基于特征信息定位的 P2P 网络模型:Barnet*

王庆波+, 代亚非, 田 敬, 赵 通, 李晓明

(北京大学 计算机科学技术系,北京 100871)

An Infrastructure for Attribute Addressable P2P Network: Barnet

WANG Qing-Bo+, DAI Ya-Fei, TIAN Jing, ZHAO Tong , LI Xiao-Ming

(Department of Compter Science and Technology, Peking University, Beijing 100871, China)

+ Corresponding author: Phn: 86-10-62751799 ext 8025, Fax: 86-10-62765813, E-mail: wangqb@net.cs.pku.edu.cn

http://net.cs.pku.edu.cn/~wangqb/

Received 2002-10-16; Accepted 2002-12-16

Wang QB, Dai YF, Tian J, Zhao T, Li XM. An infrastructure for attribute addressable P2P network: Barnet. *Journal of Software*, 2003,14(8):1481~1488.

http://www.jos.org.cn/1000-9825/14/1481.htm

Abstract: Barnet, an infrastructure for attribute addressable P2P network, is brought forward in this paper. This infrastructure aims to provide high availability, collaborative, large-scale wide area information service. The architecture, information organization and access model of Barnet are described in this paper. NetShot is a peer-to-peer based Naming, Locating and Lookup algorithm. As proof that NetShot is useful in developing decentralized applications. After discussing the naming, joining, departing and neighboring model of NetShot, this paper proposes the concept of attribute addressable P2P network, and describes the details of lookup and location strategy which is adopted in Barnet.

Key words: Peer-to-Peer; information service; routing model; location technology

摘 要: 提出了 Barnet,一个基于特征信息定位的 Peer-to-Peer 网络模型.该模型的目标是为广域网络构建一个高性能、高可用、协同、负载均衡的海量信息资源服务平台.分别描述了 Barnet 原型系统的构建目标、系统结构、信息资源的组织及信息资源的定位策略.描述了 Barnet 中所采用的一个基于 Peer-to-Peer 的分布式命名、定位、查找算法 NetShot,讨论了 NetShot 中节点的命名、节点加入离开、节点间邻接关系和节点间消息传递方式等基本问题.提出了基于特征信息定位技术的概念,并讨论了在 Barnet 中使用基于特征信息定位技术对具体信息资源进行查找、定位的具体策略.

关键词: Peer-to-Peer;信息服务;路由模型;定位技术

中图法分类号: TP393 **文献标识码**: A

网络技术的飞速发展与迅速普及使其成为现代社会传承文明的重要手段,网络的规模越来越大,连入网络

* Supported by the National High-Tech Research and Development Plan of China under Grant No.2001AA115126 (国家高技术研究发展计划(863))

第一作者简介: 王庆波(1978－),男,黑龙江建三江人,硕士,主要研究领域为移动互联网,分布式存储.

中的设备、计算单元的数量和种类也越来越多.网络本身也因“无处不在的计算”而蕴含了大量的资源,如计算资源、信息资源等,这些信息资源在互联网中存在着大量的复制和冗余.如何组织、定位和传输这些资源从而合理有效地利用它们为人们提供信息资源服务成为人们所关注的焦点问题.

伴随着网络技术的发展,计算模型也相应地从传统的单机计算模型转变为网络计算模型,目前的网络计算模型主要是 C/S 和 B/S 计算模型,这两种计算模型在海量信息的组织、访问等方面都不同程度地存在着如单点服务瓶颈、无法抵抗 DoS 攻击等问题.Peer-to-Peer(P2P)计算模型正是在此情况下,为了解决海量计算单元及其信息资源的合理利用问题而提出的分布式计算模型.在 P2P 计算模型中,系统的所有节点是对等的,各节点具有相同的责任,系统中的各个节点互相协同以共同完成计算任务.

文中提出了一个基于特征信息定位的 P2P 网络模型:Barnet.该网络模型中节点之间的拓扑互联关系、节点的加入、离开以及节点之间消息的路由方式采用的是我们所提出的 P2P 路由模型 NetShot 完成的.该网络模型中的信息资源定位方式采用的是基于特征信息进行定位.这种定位技术避免了传统定位方法的单点瓶颈问题,同时提高了信息资源定位以及访问的效率.

本文第 1 节描述了 Barnet 原型系统的构建目标以及网络结构模型.第 2 节介绍了 P2P 路由模型 NetShot.第 3 节讨论了基于特征信息的定位技术的概念.第 4 节讨论了 Barnet 中数据资源的组织方式以及访问策略.第 5 节分析了相关的研究工作.第 6 节阐述了目前的工作进展情况并进行了总结.

1 Barnet 原型系统的构建目标

信息服务技术是支撑互联网发展的若干技术中的重要组成部分,人们不断尝试新的技术来提供新的信息服务或提高信息服务的质量,但是目前的很多应用服务是彼此孤立存在的,各个应用服务之间不能够有效地共享信息资源,从而造成了信息资源的大量冗余,而且也不能够提高系统的效率.Barnet 构建的目标是为广域互联网提供高性能、高可用的信息服务的一个 P2P 基础计算平台.图 1 是 Barnet 平台的基本层次结构,该平台将融合存储服务、索引服务、共享服务以及代理服务来提供综合的信息服务,并通过各个应用之间的数据共享来提高系统资源的利用率.该平台以 NetShot 作为底层节点之间的 P2P 路由模型,采用分布式的数据资源组织模型以及基于特征信息的信息定位策略来定位和访问相应的数据资源.

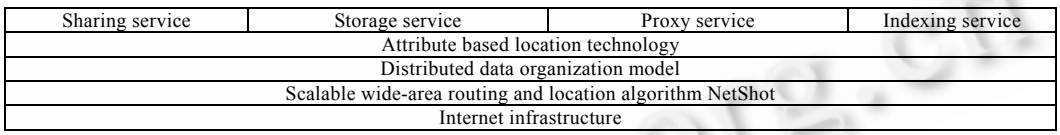


Fig.1 The hierarchy view of Barnet platform

图 1 Barnet 平台的基本层次结构

Barnet 网络模型的功能结构如图 2 所示,在该结构中的计算单元按功能划分为 3 种类型节点:胖节点、瘦节点和外节点.胖节点和瘦节点属于内节点,它们是通过 NetShot 路由模型来确认节点之间的逻辑邻接关系,每个节点在系统中都拥有惟一的逻辑标识,各个节点之间直接通过该标识进行数据通信.胖节点和瘦节点是 Barnet 系统中信息资源的载体,它们为系统提供存储空间、索引空间、代理空间以及共享空间.胖节点与瘦节点的不同之处在于胖节点为外节点提供外部开放的访问标识以及访问接口,外节点在 Barnet 中不存在惟一标识.外节点仅可以选择某些胖节点作为系统资源的访问代理,通过胖节点所提供的开放的访问协议(HTTP,FTP,RMI,RPC 等)对内节点进行访问,从而间接实现对整个系统中的数据资源的访问.Barnet 系统中的内节点也可以通过现有网络应用系统的访问协议来访问现有网络中的数据资源,从而也实现了与现有网络的应用系统的融合.

2 NetShot 路由模型

Barnet 的网络结构模型中维护着大量的节点,节点可以动态地加入和离开系统,系统中任意节点之间可以互相进行通信.NetShot 路由算法是 Barnet 系统中所采用的逻辑网络拓扑结构和路由模型.NetShot 路由模型针对 P2P 网络的动态性为 Barnet 系统提供一个可扩放的动态无冲突节点命名方式、节点间邻接关系、定位以及

查找模型.系统为每一个节点指定惟一的逻辑标识,并通过路由表、引入表来确定节点之间的邻接关系,节点利用这种邻接关系向其他节点传递消息,一个消息通过多个节点的协同传递后到达目标节点,从而实现任意节点之间的通信.在该路由模型中,对于一个具有 n 个节点规模的网络,只需要维护 $O(\log^n)$大小的路由表,节点之间的逻辑路径长度也是 $O(\log^n)$大小的,该路由模型在开销和性能之间取得了较好的结合.接下来简单介绍了 NetShot 路由算法的基本设计:节点命名、加入、离开方式以及节点之间的邻接关系,在文献[1]对 Netahot 路由模型的性能进行了详细的分析.

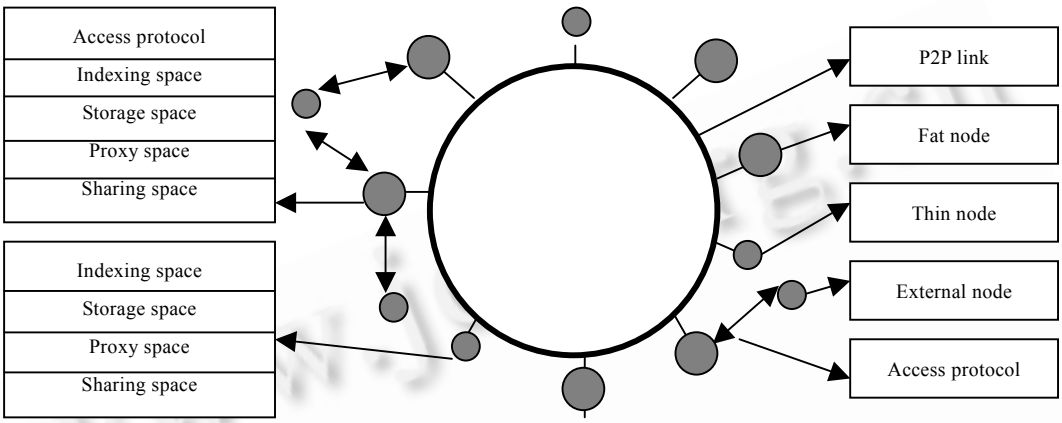


Fig.2 The functional view of Barnet platform

图 2 Barnet 网络模型的功能结构

2.1 节点的命名

如果要实现系统中任意节点之间都可以互相通信,必须为该系统中的每一个节点指定一个全局惟一的逻辑标识,通过该标识来实现从逻辑节点到物理节点之间的映射.在 NetShot 路由模型中节点的全局命名空间 GNS(global name space)是大小为 1 的[0,1]数值空间.系统中的每一个节点都将 GNS 的子空间作为自己的私有命名空间 PNS(private name space),PNS 可以表示为[UCB,VCB],其中 UCB 称为该节点的不变界,VCB 称为变界.在 NetShot 中,UCB,VCB 采用 0 或首位是 0 末位字符是 1 的 0,1 字符串来表示,各个节点的 PNS 所对应的逻辑空间是互不相同的,并将 PNS 定义为该节点的惟一逻辑标识.

对于一个字符串总长度为 K+1 的 xCB(UCB 或 VCB),其所对应的数值 $xCBV$ 为

$$xCBV=\sum_{r=1}^{K+1}xCB_r\times 2^{-(r-1)}. \tag{1}$$

对于一个具有 m 个节点的系统,其各个节点的子空间 PNS_i 之间满足下面的关系:

$$\bigcup_{1\le i\le m}PNS_i=GNS. \tag{2}$$

$$PNS_i\cap PNS_j=\varnothing,\ i\ne j \text{ 且 } 1\le i,j\le m. \tag{3}$$

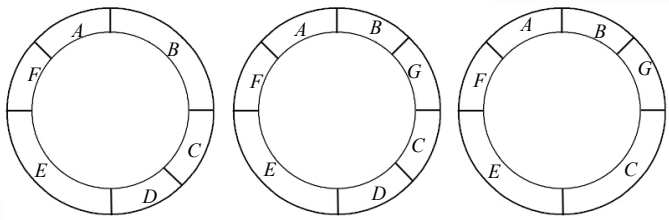
2.2 节点的加入

当一个新节点要加入 NetShot 时,系统通过将系统中某一个节点的 PNS 分割来为该节点分配一个新的 PNS.新节点首先在[0,1]之间随机均匀地产生一个固定长度(例如:128 位)的 0,1 字符串,将这个字符串所对应的数值作为该节点的目标地址,接下来这个新加入的节点至少找到系统中的一个节点作为其引导节点(bootstrap node),并向这个节点发出一个加入请求消息,引导节点根据情况将加入节点请求消息前递(forward)直到目标地址所在的节点.目标地址所在的节点将自己的 PNS 均匀地划分为两个部分,自己保留不变界所在的那一部分,并将另一部分作为新加入节点的 PNS.

2.3 节点的离开

当一个节点要离开 NetShot 时,它首先向与其不变界相邻的节点发送一个离开请求消息,相邻的节点接收到消息之后,将离开节点的 PNS 加入自己的 PNS 中,系统中剩余节点的子空间仍然满足关系式(2)和式(3).

图 3(a)是系统所处的某一状态,图 3(b)是节点 G 加入后的系统状态,图 3(c)是节点 D 离开后的系统状态.采用这种节点的命名、加入、离开策略可以带来一些好处:(1) 节点不存在固定的逻辑标识,采用加入后确定逻辑标识 PNS 的方式可以避免先确定逻辑标识再加入方式所产生的命名冲突问题;(2) 节点的 PNS 采用非确定长度的 0,1 字符串标识可以保证系统命名的可扩放性;(3) 当节点加入所采用的随机函数所产生的随机数分布足够均匀时,节点可以根据它的 PNS 大小来估计系统中节点的总数目,其 PNS 大小与系统中总的节点数目成线性反比.



(a) Before node G joins (b) After node G joins (c) After node D departs

(a) 节点 G 加入前 (b) 节点 G 加入后 (c) 节点 D 离开后

Fig.3 Node joining and departing

图 3 节点的加入及离开

2.4 节点之间的邻接关系

由于在系统中存在着大量的节点,因开销和可扩放性的需要任何节点都不可能与所有节点保持邻接关系而只能与一部分节点保持邻接关系,并通过采用这种邻接关系所形成的逻辑互连网络完成节点之间的通信任务.在 NetShot 中节点一定和与其 PNS 所邻接的两个节点保持邻接关系,与其他节点之间的邻接关系是通过引入路由表 IRT(inbound routing table)、引出路由表 ORT(outbound routing table)来维护的.在一个具有 N 个节点的 NetShot 系统中,一个节点所维护的引出路由表大小是 $O(\log N)$量级的,由于没有节点记录着系统中节点的总数目,对于某个节点 Q 其引出路由表的大小 $ORTsize_Q$ 是通过其节点的 PNS_Q 大小进行估算的:

$$ORTsize_Q=\left\lfloor \log((VCB_Q-UCB_Q)^{-1})\right\rfloor. \quad (4)$$

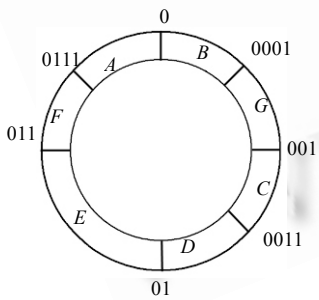


Fig.4 Node naming example

图 4 节点命名实例

一个引出路由表项(ORTEntry)对应于一个与之邻接的节点,对于一个 UCB 为 ucb 的节点,其引出路由表项 C 所对应的节点 $ORTEntry_C$ 是

$$ORTEntry_C=Owner((ucb+2^{-c})\bmod 1.0). \quad (5)$$

节点的引入路由表记录了哪些节点的引出路由表指向了该节点,一个引入表项(IRTEntry)记录着引出路由表项指向该节点的源节点以及其所指向的目标地址.图 4 是 Barnet 中节点的命名示例,表 1,表 2 则是不变界 UCB 为 0001,变界 VCB 为 001 的节点 G 所对应的引出路由表和引入表结构.

采取大小可以动态变化的路由表与采取固定大小的路由表相比可以带来一些好处:若采取固定大小的路由表,如果开始的路由表选择得比较小,当系统中的节点数目增多时则不能够实现系统的可扩放性,如果开始的路由表比较大,则会有很多路由表项是重复的或者是空的.

2.5 消息前递模型

根据第 2.4 节所述的节点之间的邻接关系,Barnet 中的节点与系统中的一部分节点保持邻接关系,节点之间

的通信是依靠这种邻接关系采用消息前递的方式进行的.当节点需要向某一目标节点发送消息的时候,节点首先检查其路由表以及引入表中是否有目标节点,如果有就直接发送给目标节点,否则,在路由表项和引入表项中选择与目标数值最接近的路由表项所对应的节点,并将消息发送给该节点,该节点接收到消息后同样根据上述规则继续进行前递,直到消息传递到目标节点为止.

Table 1 The outbound routing table of node *G*

表 1 节点 *G* 的引出路由表结构

	Source node	Destination address	Destination node
1	*G*=[0001,001)	0001+01=0101	*E*=[01,011)
2	*G*=[0001,001)	0001+001=0011	*D*=[0011,01)
3	*G*=[0001,001)	0001+0001=001	*C*=[001,0011)

Table 2 The inbound routing table of node *G*

表 2 节点 *G* 的引入路由表结构

	Source node	Destination address	Destination address
1	*B*=[0,0001)	0+0001=0001	*G*=[0001,001)
2	*A*=[0111,0)	0111+001=0001	*G*=[0001,001)

3 基于特征信息的定位技术

信息资源定位是指确定信息资源所在的位置从而获取该信息资源内容的操作.Barnet 中的信息资源定位问题分为两个层次:一个是确定信息资源在系统中的逻辑存储位置,一个是根据前面所描述节点之间的逻辑网络拓扑结构和路由模型查找到信息资源的物理存储位置.特征信息(attribute)就是信息资源所具有的一组属性信息,我们可以使用某一特征信息提取方法 *G* 来获得数据资源 *O* 的某一特征信息 *g*:

$$g=G(O). \tag{6}$$

切词技术是在信息检索中常用的自动特征信息提取方法,切词技术可以根据字典将输入信息的特征信息自动提取出来.如果一个信息资源 *O* 的某一特征信息 *g* 采用某一种 Hash 算法(MD5,SHA-1)进行运算后得到 *L* 位的 0,1 的二进制字符串 *h*,则特征信息 *g* 在 Barnet 系统中所对应的特征值 g_v 为

$$g_v=\sum_{i=1}^{L} h_i \times 2^{-i}. \tag{7}$$

在目前的应用系统中,用户可以使用多种方法来定位信息资源,例如:我们在浏览 Web 网页的时候可以根据该网页的特征信息 URL[2]来定位该网页,在使用搜索引擎进行检索时可以根据所需信息资源的一些特征信息如关键字(keyword)来定位信息资源.Barnet 系统希望能够对这些信息资源的定位方式进行统一,实现定位策略与具体应用的独立,用户不必关心信息资源物理存储位置等信息.在 Barnet 中,我们将一个信息资源的位置信息、信息资源内容、关键字等统一定义为该信息资源的特征信息.如式(8)所示,当用户提交给系统一组特征信息 g_s 后,Barnet 系统能够根据这组特征信息定位到与这组特征信息相关的信息资源 R_s,我们称之为基于特征信息的定位技术.对于非确定性定位请求,由于采取不同的特征信息提取方法所提取的特征信息不同,所以同一信息资源的不同特征信息对同一信息资源描述的准确程度也不相同,也就决定着信息资源定位的准确程度.

$$R_s=\text{Location}(g_s). \tag{8}$$

4 数据资源的存储组织及访问模式

在 Barnet 系统中,数据资源的存储组织模式分为两个层次:一个是用户视图中的最小逻辑单元的物理存储模式、冗余控制,另一个是最小逻辑单元在网络中的组织方式.前者主要是指最小逻辑单元的物理存储模式等,后者则包括提供面向存储、代理应用中的用户目录、文件的管理和访问,以及提供面向共享、索引应用中信息资源索引的建立、访问和管理.

4.1 最小逻辑单元的物理存储模式

Barnet 系统中内部节点的信息资源的物理组成单元是以文件的形式存储在 Barnet 系统中的,Barnet 系统中用户检索、访问的信息资源逻辑最小组成单位定义为 Item,一个 Item 对应于用户的一个文件,目录递归地定义为一组文件或目录的集合,Barnet 信息资源存储组织层次模型如图 5 所示.

系统采用 XML 表示的附属描述文件将 Item 映射到物理存储单元上,在附属描述文件中描述了 Item 的基本信息,如作者、大小、创建、修改时间,同时还描述了 Item 的物理存储模式.所有信息资源的访问均需通过附

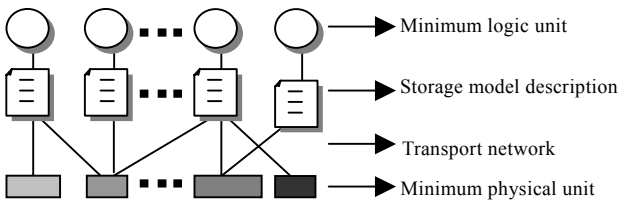


Fig.5 Hierarchy view of information storage and organization

图 5 信息资源存储组织的层次结构视图

属描述文件来映射到网络中的物理存储单元,为了提高文件的可用性及访问效率,我们对 Item 采用 RAID[3]、Reed-Solmen[4]等算法进行软件冗余,将信息划分为若干信息片段,获得若干信息片段中的一些就包含了足够的信息以重构整个信息资源,将信息片段分散在系统中的不同节点来提高可用性、可靠性,同时也协调了系统的负载平衡.为了充分利用网络 I/O 带宽,对于 Item 采用了并行+流水(Parallel+Pipeline)的方式进行访问[5]. 当外节点所访问的信息资源分散在多个内节点时,信息资源的重构操作根据具体应用的不同既可以由胖节点代理完成也可以由外节点完成.

4.2 最小逻辑单元的冗余存储控制策略

Barnet 系统可以为用户提供存储、索引、代理以及共享服务,如果能够识别相同的信息资源,并对信息资源的冗余情况进行控制,则可以提高系统中信息资源的可靠性、可用性及访问效率,同时也可以节省系统的存储空间.在式(9)中定义了一个信息资源,内容为 Content 的信息资源的 *CID*,其中 H_c 为系统中某一指定的 Hash 算法.

$$CID=H_c(\text{Content}). \tag{9}$$

我们为系统中的每一个信息资源建立信息资源内容的 *CID* 索引,Barnet 中的内部节点为系统提供 *CID* 索引的存储空间,节点的 *CID* 索引空间重合于节点的私有命名空间,对于一个节点私有命名空间为 PNS_i 的节点,其索引空间 PCS_i 为

$$PCS_i=PNS_i. \tag{10}$$

对于一个 *CID* 为 *cid* 的内部或外部信息资源,其索引的建立节点为 *cid* 所对应的特征值 *cidv* 所在的节点.通过建立 *CID* 的索引,具有相同信息资源内容的信息资源就同时将索引建立在同一个节点上,系统再对具有相同 *CID* 的信息资源内容进行比较,这样系统就可以快速地识别出所有相同内容的信息资源.系统可以根据具体的应用来控制系统中信息资源的冗余程度以提高信息资源的可靠性以及数据资源的访问效率.

4.3 存储应用中的数据组织

Barnet 原型系统的构建目标包括提供代理、存储等服务,在这些服务中都需要把信息资源存储在 Barnet 中.Barnet 中的内部节点为系统提供存储空间,节点的数据存储空间重合于节点的私有命名空间,对于一个节点私有命名空间为 PNS_i 的节点,其存储空间 PSS_i 为

$$PSS_i=PNS_i. \tag{11}$$

一个信息资源的逻辑位置 *Logic_Location* 可以描述为该信息资源的各个属性的连接:

$$Logic\_Location=Username–Password–Directory–Filename. \tag{12}$$

如果一个信息资源的逻辑位置 *Logic_Location* 为 *logic_location*,H_c 为系统中某一指定的 Hash 算法,则这个数据资源在系统中所对应的物理存储位置 *Storage_Location* 为:

$$Storage\_Location=H_c(logic\_location). \tag{13}$$

这样,在用户访问信息资源的时候只需给出逻辑位置 *logic_location* 的描述,就可以得到信息资源的物理存储位置,从而获得信息资源的内容.

4.4 共享、索引应用中的索引组织

由于 Barnet 原型系统的构建目标包括提供索引、共享等服务,在这些服务中,信息资源存储在这些信息资源的宿主节点上,信息资源的宿主节点可以是内部节点也可以是外部节点,Barnet 中的内部节点为系统提供 *CID* 索引的存储空间,这些宿主节点需要在 Barnet 系统中的内节点上为这些信息资源建立索引后,其他节点才

可以对这些节点的数据进行检索、访问.对于一个节点私有命名空间为 PNS_i 的节点,其索引空间 PIS_i 为

$$PIS_i=PNS_i. \tag{14}$$

下面是索引组织的具体策略:

(1) 使用自动或者人工的方法获得系统中信息资源的最小逻辑单元的用户检索请求的特征描述集 D,一个特征信息对应于一个⟨属性分类 k,属性 v⟩对.如果在特征提取的过程中不能够确定一个特征提取方法所获得的属性内容的属性分类,则将该属性的属性分类定义为 Keyword.

$$D=\{\langle k_1,v_1\rangle,\langle k_2,v_2\rangle,\langle k_3,v_3\rangle,\ldots,\langle k_n,v_n\rangle\}.$$

(2) 将切词特征描述集 AS 中的每一个特征描述词$\langle k_i,v_i\rangle$的连接 k_i–v_i 采用系统中某一指定的 Hash 算法 H_c 提取检索散列特征信息 g_i,并最后获得特征值 gv_i:

$$g_i=H_c(k_i–v_i).$$

(3) 一个信息资源所对应的索引项 E 定义为

$$E=\langle D,CID,\text{定位信息}\rangle.$$

(4) 将索引项信息注册到每一个特征值 gv_i 所在的节点.

4.5 共享、索引应用中的检索定位过程

经过上述的信息资源的索引建立操作,用户就可以检索定位所需的信息资源了,系统中的检索定位过程为:

(1) 将检索特征信息集 $F=\{\langle k_1,v_1\rangle,\langle k_2,v_2\rangle,\langle k_3,v_3\rangle,\ldots,\langle k_m,v_m\rangle\}$提交给系统.

(2) 系统将切词检索描述集 F 中的每一个特征描述词 k_j–v_j 采用系统中指定的 Hash 算法 H_c 提取检索散列特征信 g_j,并最后获得特征值 v_j:

$$g_j=H_c(k_j–v_j).$$

(3) 系统根据检索特征信息集 F 的每一个检索特征信息项$\langle k_j,v_j\rangle$所对应的特征值 v_j,将检索特征信息集 F 定位到相应的索引空间所在的节点,该节点根据检索特征信息集 F 检索出相应的信息资源.

(4) 使用信息资源内容索引项中的 CID 来确定相同的信息资源.

(5) 当若干个节点具有相同的信息资源时,可以采用 Parallel+Pipeline 方式来传输相应的信息资源.

(6) 当节点获得信息资源后,使用系统中指定的 Hash 算法 H_c,对信息资源内容的正确性进行校验.

5 相关的研究工作

计算资源、信息资源的服务是伴随着网络的发展而不断发展的.WWW,FTP 是目前 Internet 上主要的信息服务方式.天网[6]、Google[7]提供的是基于 Internet 的信息资源搜索引擎服务,GRID[8]计算技术是组织和利用网络中分布的计算单元进行协同计算的计算资源服务模型之一.Napster[9],Gnutella[10]是在广域网络中共享存储在不同的计算机上的 MP3 等类型文件的 P2P 应用系统.OceanStore/Tapestry[11,12]是提供面向整个互联网的持久化存储服务,该项目中的路由算法采用的是 Tapestry.CFS/Chord[13,14]也是提供面向整个互联网的持久化存储服务,该项目中的路由算法采用的是 Chord.PAST/Pastry[15,16]也是提供面向整个互联网的持久化存储服务,该项目中的路由算法采用的是 Pastry.Grass/CAN[17,18]是一个在广域网络中对文件内容进行分布的信息服务系统,该项目中的路由算法采用的是 CAN.Netshot 路由模型结合了 CAN 动态命名以及 Chord,Pastry,Tapestry 节点间消息传递逻辑路径比较短的优点,Barnet 系统则结合了多种应用之间的信息共享的优点来提供综合的信息服务.Farsite[19]讨论了在没有中心服务器的情况下将信息资源存储在一组 PC 上的可行性,并在 SIS[20]中描述了如何保证冗余文件存储的惟一性.Publius[21],Freenet[22]是提供匿名性支持的信息发布和存储服务系统,这些系统保证了用户数据访问的隐私性.在文献[23~25]中对在 P2P 网络中如何采用关键字特征信息进行高效的信息资源检索进行了讨论.

6 工作进展以及总结

本文对 Barnet 的设计目标、系统结构、数据资源组织模型、信息定位策略以及我们根据该结构所实现的

原型系统分别进行了介绍,并比较了相关的研究工作.我们已经完成了 Netshot 算法的理论分析以及模拟工作,目前正在构建 Barnet 的原型系统.

References:

[1] Wang QB, Dai YF, Li XM. NetShot: An infrastructure for scalable wide-area location and routing. Technical Report, 2002. http://net.cs.pku.edu. cn/~wangqb/.

[2] Lee TB, Masinter L, McCahill M. RFC1738: Uniform resource locators (URL)/ 1994. http://www.faqs.org/ rfcs/rfc1738.html.

[3] Patterson DA, Gibson G, Katz RH. A case for redundant arrays of inexpensive disks (RAID). In: Proceedings of the 1988 ACM SIGMOD Conference on Management of Data. 1988.

[4] Rabin MO. Efficient dispersal of information for security, load balancing, and fault tolerance. Journal of the Association for Computing Machinery, 1989,36(2):335~348.

[5] Rodriguez P, Kirpal A, Biersack EW. Parallel-Access for mirror sites in the internet. In: Proceedings of the IEEE Infocom 2000, Vol.2. 2000. 864~873.

[6] TianWang. http://e.pku.edu.cn/.

[7] Google. http://www.google.com/.

[8] Global Grid Forum. http://www.gridforum.org/.

[9] Napster. http://www.Napster.com/.

[10] Gnutella. http://www.gnutella.com/.

[11] Kubiatowicz J, Bindel D, Eaton P, Chen Y, Geels D, Gummadi R, Rhea S, Weimer W, Wells C, Weatherspoon H, Zhao B. OceanStore: An architecture for globalscale persistent storage. ACM SIGPLAN Notices, 2000,35(11):190~201.

[12] Zhao BY, Kubiatowicz J, Joseph AD. Tapestry: An infrastructure for fault-tolerant wide-area location and routing. Technical Report, UCB//CSD- 01-1141, Berkeley Computer Science Division, University of California, 2001.

[13] Dabek F, Kaashoek MF, Karger D, Morris R, Stoica I. Wide-Area cooperative storage with CFS. In: Proceedings of the 18th ACM Symposium on Operating Systems Principles (SOSP 2001). Banff: Chateau Lake Louise, 2001.

[14] Stoica I, Morris R, Karger D, Kaashoek MF, Balakrishnan H. Chord: A scalable peer-to-peer lookup service for Internet applications. In: Proceedings of the ACM SIGCOMM 2001 Conference. 2001.

[15] Druschel P, Rowstron A. PAST: A large-scale persistent peer-to-peer storage utility. In: Proceedings of the 8th IEEE Workshop on Hot Topics in Operating Systems VIII. 2001.

[16] Rowstron A, Druschel P. Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems. In: Proceedings of the IFIP/ACM International Conference on Distributed Systems Platforms (Middleware). 2001. 329~350.

[17] Ratnasamy S, Francis P, Handley M, Karp R, Padhye J, Shenker S. Grass-Roots content distribution: RAID meets the Web. 2001. http://www.aciri.org/sylvia/.

[18] Ratnasamy S, Francis P, Handley M, Karp R, Shenker S. A scalable content-addressable network. In: Proceedings of the ACM SIGCOMM Symposium on Communication, Architecture, and Protocols. ACM SIGCOMM, 2001. 161~172.

[19] Bolosky WJ, Douceur JR, Ely D, Theimer M. Feasibility of a serverless distributed file system deployed on an existing set of desktop PCs. In: Measurement and Modeling of Computer Systems. 2000. 34~43.

[20] Bolosky WJ, Corbin S, Goebel D, Douceur J R. Single instance storage in Windows 2000. http://research.microsoft.com/farsite/WSS2000.pdf.

[21] Waldman M, Rubin AD, Cranor LF. Publius: A robust, tamper-evident, censorship- resistant, web publishing system. In: Proceedings of the 9th USENIX Security Symposium. 2000.

[22] Clarke I, Sandberg O, Wiley B, Hong T. Freenet: A distributed anonymous information storage and retrieval system. In: ICSI Workshop on Design Issues in Anonymity and Unobservability. 2000.

[23] Yang B, Garcia-Molina H. Efficient search in peer-to-peer networks. In: Proceedings of the International Conference on Distributed Computing Systems (ICDCS). 2002.

[24] Reynolds P, Vahdat A. Efficient peer-to-peer keyword searching. Technical Report, CS Department, Duke University, 2002. http://issg.cs.duke.edu/search/.

[25] Prinkey MT. An efficient scheme for query processing on peer-to-peer networks. http://aeolusres.homestead.com/files/index.html.