据,恢复信息 E ,并对数据块 e_i 的服务进行评价,评价结果 ω_1 与初始服务评价 ω_0 加权的值赋给 ω_1 ,加权方法称为本地策略 Π ,同时 ω_1 与数据块内容更新存储于数据

收稿日期:2009-11-26;修回日期:2010-01-29。 基金项目:国家“十一五”项目(FIB070335-A8-18)。

作者简介:秋兴国(1965-),男,陕西乾县人,教授,主要研究方向:网络集成、数据库、嵌入式系统; 王博辉(1986-),男,陕西礼泉人,硕士研究生,主要研究方向:网络集成、数据库; 龚尚福(1954-),男,宁夏平罗人,教授,主要研究方向:计算机网络、信息安全。

块 e_i , B 节点及其邻域节点对数据块 e_i 进行分布式代理记忆。

由于数据重构需要搜索若干数据块, 当访问节点 D 访问数据时, 根据本地策略 Π 对可提供可靠服务的数据块进行搜索。若源节点 A 在线, 其节点、后继节点及邻域节点均满足条件, 则对所有符合条件的数据块进行权衡, 选取最优节点进行服务。文献[6]指出, 近期行为的可靠性比较大, 不考虑节点距离时, 一般情况下 B 节点及邻域节点的数据块和服务评价比 A 节点最优。服务行为发生后 D 节点对服务进行评价, 评价结果 ω_2 与 ω_1 或 ω_0 进行加权, 加权结果赋值给 ω_2 , 同时 ω_2 与数据块内容更新存储于数据块 e_i , D 节点及其邻域节点对数据块 e_i 进行分布式代理记忆, 以供后续节点依次访问。

每个用户在一定时期内自愿共享的信息量是有限的, 所以对数据进行分布式代理存储是可行的。当用户信息量积累到一定程度, 又会影响和制约计算机的存储和计算能力, 所有有必要对早期记忆进行考核, 针对记忆进行时间段分组, 使较早不关联服务评价和服务评价低的数据块符合分布式代理记忆规律进行遗忘或删除, 以减轻分布式存储压力。

其思想有以下主要特点: 1) 数据以数据块的形式被分布式代理记忆, 访问数据时搜索其所组成的数据块以重构数据信息; 2) 数据块在后继和邻域节点分布式代理记忆, 不会大范围的传播; 3) 数据块分布式代理记忆处于动态的更新与清除中, 不会造成持久性存储积压; 4) 用户对信息的搜索不再是单一的源节点提供, 而可以看作是多个对等的节点, 有效地减少了搜索时间, 提高了搜索效率。

2 相关定义

2.1 基础定义

定义1 数据可用性。数据在时间 t 内可以被访问到的概率。假设数据块的访问概率为 $p_{\text{可用性}}$, 则:

$$p_{\text{可用性}} = \left(\sum \frac{n-m}{n} \right) \frac{1}{t}$$

其中: n 为访问次数, m 为离线次数, t 为统计时间。若 i 为每份数据分成的数据块个数, 则 i 为数据重构所需最少的数据块个数, 当 $p_i \geq r, r \in [0, 1], r \in \mathbb{N}$, 数据方可重构使用。

设数据可用性用 y 表示, 则 p 越高, 数据可用性 y 越大。

定义2 数据可靠性。数据的可信性, 即在使用过程中可能会产生的危险程度。设数据块危险度集合为 x , 则 $x = \{\text{高度危险}, \text{中度危险}, \text{低度危险}, \text{无危险}\}$ 。若数据可靠性用 k 表示, 则危险度 x 越低, 数据可靠性 k 越高。

定义3 数据完整度。数据上传或下载是否完整。设数据完整度集合为 w , 则 $w = \{\text{完整}, \text{部分}, \text{丢失}\}$ 。完整度越高, 数据的可用性和可靠性越高。

定义4 搜索最优化。在同条件下搜索, 路径最短, 资源消耗最少的搜索方法。设搜索速度用集合 z 表示, 则 $z = \{\text{很快}, \text{快}, \text{普通}, \text{慢}\}$ 。搜索越快, 说明数据的完整度、可用性和可靠性越高。

定义5 服务评价。对数据块的可用性 y 、可靠性 k 、搜索最优 z 和数据的完整度 w 等指标的综合服务评价。假设服务评价为集合 ω , 则:

$$\begin{cases} \omega = \{\omega_0, \omega_1, \omega_2, \dots, \omega_n\} \\ \omega_i = \sum_0^i (ya + kb + wc + zd) \\ a + b + c + d = 1 \\ a, b, c, d \in [0, 1] \end{cases}$$

其中: ω 是对 y, k, w, z 等指标的综合服务评价; a, b, c, d 为权重; 根据普通用户行为需求, y, k, w, z 的值域分别为 $[0.5, 1.0]$ 、 $[0.6, 1.0]$ 、 $[0.5, 1.0]$ 、 $[0.3, 1.0]$, 特定用户可以根据自己需求, 对权重进行修改, 获取相应的数据约束范围, 改变本地策略 Π 的默认值。

2.2 DAMM 定义

通过对以上知识的约束定义, 本模型可以用一个五元组 (S, ω, L, E, Π) 来表示。其中 S 表示节点集合, $S = \{S_1, \dots, S_n\}$, S_i 表示第 i 个子网络节点, 分别包括 n 个节点, 每一个节点含有多个邻域节点, 同一个节点可以归属于不同的子网域。 ω 为数据块服务评价集合, 可以根据数据块的服务评价的时期分为不同的服务评价信息集 $\psi, \psi = \{1 \text{ 日内评价 } \psi_1, 3 \text{ 日内评价 } \psi_3, 7 \text{ 日内评价 } \psi_7, 1 \text{ 月内评价 } \psi_m, 1 \text{ 年内评价 } \psi_n\}$ 。 L 为服务评价信息列表, 可以对其进行查询、添加、删除等操作。 E 表示信息集, E 内容包括数据块和服务评价, 可分为 $\{\text{病毒或恶意类信息}, \text{服务评价低的信息}, \text{服务评价较低的信息}, \text{信任信息}\}$ 。 Π 称为本地策略, 包括用户对P2P网络服务所需的语言变量, 本模型中使用 y, k, w, z 共4个变量, 用户也可以根据需求进行语言变量的增减, 以达到更确切的需求。用户访问信息时通过这些变量条件之间的约束关系产生, 用户每一次的代理记忆信息更新存储于资源列表 L , 计算机根据用户行为对资源列表存储的记忆进行考核与更新。

3 模型介绍

DAMM模型的框架如图1所示。

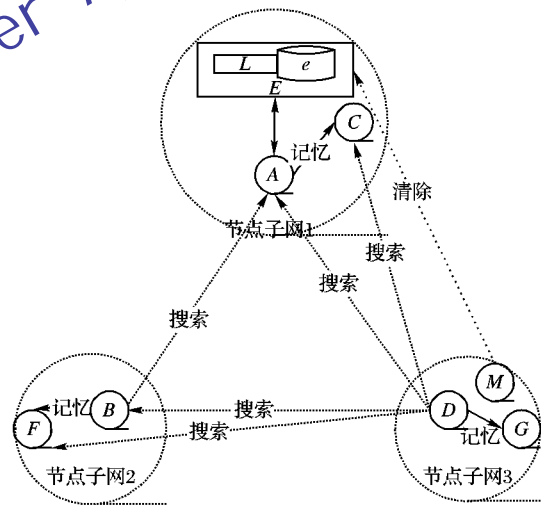


图1 简单框架图

本框架由以下几部分组成: 1) 互联网由多个子网络域组成, 每一个子网络域包含多个节点, 每一个节点代表一个终端, 各终端之间可以相互访问信息。2) 源节点 A 发布的信息 E 以数据块 e 的形式汇入子网络 S_1 , 与源节点 A 处于同一子网络的邻域节点对其数据块 e 进行分布式代理记忆。若后继节点 B 在线(后继节点 B 可以和 A 节点同位于子网络 S_1 , 也可以位于子网络 S_2 或子网络 S_3 , 节点是否在同一子网络不影响对数据块的分布式代理记忆), 则对源节点 A 的服务进行评价, 新的评价与数据块内容更新存储于数据块 e , 其节点和邻域节点进行分布式代理记忆; 当位于子网络 S_3 (假定位于与 A, B 节点不同子网络的节点)的节点 D 访问信息时, 根据本地策略对所有子网络域数据块进行搜索, 对可提供信息服务的节点路径进行优化搜索, 访问结束后对其信息服务进行评价, 新的评价与数据块内容更新存储于数据块 e , 其节点和邻域节点

进行分布式代理记忆,以供后续节点访问。3) 对于在访问过程中服务评价严重低于本地策略的数据块进行记忆清除;对不影响服务评价加权的数据信息进行记忆遗忘。对于重复记忆进行覆盖,不重复存储;对重复记忆删除只搜索信息源进行

删除。

4 算法分析

DAMM 算法流程如图2所示。

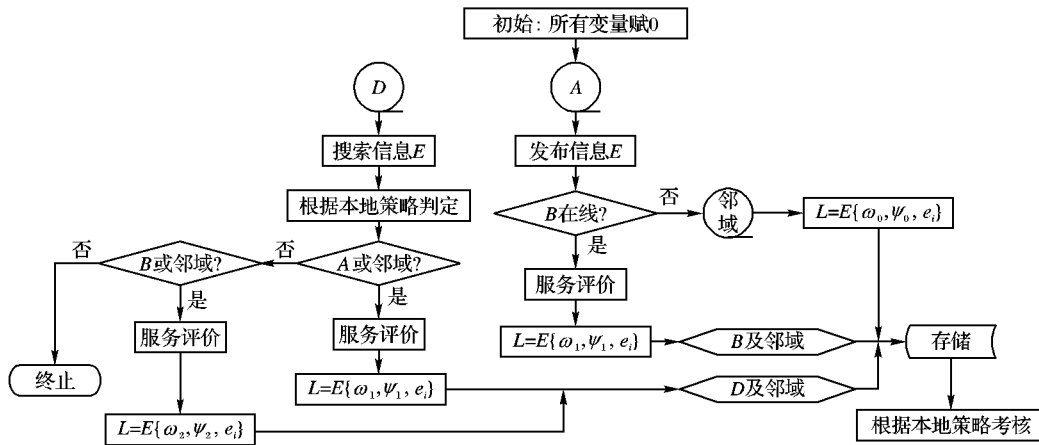


图2 算法流程

其中每次访问 ω 都会被更新存储于信息集 ψ 和资源列表 L ,计算机根据本地策略,对资源列表进行定期的更新删除,以减少存储空间和提高搜索效率。

分布式代理记忆算法的步骤如下:

- 1) $y, k, w, z, i, p, L, \omega, \psi, r \rightarrow 0$ // 变量初始化;
- 2) $A \rightarrow E_0 \rightarrow e\{e_i, \omega_0\}$ // 源节点A发布信息 E_0 , 信息以数据块 $e\{e_i, \omega_0\}$ 的形成存放,每份数据通常会被分为 i 个数据块*//;
- 3) $S_{1\mathcal{A}} \rightarrow E_0 \rightarrow L$ // A节点的邻域节点分布式代理记忆数据块 e ,存储于本地资源列表 L *//;
- 4) if ($B \neq 0 \cup pi \geq r$) // *B节点获取到足够的数据重构所需数据块*//;
 $B \rightarrow e_r \rightarrow E_0\{e_i, \omega_0\} \rightarrow E_1\{e_i, \omega_1\} \rightarrow L$ // *B重构数据,并对数据块 e 服务进行评价*//;
 $\omega_1 = q * \omega_0$ // q 为加权方法;
- 5) $D \rightarrow E$;
 if ($A \mid S_{1\mathcal{A}} \neq 0 \cup pi \geq r$) // D节点访问信息;
 $D \rightarrow A \mid S_{1\mathcal{A}} \rightarrow e_r \rightarrow E \rightarrow L$ // 重建数据,对服务评价存储;
 $\omega_1 = q * \omega_0$;
 else if ($B \mid S_{2\mathcal{B}} \neq 0$) // B节点在线;
 $D \rightarrow B \mid S_{2\mathcal{B}} \rightarrow e_r \rightarrow E_2\{e_i, \omega_2\} \rightarrow L$;
 // 从B节点及其邻域获取更加可靠的数据块;
 $\omega_2 = c * \omega_1$;
- 6) if $somenode \rightarrow E$; goto 4);
- 7) 根据数据代理记忆算法定期对信息集进行定期更新;

数据代理记忆更新算法的步骤如下:

- 1) if ($\omega' < 0.3$) delete e_1 ;
 // 评价病毒等高危害数据块进行删除;
- 2) if ($0.3 < \omega' < 0.5$) $\psi_1 = e_1$; // 评价低的数据可以缓存;
 if ($\omega'' < 0.3$) delete e_1 ;
 // 再次服务评价过低予以删除;
 else $\psi_3 = e_2$;
- 3) if ($0.5 < \omega' < 0.6$) $\psi_1 = e_1$; // 评价较低的数据缓存;
 if ($\omega'' < 0.5$) $\omega' = \omega''$; goto 1); // 重复1) ~ 2);
 else if ($\omega'' > 0.6$) goto 4); // 可信任,跳转至4);
 else goto 3); $\psi_3 = e_2$; // 否则继续评价;
 if ($\omega''' = 0$) $\psi_7 = e_2$; $\psi_m = e_2$;
 // 再次服务评价判定;
 else $\omega' = \omega'''$; goto 3); // 重复3);
- 4) if ($0.6 < \omega'$) $\psi_1 = e_1$; // 信任信息;

if ($\omega'' < 0.6$) $\omega' = \omega''$; goto 1);

// 再次服务进行评价

else $\psi_n = e_2$;

if ($0.6 < \omega'''$) $\omega' = \omega'''$; goto 4);

else $\omega' = \omega'''$; goto 1); // 重复1) ~ 3) 考核;

5 性能分析

由于大量用户的参与,每个用户可能设置的本地策略 Π 不同,从而会影响P2P网络的稳定性,因而本地策略应由设计者根据安全性原则给出默认值,普通用户使用默认值。为了对本地策略 Π 的取值范围进行说明,根据一般安全原则,给出了分析图,设计者可以根据加权结果来测定具体数值。

本模型取4个语言变量 y, k, w, z 对服务评价进行分析,其权值均取25%,具体如图3所示。

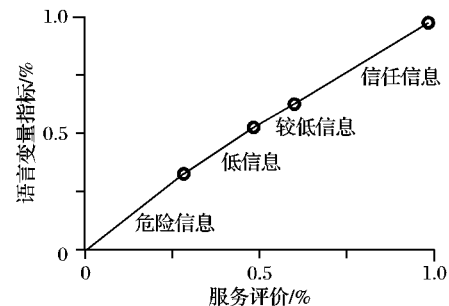


图3 服务评价分析

当信息发布后,信息首先被分成若干个数据块,分布式代理存储于邻域节点。此时后继节点访问时需要和数据块进行搜索。由于本地资源列表存储着大量数据块信息,所以后继节点只需要搜索到 r 个数据块, r 是小于 i 个数据块总数的一个常数,其值大小受本地资源列表所存储的信息量制约,其搜索效率受到数据块可用性、可靠性 k 、搜索最优 z 和数据的完整度 w 等指标的综合影响,它们的阈值一般设定为0.5,0.6,0.5,0.3。当它们均低于此值时,数据服务评价低于0.5,此时对数据进行缓存或者清除;否则,数据安全可用。

由于节点的活跃性越高,节点被评价的次数就越多,进而所存储的数据块共享度越大,综合服务评价趋于可靠,有效地监管了数据的安全性,增强了系统的稳定性。当用户需要搜索到尽可能多且共享度不高的信息时,可对阈值进行修改。

(下转第1497页)

字,故每一个节点共需要 n 位存储开销。

2) 通信开销。任意两个节点之间可直接建立链路,设其每一跳的通信开销为 1,则对距离为 L 的任意两节点,其通信开销为 L ,故其平均通信开销也为 L 。

4 仿真实验

4.1 任意两个节点之间建立直接密钥的概率的仿真实验

在簇内的任意两个节点之间,由于其编码的独特性,必定存在不同位的编码,因此任意两个节点之间建立通信链路的概率为 100%。

4.2 通信开销的仿真实验

对随机密钥模型和 DNA 模型在密钥建立过程中的通信开销仿真实验数据比较如图 2 所示。

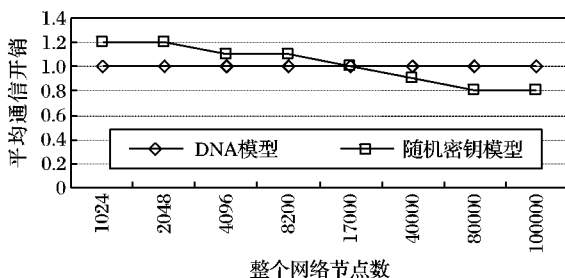


图2 密钥建立过程通信开销的比较

从图2可以看出,当节点数目比较少时,DNA模型的通信开销小于随机密钥模型,原因在于DNA模型中节点之间的密钥路径建立只需要一次通信即可,为了提高安全性能,随机密钥模型在路径建立过程中,不是一次性建立路径,而是通过互相产生随机数字来进行新的密钥生成,在簇中节点之间的第一次通信中,将多进行2次通信,相对于其安全性能的提高,簇中节点之间通信开销的少许增加是可以接受的;当节点数目的急剧增加时,随机密钥模型的通信开销小于DNA模型,因为随着节点数增加,簇间通信大大增加,随机密钥模型在簇间通信的直接一次路径建立概率得到了提高。

5 结语

在基于DNA模型的对偶密钥建立算法基础上提出了一

种新的基于对偶编码的随机密钥建立算法。理论分析和仿真实验结果表明,新算法有效提高了任意两个节点之间直接对偶密钥建立的概率和的安全性能,同时,也在簇与簇之间节点的通信中节省了通信开销。因此,可以认为这是一种性能更好的传感器网络密钥建立算法。

参考文献:

- [1] POTTIE G, KAISER W. Wireless sensor networks [J]. Communications of the ACM, 2000, 43(5): 51-58.
- [2] ESTRIN D, GOVINDAN R, HEIDEMAN J, et al. Next century challenges: Scalable coordination in sensor networks [C]// Proceedings of the ACM/IEEE International Conference on Mobile Computing and Networking. New York: ACM Press, 1999: 263-270.
- [3] AKYILDIZ I F, SU W, SANKARASUBRAMANIAM Y, et al. A survey on sensor networks [J]. IEEE Communications Magazine, 2002, 40(11): 102-114.
- [4] EESCHNAURE L, GLIGOR V D. A key-management scheme for distributed sensor networks [C]// Proceedings of the 9th ACM Conference on Computer and Communication Security. New York: ACM press, 2002: 41-47.
- [5] CHAN H, OERRIG A, SONG D. Random key predistribution schemes for sensor networks [C]// Proceedings of IEEE Symposium on Research in Security and Privacy. Los Alamitos: IEEE Computer Society Press, 2003: 191-213.
- [6] BLANDO C, DESANTIS A, KUTTEN S, et al. Perfectly secure key distribution for dynamic conferences [C]// Proceedings of the Advances in Cryptology. Berlin: Springer-Verlag, 1993: 471-486.
- [7] LIU DONGGANG, NING PENG, LI RONGFANG. Establishing pairwise keys in distributed sensor networks [J]. ACM Transactions on Information and System Security, 2004, 20(10): 1-35.
- [8] 蔡立军,王雷,林亚平,等.传感器网络中基于DNA模型的对偶密钥建立算法研究[J].电子学报,2008,36(1):171-176.
- [9] 胡宇舟,王雷,陈治平,等.传感器网络中对偶密钥的动态密钥路径建立机制及算法[J].通信学报,2008,29(2):52-58.
- [10] WANG G, CHO G. Compromise-resistant pairwise key establishments for mobile Ad Hoc networks [J]. ETRI Journal, 2006, 28(3): 375-378.

(上接第1485页)

对于数据的完整度而言,一般情况下会影响数据块的存储转发,所以可以进行适当的修改,以获取更多的稀缺资源。

6 结语

本文结合数据复制,分布式代理存储和服务评价模型等方面的研究成果,提出了一种分布式代理记忆机制。通过对本地策略的设定,可以有效地阻止病毒等服务评价低的文件的传播;对数据分布式代理记忆的信息的动态更新与遗忘,可以减缓数据存储压力,减少带宽需求;当用户只下载不上传时,其余节点无法对其本身节点进行服务评价和代理记忆,熟识度降低后会被列入遗忘节点队列,提供服务也会逐渐减少,这样可以激励用户上传资源。一旦访问节点开始资源搜索时,等同于多个节点提供信息资源,搜索更加便捷,从而传输更加快速;局域网用户访问数据时不需要远距离搜索。对需要保护的数据,用户发布时可将数据安全性划分为隐私、保护和公开3个等级以对数据进行安全性约束,当选择隐私级别时,后继节点和邻域节点不进行代理记忆;当用户选择保护

时,只有邻域节点代理记忆;在公开级别时默然为节点信息互访不受限制,只根据本地策略进行数据代理记忆。记忆的隐私程度越低,传播程度越广。

参考文献:

- [1] 周文莉,吴晓非. P2P技术综述[J]. 计算机工程与设计, 2006, 27(1): 76-79.
- [2] 王意洁,张小明,周婧. P2P系统中数据复制算法研究[J]. 国防科技大学学报, 2007, 29(3): 61-70.
- [3] 刘翔,汪海玲. 分布式存储中的一种数据放置策略[J]. 计算机与数字工程, 2009, 37(5): 27-29.
- [4] 赵恒,权义宁,胡予濮. 一种新的P2P数据共享解密授权方案[J]. 西安电子科技大学学报:自然科学版, 2005, 32(5): 781-785.
- [5] 李佳伦,谷利泽,杨义先. 一种新的P2P网络的信任管理模型[J]. 北京邮电大学学报, 2009, 32(2): 71-74.
- [6] 窦文,王怀民,贾焰,等. 构造基于推荐的Peer-to-Peer环境下的Trust模型[J]. 软件学报, 2004, 15(4): 571-583.