第 33 卷 第 4 期 计 算 机 工 程 2007 年 2 月
Vol.33 No.4 Computer Engineering February 2007
·网络与通信· 文章编号：1000—3428(2007)04—0088—03 文献标识码：A 中图分类号：TP391

面向分布式协同设计的对等信息共享系统

王 征，刘心松，李美安
(电子科技大学计算机科学与工程学院 8010 研究室，成都 610054)

摘 要：为了实现分布式协同设计中的共享信息快速检索以及多副本同步，提出了基于对等网结构的信息共享系统，给出了该信息共享系统的节点模型、管理策略及信息检索模型，提出了一种结合分布式哈希表和聚类的检索方法，保证了用户能够在协同设计系统中快速地精确检索和“盲目”检索，实现了系统的用户透明。为了保证分布式多副本同步，提出“对等锁”作为一致性维护方法。该文给出了系统的具体实现方法，并给出了实例。

关键词：分布式协同设计；对等；信息共享系统；检索；分布式哈希表；多副本同步

P2P Information Sharing System in Distributed Collaborative Design

WANG Zheng, LIU Xinsong, LI Meian
(8010 R&D, Computer Science and Engineering College, University of Electronic Science and Technology, Chengdu 610054)

【Abstract】 Aiming at the information retrieval and the multi-replica synchronization in distributed and collaborative design systems, an information sharing system based on P2P (peer to peer) structure is presented. The node model, the management scheme and the information retrieval model are given. In order to ensure the exact retrieval and the blind retrieval, an information search method based on distributed hash table and information clustering is given. In order to ensure the distributed multi-replica synchronization, a novel P2P multi-replica synchronization algorithm “peer-to-peer lock” is proposed. The implement method and an instance system are given.

【Key words】 Distributed and collaborative design; Peer to peer (P2P); Information sharing system; Retrieval; Distributed hash table (DHT); Multi-replica synchronization

在计算机支持的分布式协同设计(简称分布式协同设计)方式下，分布在不同地点的产品设计人员通过网络采用计算机辅助工具协同地进行产品设计活动[1]。目前的分布式协同设计系统，如Liao等的ASP模式的协同设计系统[2]等，处理共享信息的主要方法是将其设置为“共享”，用户手动在服务器间切换，并在所登录的服务器上查找/访问。这种方法效率低下、缺乏统一的接口，而且无法保证分布式环境下的多副本同步[3]。

目前，基于对等网P2P(Peer to Peer)的系统被广泛的研究应用[4,5]，例如Chord、PASTA等。P2P系统用对等互助模式替代Client/Server(C/S)模式，系统中的节点存储信息也充当信息检索中的资源路由。本文针对协同分布式设计系统的特点，提出了面向分布式协同设计的对等信息共享系统，解决传统信息共享系统的诸多问题。

1 系统模型与信息共享流程

1.1 对等信息共享系统模型

面向分布式协同设计的对等信息共享系统可以定义为：(1)系统由具有服务器质量的计算机组成，向协同设计用户提供统一透明的信息共享服务；(2)具有独立于 DNS 的信息寻址系统，通过资源路由高效快速地检索/获取共享信息；(3)具有可变连接合作的能力。

面向分布式协同设计的对等信息共享系统的节点模型各主要模块有：

(1)资源浏览器(Resource Browser)/客户接口(Client Service Interface)/服务接口(Server Service Interface)：提供接口/工具给协同设计系统中的用户/子系统。其中，登录 P2P 节点的用户通过资源浏览器检索/获取系统中的可用信息资源；CSI 提供本地/远程调用接口给客户端程序；管理员通过 SSI 配置检索/资源路由器的初始化路由等信息。

(2)检索器/资源路由是节点模型的核心部分。它主要处理经过哈希化的信息(节点号、信息的 Key 标识等)；同时，它负责对收集到的信息进行分类，协调其它模块的工作；本地共享信息通过分布式哈希表(Distributed Hash Table, DHT)获得对应的关键字 Key，由资源路由将其 Key 标识扩散到同一关键字空间中的相关节点；异地共享信息的 Key 通过资源路由分类存储在本节点中。

(3)路由表/叶子集/邻居集：每个节点在进入系统时都通过 DHT 分配了一个 128 位的节点号；本地节点获得的异地节点号/IP 地址存入 3 个集合中：通常，叶子集保存了节点号最接近(通常是同一广播域内的节点)的信息；路由表/邻居集保存了其它节点信息；本节点模型根据分布式协同设计中信息聚类分布/访问的特点，对上述 3 个集合作了按用户访问兴趣/资源特征聚类的改进。

(4)对等节点管理/对等通信：前者控制本地节点，主动将节点号/IP 地址扩散至同广播域的节点，同时也通过广播获得其它节点的信息；对于不在同一个广播域的节点，本地节点需要系统管理员配置或从其它节点获得；同时，该模块还负

基金项目：四川省应用基础研究项目(04JY029-017-2)；科技型中小企业技术创新基金资助项目(04C26225110223)
作者简介：王 征(1979－)，男，博士生，主研方向：分布式系统；刘心松，教授、博导；李美安，博士生
收稿日期：2006-03-31 **E-mail：**wangzheng151400@163.com

责分布式多副本一致性的维护工作。后者发送消息、建立信息传输通道，完成信息检索、资源路由、文件共享等一系列工作。信息共享系统在分布式协同设计系统中的位置及节点模型如图 1 所示。

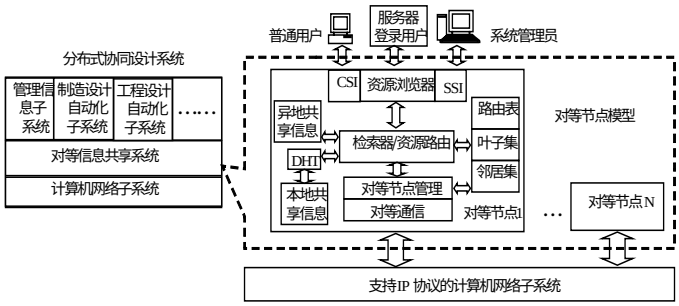

图 1 系统与节点模型

1.2 信息共享系统运行流程

面向分布式协同设计的对等信息共享系统的运行是一个 3 段流程，包括：(1)信息的搜集/发布，主要将共享信息的元数据、用户兴趣、聚类特征等检索信息在系统中收集/规范化/扩散；(2)预处理，主要是关键词提取、索引化/聚类操作等；(3)服务，包括检索、获取信息、多副本同步等。下文以一个共享设计图的处理流程说明该系统的运行流程。

(1)在单个节点上生成设计图，将该图设置为“对等共享”或者保存在对等共享的文件夹中；本地节点的信息共享系统自动提取该图的元数据(文件名、类型等属性)；由 DHT 运算生成该图的 Key 值，封装<Key,URL>二元组，其中 URL 是本地节点号；检索器/资源路由在将其保存入本地共享信息的同时，也将其扩散至叶子集中的各个节点，以及节点号和该 Key 接近的节点中。

(2)节点接收到其它节点扩散的<Key,URL>二元组，将其按照类型保存在异地共享信息中。

(3)系统中某节点发起协同请求，遍历设计表单，发现需要该设计图时，同样通过该文件的元数据信息(文件名等)由 DHT 生成 Key 值。首先查找异地共享信息，如果有 Key 值相等的表项，则可以直接从 URL 表征的节点中获取该图；如果没有，进一步向节点号最接近 Key 值的节点发送检索消息<Key,URL>。

(4)当某节点接收到检索消息后，检查本/异地共享信息中是否存在接收到的 Key，如果有返回查询结果给检索节点；否则继续前递(Forward)消息给叶子集/路由表中的关键字空间距离最近的节点。为了限制搜索深度，消息中带有 TTL(Time to Live)标识：每次路由该消息时，将 TTL 减 1；当 TTL=0 时，清除该消息。

(5)当检索设计图的节点接收到检索应答消息时，一方面更新异地共享信息，插入<Key,URL>项，必要时也修改路由表/邻居集；另一方面该节点的对等通信模块自动从设计图所在节点获取该图的副本。

(6)当该图的多个副本扩散在系统中，多个用户对该图进行写操作(插入、删除等)，则副本所在节点通过节点管理模块进行分布式互斥操作，保证多副本的一致性。

2 关键技术与系统实现

2.1 基于 DHT 和聚类的检索

本系统采用了基于 DHT 的资源路由基；通过引入聚类方法加快搜索收敛速度，并能进行精确搜索及基于聚类的“盲目”检索。

路由基(Routing Substrate)为对等信息共享系统提供了依据关键字的定位服务；目前最为流行的对等路由基是分布式哈希表(Distributed Hash Table，DHT)技术；这种技术给每个节点分配一个全局唯一的伪随机数作为节点号，该随机数用于指示该节点在一个关键字空间中的位置；检索消息被路由到具有相同关键字的位置上，或者被转发到关键字空间中距离最近的节点上，如此循环直到找到所需信息或TTL失效为止。本系统采用SHA-1(Secure Hash)函数生成一个 2^{128}容量的环形关键字空间，节点号的分配、资源Key生成都在该空间内进行；节点将自身共享信息的元数据哈希化，生成对应的关键字Key，扩散在系统中；节点/关键字的管理查找方式类似于Pastry模型。

但普通的 DHT 路由基只能“先知”的精确检索，即检索者需要提供检索对象的精确元数据，例如文件名等。用户通常需要根据模糊条件进行“盲目”检索、搜索一类信息，而后从中选择合适的对象；这种检索无法用传统的 SHA 算法实现。为解决这个问题，系统在生成共享信息/路由表/叶子集时，对它们进行本地聚类；具体做法是：资源路由将转发的“应答”和其他节点扩散的 Key 转化为异地共享信息，保留在本地作为检索依据；信息表行内按照关键字的相似度存储，表列按照系统指定的类别进行索引化；在表列上形成了“倒排表”(Inverted File)索引结构。这种做法相当于将二维环形关键字空间进行了聚类多维化处理，将其转化为聚类关键字空间：一个关键字可以被分配在多个类空间中，当系统提出一类信息的检索时，可以从表列索引进入，从而获取相关信息。图 2 显示了关键字空间模型与精确/“盲目”检索的运行过程。

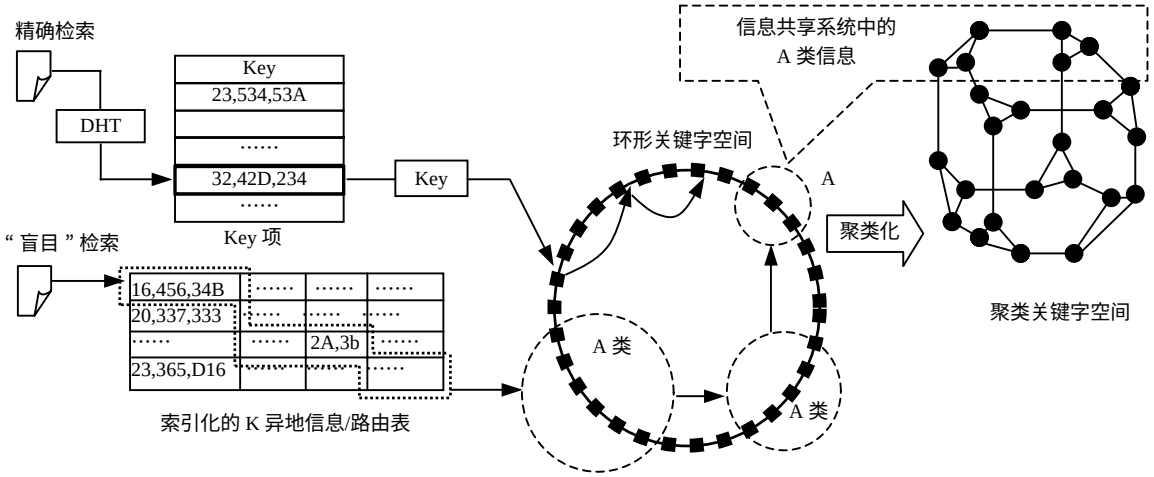

图 2 基于对等信息共享系统的信息检索

2.2 P2P 下的分布式多副本同步

异地协同设计等应用环境需要进行分布式多副本同步；这也是分布式协同设计系统的一个难点。通过在信息共享系统中引入 P2P，使该问题变得容易解决。

2.2.1 模型与算法

分布式多副本同步算法的实质是分布式互斥，用以保证分布式系统的一致性，即确保每次只有一个节点进入临界区，其关键在于如何选择决策节点。传统分布式互斥的算法很多，例如 Lamport 的时间戳算法、Maekawa 的请求集算法。本系统中将传统的“ 锁 ”机制扩展为“ 对等锁 ”，解决一致性问题，“ 对等锁 ” 算法描述如下：

Step1 需要进入临界区的节点P_m通过DHT路由基发送“ 请求 ” 消息给主本所在节点P_0。注意：系统中可能已经存在多个副本，部分副本可能由其他副本复制生成，这些副本不能视为主本；因此本算法中节点将“ 请求 ” 消息发送给主本所在节点。

Step2 P_0接收到多个“ 请求 ”，P_0对这些消息进行排序，存入请求队列；主P_0给优先级最高的节点P_m发送“ 应答 ” 消息，允许它进入临界区，P_0即被点P_m锁住；接到“ 释放 ” 消息前，P_0不得发送“ 应答 ” 给其他节点。

Step3 优先级最高的节点P_m退出临界区后，不需再次查找，直接发送“ 释放 ” 消息给P_0。

Step4 主本所在节点 P0 接到“ 释放 ” 消息时，释放锁，删除消息队列中 Pn 产生锁的记录，并发送“ 应答 ”给下一个节点；重复上述 Step2，Step3，直到请求队列为空。

2.2.2 消息复杂度

传统副本同步算法需要在系统中广播/组播大量的消息；即使在节点全互连的网络中，Lamport算法也具有O(N)的消息复杂度，Maekawa算法具有$O(\log_2 N)$消息复杂度。在多跳网络中，这些算法的消息复杂度还将上升；而在本系统中，为获取一次进入临界区的机会，节点最多将传递$\log_2{}^b N$个消息用于查找主本所在节点；返回“ 应答 ”和“ 释放 ”，最多各需要$\log_2{}^b N$个消息(全互联网络仅需 1 个消息)，因此其消息复杂度为$O(\log_2{}^b N)$。因为b大于 1，所以“ 对等锁 ” 算法的消息复杂度远低于其他算法。

3 系统实现与实例

如图 3 所示，基于对等信息共享的分布式协同设计系统在某企业的虚拟专用网络(Virtual LAN，VLAN)中得以实现。该系统由 4 个局域网组成，局域网间通过 Internet 互联，单个局域网由 6~20 台服务器组成，节点间通过 IP 进行直接通信；服务器安装有 Windows 2000 Server 及 Red Hat Linux 9 等操作系统。为适应异构环境，对等信息共享系统提供了若干操作系统的版本，但实现机理是一致的。该系统在实际应用中得到了应证。图 3 通过消息跟踪的方式显示了局域网 1 中的节点①的获取局域网 3 中的共享文件“ 区位设计图 0905.DWG ” 的全过程。

Step1 通过资源浏览器检索，登录节点①的用户根据“ 区位设计表单 ” 确定所需信息；请求从系统中获得“ 区位设计图 0905.DWG ”；此时节点①并不知道该文件的位置；因此节点①的消息处理模块通过 DHT 获得该文件的 Key，并将检索消息<Key, URL① >扩散至关键字空间距离最近的节点③。

Step2 节点③的消息接收/发送线程通过消息缓冲将<Key, URL① >转交给消息处理模块；消息处理线程在本地检索后，未发现<Key, URL①>的相关信息；因此节点③将该消息转发给关键字空间距离更近的节点④。由于整个系统采用 VLAN 技术，分属不同局域网的节点，因此可通过 IP 通道直接发送消息。

Step3 节点④继续转发消息给节点⑥，节点⑥有“ 区位设计图 0905.DWG ” 文件的主本，因此其本地共享信息中有该文件的 Key 与<Key, URL①>一致；节点⑥的消息处理线程启动共享对象处理线程，主动将该文件的副本扩散至节点①。特别注意：该实例中，节点④和节点⑥虽然在一个局域网中，但是不在同一广播域，因此节点④中并没有“ 区位设计图 0905.DWG ” 的信息。如果节点④中有该 Key，则不用转发消息给节点⑥；节点④将直接发应答给节点①，由节点①中的共享对象处理线程发起扩散操作，从节点⑥下载副本。

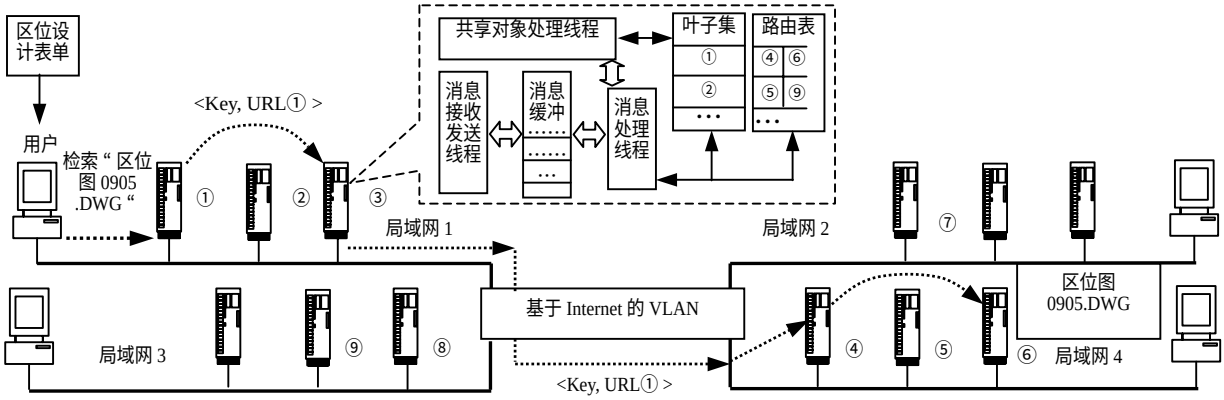

图 3 面向分布式协同设计的对等信息共享系统实例

4 结束语

综上所述，面向分布式协同设计的对等信息共享系统能够满足“ 信息共享，分散存储 ” 的需求；该系统通过 DHT 及聚类向分布式协同设计人员提供了更快捷地自动信息检索；实现了真正的用户透明；同时系统通过“ 对等锁 ” 算法保证了分布式环境下多副本信息的一致性。

参考文献

1 高曙明, 何发智. 分布式协同设计技术综述[J]. 计算机辅助设计与图形学学报, 2004, 16(2): 149-157.

2 廖 敏, 殷国富, 罗中先. ASP 模式的机械产品分布式协同设计的研究[J]. 计算机辅助设计与图形学学报, 2004, 17(6): 1341-1346.

3 郭学旭, 王云鹏, 潘 翔. 计算机辅助协同设计系统并发控制机制的研究[J]. 计算机辅助设计与图形学学报, 2004, 16(2): 201-205.

4 鄢 萍, 王逢春, 刘 飞. 一种基于对等网技术的企业信息分布式集成方法[J]. 计算机集成制造系统——CIMS, 2005, 10(5): 492-496.

5 吕建明, 刘 悦, 丁 林. P2P 与信息检索[J]. 信息技术快报, 2005, 3(2): 1-12.