基于 Region 的多层结构 Peer-to-Peer 网络模型与搜索算法研究*)

乐光学[1,2]　李仁发[2]

(怀化学院计算机系　湖南怀化 418000)[1]　(湖南大学计算机与通信学院　长沙 410082)[2]

摘　要　以小世界模型为理论基础，以 Region 为基本逻辑管理单位、按用户需求和共享目的组织 Region。提出了基于 Region 的多层结构 Peer-to-Peer 网络模型和构造规则，给出了 Region 的划分策略和数学模型，证明了模型的正确和合理性；对模型中的层和域、中心节点、普通节点和汇聚点进行了明确的定义，给出了节点加入、离开、中心节点选取策略和算法描述；使定位某种服务的工作量和查询范围从网络中的所有结点数降低到 Region 的节点数，有效地防止了恶意请求引发的洪，网络系统开销为常数。模拟分析表明，该模型可有效解决可扩展性、性能与效率不高问题，且网络规模越大，其综合性能的优越性越明显，因此，模型是合理有效的。

关键词　对等网，层和域，中心和普通节点，多层结构，搜索包扩散

Study on Peer-to-Peer Network Model and Search Algorithm with Multi-layer Architecture Based on Region

YUE Guang-Xue[1,2]　LI Ren-Fa[2]

(The Department of Computer Science, Huaihua Institute, Huaihua, Hunan 418000)[1]

(College of Computer and Communication, Hunan University, Changsha 410082)[2]

Abstract　By using the "small world" model as the theoretical foundation, and in the light of the users' requirement and a shared organization Region of logic manage-unit, puts forward multi-layer architecture Peer-to-Peer Network model based on Region, provides a classifying policy and a mathematic mode of Region, which proves the exactitude and rationality of the model. At the same time, it makes an explicit definition topology, Leader, Host, Layer, Region and rendezvous point of the system structure. It also gives algorithm policy about the new host joins, host departure and Leader selection. It has brought about a number of nodes locating and querying service down to that of Region, so as to effectively control the request Flood produced by network, and the control overhead of network system tends to be a constant. Simulation results show that it could effectively solve the above problems, and the larger the network size is, the more obvious the superiority of its comprehensive performance is, so the model is reasonable and effective.

Keywords　Peer-to-Peer network, Layer and region, Leader and host, Multi-layer architecture, Diffusing of search packet

1　引言

P2P(Peer-to-Peer)网络中所有的节点是对等的，对等机具有相同的责任与能力并协同完成任务，兼有客户机和服务器的功能，对等机通过直接互连实现计算机资源和服务的全面共享，消除了信息资源孤岛和 C/S 模型中的服务瓶颈问题[1]。从技术上讲，P2P 的拓扑结构有 3 种不同的形式[2]。

(1)以 Napster 为代表的中心文件目录/分布式文件系统结构模式，如图 1，通过中央服务器进行目录管理，实现文件共享，数据传输主要在节点之间进行，避免了网络的拥塞。但仍存在单点瓶颈问题。

(2)以 Gnutella 为代表的纯 P2P 模式，如图 2，系统没有中间服务器，接近于绝对的自由。这样形成的 P2P 网络很难进行诸如安全、身份认证、流量等控制。

(3)基于超级节点的两层结构 P2P 虚拟网络，是 Napster 和 Gnutella 模型的折中，如 Sun 推出的 Project JXTA 2.0 Super-Peer Virtual Network[3,4]，如图 3。通过分布式文件系统，建立完全开放的可共享文件目录，运用相对的自由来兼顾安全和可管理性。

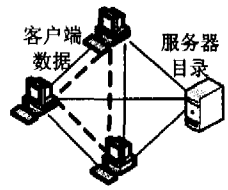

图 1　中心文件目录/分布式文件系统

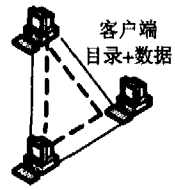

图 2　纯 P2P 系统

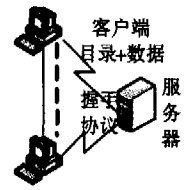

图 3　中间服务器/分布式文件目录系统

2　基于 Region 的多层次结构 P2P 网络模型描述

网络体系结构模型是构建 P2P 系统的基础，与 P2P 系统所提供的功能和整体性能直接相关。构建网络模型应遵循：(1)尽量减少系统中需要远距离交互的对等点数目，检索尽可能少的对等点就能够满足尽可能多的请求，以降低每一检索过程在网络中产生的总负载量，避免引发请求洪；(2)易于扩

*)基金项目：本文受到国家自然科学基金项目资助(60273075)；湖南省教育厅自然科学基金项目重点资助(03A036)。**乐光学**　硕士，副教授，主要研究方向为：网络技术与分布式计算，虚拟技术。**李仁发**　博士，教授，博士生导师，主要研究领域为：网络技术与分布式计算，嵌入式计算。

展、鲁棒性高、有利于信息传输，系统维护和控制开销尽量小；(3)有利于数据流向对其更感兴趣的区域(即请求活跃区)。结合 P2P 和 C/S 结构各自的特点，以小世界模型[5]为理论基础，提出了以域(Region)为基本逻辑管理单位，按用户需求和共享目的组织域，将网络中的对等点按域进行划分的多层结构模型。网络拓扑由 m 层(Layer)构成，以 $L_i(i=0,1,\cdots,m)$ 表示，每一层由 $n(n=1,2,\cdots,n)$ 个域组成，每个域含 $k(k=1,2,\cdots,k)$ 个节点或主机(Host)，网络中的任一主机必属于一个特定的域，域中节点以其性能和在网络中的角色分为中心节点(Leader)和普通节点(Host)，最高层只有一个节点，称为汇聚点 PR(Rendezvous Point)。从第一层开始，第 i 层的节点由第 $i-1$ 层的中心节点组成，以此类推，就构成了多层次结构的 P2P 网络拓扑。对任意网段抽象可得如图 4 所示的网络体系结构，其中 h 为请求加入节点。严格意义上讲，该结构属于图 3 所示的 P2P 模式。并简称为 RLP2P。

因此，分层就是构造一种逻辑拓扑结构，使域内节点有很大的相似系数，平均距离很小，赋予那些性能高的节点更多的职责，以提高整个网络的性能，实现信息查询的快速定位和资源的全面共享。

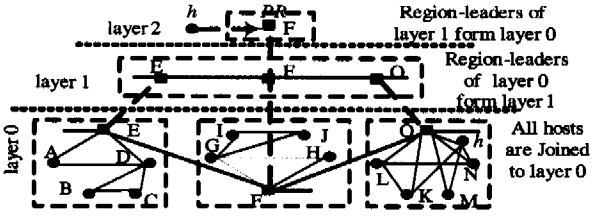

图 4　RLP2P 网络拓扑图

2.1　相关定义

2.1.1　中心节点(Leader)　由域内性能最高节点充当，位于域的中心，与域内节点距离最短，是域的对外端口，类似于 C/S 结构中央服务器和 JXTA 中的集合点[3,4](Rendezvous Peer)，但服务功能已被弱化，负责域内节点的管理，完成对等点的"握手"、工作结果的备份和归档，提供：管道、成员资格、访问、发现和解析器服务。中心节点维护着一个域间和一个域内活动节点的信息列表，节点加入，须向中心节点注册包括节点名和地址等信息；负责收集和反馈域内外的状态信息，及时处理节点的加入和离去，更新服务和网络状态信息。

2.1.2　普通节点(Host)　类似于普通 C/S 结构中的客户机和 JXTA 中的普通节点(Simple peer)[3,4]，但功能已被加强。同域内节点的性能和任务特征基本相似；提供简单的管道和访问等服务，实现对等发现、对等解析器、对等成员资格、对等信息、管道和端点路由协议，构成一个信息资源和计算中间结果全面共享的 P2P 计算网络。节点请求加入或离开网络时，由中心节点向所辖 P2P 计算网络发布消息，建立或撤销连接，更新资源和服务目录，普通节点通过中心节点实现跨域访问。

2.1.3　域(Region)　是按节点对资源需求、共享目的和任务的相似性进行划分的逻辑管理单位，是由中心节点所辖区域内多个属性相似、性能相对较弱的主机组成的一个闭集。域内节点有很大的相似系数，在域内找到满足服务的概率很大，节点跨域请求服务的概率很小。

域相当于 JXTA 中对等组的作用，作为网络分区协议(Network Partitioning Protocol)，它确保多播只被中继到能访问这消息的域的成员[3,4]。消息在中心节点间实现传递并依照自身及其域内节点的内容决定是否响应查询请求。理论上中心节点和普通节点在搜索、资源和信息共享上处于同等位置，但中心节点同属于两个不同的层和域，为临界点，是域间的接口，其任务特征和性能比节点更广。域的规模为，$k\leqslant RL_i(x)\leqslant 3k-1$，$k$ 为常数，任一时刻，域内至少有一个中心节点；若中心节点因故障或异常退出网络，其所辖域内节点在新中心节点产生之前，将暂时与网络断开。为增强网络的鲁棒性，可在域内增加一个冗余的"准"中心节点，或设计一个产生中心节点的代理(Agent)运行于网络系统中，当中心节点失效时，系统及时启动该代理，在域内选取一个新的节点取代失效的中心节点。域内中心节点和节点依据性能高低，可实现角色互换。

2.1.3　汇聚点 RP(Rendezvous Point)　为协议模型中最上层的节点(如图 4 中的 F)，类似于网络中的路由，保存有网络中每个节点提供的服务和状态信息列表，负责引导节点加入和消息路由，对整个网络的节点进行管理和控制。

3　协议描述与性能分析

3.1　域的划分策略

域的划分是为了便于对网络的控制和管理，实现网络资源的全面共享和工作协同。以前述三条准则为基础，按节点对资源需求、共享目的和任务相似性进行域的划分。通过设置节点属性相似系数 $\mu(0.7\leqslant\mu\leqslant1)$ 来决定该节点属于何域，以避免普通节点同属于多个域的情况发生。每层的节点分属于不同的域，域的规模为 $RL_i(X)$，$K\leqslant X\leqslant 3K-1$(即域的边界)，其中 X 为域所含的节点数，K 为常数。当 $RL_i(X)>3K-1$，将域等分为两个域；当 $RL_i(X)<K$，合并域。任一时刻，域至少有一个中心节点。如图 4 所示，根据域的划分策略把最低层的节点划分为[$ABCDE$]、[$FGHIJ$]、[$KLMNO$]三个域，构成第 0 层，记为 L_0，其中心节点分别为 E、F、O 组成一个域，构成第 1 层，记为 L_1，依此类推，就构成了一个多层次结构的 P2P 计算网络。

域划分边界取值 $K\leqslant X\leqslant 3K-1$ 的正确合理性证明。

(1) 设域的上界取一个较小的值，如 $RL_i(x)\leqslant 2K-1$。当 $RL_i(X)>2K-1$ 时，域被等分为两个域，$RL_i(X)=K$；任意时刻，若此两域中任一域有 $n\geqslant 1$ 个节点离去或失效，使 $RL_i(X)<K$，两个刚划分的域又需合并；若有 $n\geqslant K$ 个节点加入，使 $RL_i(X)>2K-1$，域又需划分为两个，导致网络波动和管理开销增大；

(2) 设域的上界为 $RL_i(X)\leqslant 3K-1$。当 $RL_i(X)>3K-1$，域被等分为两个域，$RL_i(X)=3K/2$；任意时刻，只有当此两个域中任意一个域有 $n\geqslant K/2$ 个节点离去或有 $n\geqslant 3K/2$ 个节点加入时，系统才进行域的合并和划分操作，使网络保持相对的稳定。因此，$RL_i(X)\leqslant 3K-1$ 可避免域的频繁分割和合并；

(3) 设域的上界取一个较大的值，如 $RL_i(X)\leqslant 4K-1$。域的规模增大，其划分和合并操作相对减少，但系统的管理和控制开销急剧增加，加重中心节点的负担，引发信息瓶颈，甚至导致中心节点失效，使系统的交互性和鲁棒性急剧下降。

域的划分和合并算法描述如表 1、2 所示。

表 1　域的划分算法描述

Procedure:*RegionSplit*(C)
{|C|≥3K};/* |C|=$RL_i(X)$ */

$\partial \Leftarrow \{Q/Q \subset C \cap |Q|, |C-Q| \geqslant |\frac{3K}{2}|\}$
Let $R(Q)=\max(\text{radius}(Q), \text{radius}(C-Q))$
Find Q^* s. t. $R(Q^*) \leqslant R(Q)$; /where $Q, Q* \in \partial$/
LeaderTransfer(Ldr(C), Q^*, Ldr(Q^*))
LedderTransfer(Ldr(C), $C-Q^*$, Ldr($C-Q^*$))

表 2　域的合并算法描述

Procedure: *RegionMerge*(C)
{$|C| < K$ and L_i is the layer to which C belongs}
$l \Leftarrow$Ldr(C)
Find y s. t. dist(l, y) < dist(h, x), $x, y \in RL_{i+1}(l)$
RegionMergeRequest(l, y, L_i)
LeaderTransfer(l, C, y)

3.2　节点加入网络及调整

定义　任意时刻位于 L_i 层的节点 h 请求加入网络，h 向汇聚点 RP 发出加入请求，将 h 的属性和特征状态信息传给 RP，RP 在所辖域内确定一个距它最近且任务属性最相似的响应节点，h 查询该域中的所有节点，以确定 h 的加入点。重复上述过程，直到找到 L_0 层中的某一个合适的域加入，系统更新服务、信息列表和特征状态信息。

设 h 请求加入网络，h 向汇聚点 RP 发出加入请求，如图 4 节点 F，将 h 的属性和特征状态信息传给 RP，RP 在所辖域内确定一个响应节点，h 查询该域中的所有节点，以确定 h 的最相似和最近点。h 在 RP 的引导下搜索 L_1 层的所有成员中，发现其任务属性与 O 最相似且距离最近，h 向 O 发出加入请求，O 响应请求，h 向 O 注册包括节点名和地址等信息，O 告知 h 在 L_0 层的所辖域中有节点 K、L、M、N，h 向网络发布能提供的服务和特征状态信息，并与节点建立对等连接，系统更新服务、信息列表和特征状态信息，h 加入完成。若域中的节点发现网络中另外一个域的特性更适合自己，节点可向系统发出请求，自动进行调整。

节点 h 请求加入网络和节点 Z 从当前域转移到与自己属性更相似的域的算法描述如表 3、4 所示。

表 3　节点 h 请求加入网络算法描述

Procedure: *BsicJoinLayer*(h, i)
$RL_j \Leftarrow Query(RP, -)$
While($j > i$)
Find y s. t dist(h, y) $\leqslant$ dist(h, x), $x, y \in RL_j$
$RL_{j-1}(y) \Leftarrow Query(y, j-1)$
Decrement j, $RL_j \Leftarrow RL_{j-1}(y)$
Endwhile
JoinRegion RL_j

表 4　节点 Z 转移到与自己属性更相似的域的算法描述

Procedure: *RegionRefind*(Z)
{L_l is highest layer to which z belongs}
$l \Leftarrow Ldr(RL_l(Z))$; $C \Leftarrow RL_{l+l}(l)$
Find y s. t. dist(z, y) < dist(z, x), $x, y \in C$
if($y \neq l$)
LeaveRegion(z, l, L_l)
JoinRegion(z, y, L_l)
endif

由上分析知，RLP2P 网络模型自然满足构建网络模型应遵循的规则，并具有特点：

(1)每个层分为若干个域，每个域的规模为 $K \sim 3K-1$ 台主机；

(2)每个域至少有一中心节点。每层的所有中心节点形成上层的节点，所有主机最初都属于初始层 L_0；

(3)任一主机在任何层只能存在于一个域，但中心节点同属于相邻层中不同的两个域；

(4)如果一主机处于 L_i 层的某个域中，它一定是 $L_0 \cdots\cdots L_{i-1}$ 层的某个域中的中心节点；

(5)如果一主机不在 L_i 层中，也必不在 L_j 层中，其中 $j > i$；

(6)网络最多有 $\log_k(N)$ 层，最高层只有一个成员；

(7)路径搜索采用分布式向前咨询方式完成。

3.3　节点离开、中心节点选取策略

若节点 h 正常离开，它把离开信息在域内广播；若 h 因失效或异常离开，h 发不出离开信息，则系统通过 h 的“心率信息”来确认。若正常离开节点 h 是中心节点，则根据服务和特征状态信息，直接推举一个节点，承担中心节点之责，将网络服务、资源和状态信息直接传给新的中心节点，并做标记；若中心节点因失效或异常离开，则域内的节点暂时与网络断开连接，系统运行产生中心节点的代理(Agent)，在域内选取一个综合性能高的节点为新中心节点，重新建立系统新服务、资源和状态信息表，并做标记。新中心节点的选取操作如图 5 所示。

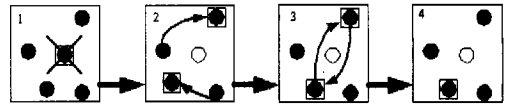

图 5　中心节点产生示意图

3.4　网络模型的性能分析

在该模型中，每个域的节点数 X 为：$K \leqslant X \leqslant 3K-1$，网络规模为 N。一般，L_i 层中任一节点与 $L_0, \cdots, L_i$ 层的域中其他节点控制信息的开销为 $O(K)$，域内节点控制开销的上界为 $O(N/K^i)$；最高层节点的控制总开销为 $O(K\log N)$。当 N 逐渐增大，网络的控制开销可近似表示为：

$$C \leqslant \frac{1}{N}\sum_{i=0}^{\log_K N}\frac{N}{K^i}K*i$$

$$=O(k)+O(\frac{\log N}{N})+O(\frac{1}{N}) \Rightarrow O(K)$$

因此，对于任一节点其控制开销的平均值为 $O(K)$，最坏情况为 $O(K\log N)$。任意两对等点间的交互开销为 $O(\log N)$，当节点交互开销达到 $O(K\log N)$时，则需减少域规模 K 的值。

由上分析可知，域中的节点具有“小世界”的特性，即网络拓扑聚集度高、特征路径长度小的特性[5]。以域为基本逻辑管理单位，将大规模节点在逻辑上划分成小区域，消息经分级扩散，每个结点掌握着大量的域内节点服务信息，使域内任意两个节点都有很小的平均距离，节点间有很大的相似系数，域文件能够很大程度代表域内部提供的服务信息。因此，在域内部找到满足服务请求的概率很大，节点出现跨域请求服务的概率很小，使定位某种服务的工作量、查询范围从网络中的所有节点数降低到域内的节点数，有效地防止了请求洪。在不增加网络管理开销的情况下，使消息的扩散效率和网络的鲁棒性得到大幅度提高。

文[5]已论证分布式算法路由链上界为 $O(\log\sqrt{n})^2$，其中 n 为网络节点规模。设网络系统中节点和服务是平均分布的，则 Gossip 算法在高概率下最多经过 $O(\log\sqrt{n})^{1+\varepsilon}$ 次消息扩散到达距离其$\sqrt{n}$的任意其他结点[6]。本文提出的网络模型中，将大规模网络消息扩散划分成小规模消息扩散，在域规模相同(节点数为 m)的理想状况下，算法路由链上界为 $O(\log\sqrt{n/m})^{1+\varepsilon}$，其扩散效率明显优于 Gossip 算法。

为了减少节点的加入延迟，当 RP 收到加入请求时，将请求加入节点视为一个临时节点，进入当前层，但不属于任何域，由当前层的中心节点向下发出请求。如图 4，h 在加入到

K、L、M、N 节点所在的域之前，h 是 F 的一个临时成员。为了提高效率，RP 需要维持一个加入请求响应信息表。

4 搜索包扩散路由算法与性能分析

4.1 搜索包扩散路由策略

基于域的多层结构 P2P 网络模型中节点具有“小世界”高聚集性特征，为了实现高效的服务定位和搜索最小冗余扩散，使控制开销最小，规定节点维护的路由表，除了保存与其邻居节点的连接关系信息外，还必须保存与其相邻节点的邻居节点的节点聚集连接关系，即节点必须了解其邻居节点的邻居连接关系信息。通过与相邻节点间周期性地交换邻居连接关系查询包或新节点加入时刷新信息表。因此，每个节点都知道谁与它的邻居节点直接相连，利用这种分布式本地节点信息列表，实现数据“扩散”冗余最小。最小冗余数据扩散算法描述如下：

Neighbor：节点 n 的邻居节点的集合；

Connection：已知的有连接关系的节点对(x,y)的集合；

Receive：已接收到的搜索消息的标识(ID)集合；

Temp：已接收到搜索消息的节点的索引集合，若 $i\in$ Temp，表示节点 b_i 已收到消息。

当节点 n 收到一个来自节点 a 的消息 m 时，做以下操作：

```
If (ID(m)∈Receive) ignore;/*说明该搜索消息原来已收到过，则忽略该消息。*/
else { for (对于节点 n 的每一个邻居节点 b_i∈Neighbor)
    {if (a=b_i) {b_i⇒Temp;标识节点 b_i 已收到消息 m;}
    else { if ((a, b_i)∈ Connection) b_i⇒Temp;标识节点 b_i 已收到消息 m;}
      /*节点 b_i 与节点 a 直接相邻，b_i 将 a 收到搜索消息 m，不需要经过节点 n 的转发。*/ }
    for (对于节点 n 的每一个邻居节点 b_i∈Neighbor)
        {for (对于节点 b_i 的每一个邻居节点 b_j，
              且(b_i,b_j)∈Connection)
    if (j∈Temp) {有两条路径可以将消息 m 转发给节点 b_i；一条经节点 b_i 转发，一条经节点 n 转发，则规定：
        if (节点 b_j 地址<节点 n 地址){节点 b_j 转发;b_i⇒Temp;标识节点 b_i 已收到消息 m;}
        else{节点 n 转发;b_i⇒Temp;标识节点 b_i 已收到消息 m;}
            节点 n 向节点 b_j 发出“心率查询”信息;
if (节点 b_j 不存在){节点 n 转发;b_i⇒Temp;
    标识节点 b_i 已收到消息 m;}}}
for (对于节点 n 的每一个邻居节点 b_i∈Neighbor)
    {if (b_i∉Temp) {转发搜索消息 m 到节点 b_i;}}
```

4.2 算法的性能分析与模拟

如图 6 所示，当节点 A 收到搜索包，且 A 不满足请求，则 A 将之扩散转发到节点 B、C、D，设 B、C、D 也不能满足请求，在 Gossip 算法中，B、C、D 分别相互转发，共产生 9 个包，其中 6 个是冗余搜索包。文[7]已论证：随着节点间连接的增加，网络中的消息冗余量呈指数增加，在具有 n 个连通节点的 Gnutella 网络中，一条消息最多可产生 $P_2^n-(n-1)$ 个冗余消息，这些冗余消息将耗费大量的机器处理时间、吞噬网络带宽。在本文提出的最小冗余扩散算法中，由于节点 B、C、D 彼此知道与节点 A 直接相邻，均直接收到从节点 A 发来的搜索包；且节点 B、C、D 知道彼此直接相邻，所以节点 B、C、D 不会相互转发该搜索包。但如果节点 A、D 联接失效或断开，按照规则，由于 C 知道 B 与 A 直接相邻，而 D 又与 B 直接相邻，D 必定可以收到 A 经 B 转发来的广播搜索包，所以 C 不向 D 转发；同理，B 也不会向 D 转发；这样就出现了 B、C 都不向 D 转发搜索包的错误结果。算法规定：若节点 B 的地址＜节点 C 的地址，则由 B 向 D 转发搜索包，否则，由 C 向 D 转发搜索包；为了防止 B 此时正好发生失效或离开网络，且 C 又未收到 B 离开网络的消息，至使 D 得不到搜索包的情况发生，规定：C 在做不转发搜索包之前，向 B 发出“心率检测”信息，以确定 B 是否存在，若 C 在规定的时间内未收到 B 的检测响应，则 C 认为 B 已离开网络，由 C 向 D 转发搜索包；但这可能因 B 此时正在忙于其他事务的处理而未能及时响应 C 的“心率查询”，B 并没有离开网络，B 一旦将事务处理完毕，立即向 D 转发搜索包和响应 C 的“心率查询”，这时会产生一个冗余搜索包。由于域是按照文中提出的准则进行构造，搜索绝大部分是在域内进行，满足概率很高，且域的规模 K 是一个常数，在网络中出现节点未能及时响应“心率查询”的情况极少，搜索包在域内的转发数趋于理想状态，冗余包趋于一个很小的常数，因此这种冗余是很有限的。且这种极有限的冗余对保证网络信息的高效搜索是有益的。

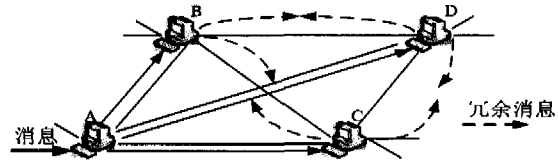

图 6 Gnutella 协议消息扩散示意图

由上分析可知，该算法能极大地减少的网络流量。以 4 个节点的全连通图为例，如图 6 所示。该算法仅转发了必要的三个搜索包，仅是 Gossip 算法转发搜索包数量的三分之一。且网络的聚集性越高，则该算法的有效性越强。本文提出的“基于域的多层结构 P2P 网络模型”中，域中的节点具有“小世界”模型的高聚集性，因此，随着网络规模的扩大，该算法能有效地降低冗余搜索包的产生，减轻了网络的压力，避免了由搜索包转发而引发的请求洪，极大地增强了网络的扩展性和鲁棒性。由算法可知，其时间复杂度为 $O(n^2)$，所以整个算法的时间复杂度为 $O(n^2)$，其中 n 为某节点 X 的邻居数，$K\leqslant X\leqslant 3K-1$，一般情况下其值并不太大。

5 仿真和分析

仿真环境：联想万全服务器 T100 一台；11 台 PC 机 IBM8624；路由器：HUAWEI Quidway R2501E；交换机：Fast Ethernet ES-3124RL 24-port 10/100Mbps；软件：Linux9.0，Windows2000，JAXT2.0；用 nem[8] 产生仿真网络拓扑，拓扑生成算法为 PLOD[9]；单独 1 台 PC 用于网络协议和流量分析。仿真网络模型为 RLP2P 和 Gnutella，提供的服务类型为：新闻、体育、娱乐、音乐、文学、财经 6 个类；网络规模 $N\leqslant 1500$，域的规模 $RL_i(x)=64\sim256$，节点的度$\geqslant3$，网络状态信息更新频率为 5s，TTL=7s，为 3 层逻辑拓扑结构，连接相对稳定，数据源端节点以一固定的速率连续产生数据流，数据包高效有序地在网络中传输，且在给定的生命周期内均能被捕获，节点可随机加入和非正常离开网络。

图 7，8 为对 RLP2P 和 Gnutella 网络连续 12 小时的流量监测，RLP2P 的网络流量基本维持在 20kbps，Gnutella 网络的流量随时间的增加而增加，到达监测终时时，其流量已达到 39kbps。

图 9 为网络查询平均耗时，$RTT_{RLP2P}\cong0.682RTT_{Gossip}$，$RTT_{RLP2P}$维持在 15ms，$RTT_{Gossip}$维持在 23ms。

图 10 为平均带宽占用量，当 N$\geqslant$700，Gnutella 平均占用带宽已达到 200kpbs，且呈急剧上升趋势；RLP2P 平均占用带宽基本维持在 180kpbs，约为 Gnutella 的 76%。

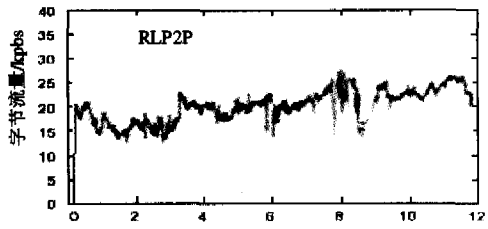

图 7　RLP2P 网络 12 小时的流量分布

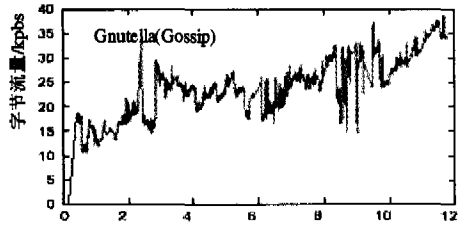

图 8　Gnutella 网络 12 小时的流量分布

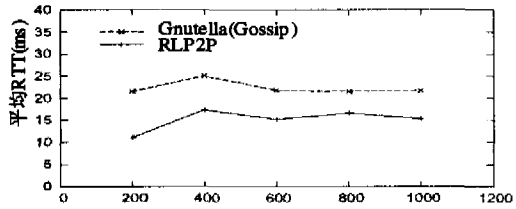

图 9　查询返回平均时间

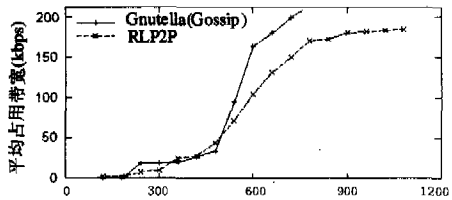

图 10　平均带带占用量

仿真结果表明：当网络规模较小时，泛洪搜索算法 Gossip 在 Gnutella 网络中的综合性能非常优异，但随着网络规模的扩大，连接拓扑中“环”的数目增加和泛洪式消息扩散，使网络中的冗余消息呈指数级增加，消耗大量的处理时间和网络带宽，随着这部分节点失效，Gnutella 网络被分片，使查询访问只能在网络很小的一部分进行，导致网络的可扩展性和搜索性能下降。本文提出的搜索算法对改善 Gnutella 网络的搜索性能非常有效。

结论　结合了传统集中式网络易于管理与分布式网络具有良好的区域自治、负载平衡以及健壮性的优点，以域为基本逻辑管理单位，将大规模节点在逻辑上划分成小区域，使域内任意两个节点都有很小的平均距离，节点间有很大的相识系数，在域内部找到满足服务请求的概率很大，节点出现跨域请求服务的概率很小，使定位某种服务的工作量、查询范围从网络中的所有节点数降低到域内的节点数，将大规模网络消息扩散划分成小规模消息扩散，在域规模相同的理想状况下，算法路由链上界为 $O(\log \sqrt{n/m})^{1+\varepsilon}$（$m$ 为域内节点数），其扩散效率明显由于 Gossip 算法；网络系统控制开销为常数 $O(K)$；提出的搜索包最小冗余扩散算法使扩散冗余包趋于一个很小的常数，算法时间复杂度为 $O(n^2)$，极大地减轻了网络的压力，有效地防止了请求洪，提高了网络的鲁棒性和扩展性，具有更高的综合性能指标。

参 考 文 献

1　Parameswaran M, Susarla A, Whinston A B. P2PN Networking: An Information Sharing Alternative. Computing Practices, Jule 2003. 1～8

2　Parameswearan M, Susarla A, Whinston A B. P2P networking: An Information-Sharing Alternative [J]. Computer, 2003. 34 (7):31～38

3　Traversat B, Arora A, et al. Project JXTA 2.0 Super-Peer Virtual Network. http://www.jxta.org/project/www/docs/JXTA2.0protocols1.pdf. 2004/08/10

4　Super-Peer Architectures for Distributed Computing. http://www.fiorano.com/whitepapers/superpeer.pdf. 2004/08/10

5　Kleinberg J. The Small-World phenomenon: An algorithmic perspective. ACM Symp on Theory of Computing, 2000. 820～828

6　Kempe D, Kleinberg J, Demers A. Spatial gossip and resource location protocols. In:Proc. of the 33rd ACM Symp on Theory of computing [C], Crete, Greece, 2002. 163～172

7　Gnutella: To the bandwidth barrier and beyond. http://www.dss.clip2.com/gnutella.html. 2004/08/10

8　Magoni D. nem: A Software for Network Topology Analysis and Modeling. In: Proc. 10th IEEE Intl. Symposium on 11-16 Oct. 2002. 364～371

9　Palmer C P, Steffan J G. Generating Network Topologies That Obey Power Laws. Global Telecommunications Conference, 2000. GLOBECOM '00. IEEE, 2000, 1:434～438

（上接第 40 页）

2　Zeng Jiazhi, Xu Jie, Wu yue, et al. Service Unit Based Network Architecture. In:Proc. PDCAT'03, 2003

3　曾家智，徐洁，吴跃，等. 服务元网络体系结构和微通信元系统构架. 电子学报，2004，32(5)：745～749

4　Tanenbaumk A S. 计算机网络(第 3 版). 北京：清华大学出版社，1998

5　RFC1633. Integrated Services in the Internet Architecture: an Overview. 1994

6　RFC2475. An Architecture for Differentiated Services. 1998

7　夏梦芹，易发胜，曾家智. 互联网区域路由 QoS 机制研究. 计算机科学，2004，31(11)：38～39

8　Perry Tang Puqi, Charles Tai Tsung-Yuan. Network traffic characterization using token bucket model. Proc IEEE INFOCOM, 1999

9　Au T M, Mehrpour H. Leaky bucket based scheduling algorithm for real time traffic. Australia National Conference Publication - Institution of Engineers, 1994