

基于网络的半结构化网络数据搜索模型^{*}

陈庆奎¹ 那丽春²

(上海理工大学计算机工程学院 上海 200093)¹ (上海立信会计学院信息科学系 上海 201620)²

摘要 为了提高大规模网络数据搜索的效率,提出了一个基于网格机制的 Internet 网络数据搜索模型。给出了数据搜索节点、数据搜索服务器、数据搜索区域、数据服务器区域和数据搜索网格(Data Searching Grid-DSG)等一系列定义。利用网络跳距离、数据搜索节点密度、计算节点能力和可信性机制以及市场服务代价原则,给出了 DSG 结构和构建算法。描述了基于 DSG 的网络数据搜索过程。该模型的控制结构是基于网格结构的,其具体搜索机制是非结构化的。理论分析表明,该模型具有良好的扩展性和较高的效率。

关键词 网格,数据搜索模型,网络数据搜索,国际互联网

A Semi-structure Data Searching Model Based on Grid

CHEN Qing-Kui¹ NA Li-Chun²

(College of Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093)¹

(Department of Information Science, Shanghai Lixin University of Commerce, Shanghai 201620)²

Abstract For improving the efficiency of large scale Internet Data searching, a Grid-Based Management Model was discussed. A series of formal definitions, such as the Data Searching Node (DSN), the Data Searching Server, the Data Searching Area, the Data searching Server Area and the Data Searching Grid (DSG), were given. Using the network-skip-based distance, the density of DSN, the ability and Reliability of Computer and the mechanism of market price, the architecture and construction algorithms of DSG was proposed. The internet data searching process was described in DSG. The control structure of this model is Grid-based, and its searching mechanism is semi-structure. The analysis shows that this model has the better expansibility and the higher efficiency.

Keywords Grid, Data searching model, Internet data searching, Internet

随着 Internet 和信息技术的飞速发展,特别是 Web 技术的广泛应用,面对呈爆炸趋势增长的 Internet 信息海洋,使得人们对网络数据的挖掘技术和搜索技术的研究越来越迫切,并已经出现许多实用的搜索技术。其中基于 P2P 技术的网络数据搜索引擎已经被广泛使用并已成为业界研究热点^[1~5]。由于 P2P 系统是一个分散的、动态的系统,系统中不存在中心控制节点,节点可以动态地加入和退出网络,是一个全分散的网络数据搜索系统,这样随着搜索网络规模的增大,效率将变得低下,并且系统的管理也变得困难。半结构化技术(如 FreeNet^[6])可以克服这些缺点。同时,网格^[7]技术可能成为另一个可行的解决方案。目前由多个计算机网络构成的 Intranet 越来越多,大量的廉价的个人计算装置随处可见,但是它们的资源利用率非常低^[8~10]。利用这些空闲计算机的计算资源、存储资源、通信资源,可以得到大量的、非专用的、廉价的、大规模的、分布广泛的网络数据搜索机。显然,如此规模的数据搜索系统可以带来可观的搜索、挖掘性能,然而其控制管理问题也是一个难题。网格系统的监控机制的研究和应用已经有经可鉴^[11~15];文[12]是全球网格论坛性能工作组提出的 GMA(Grid Monitoring Architecture)结构,是基于目录服务的网格监控框架。文[13]所描述的 PyGMA 提供了 SOAP 和 WWW 服务框架上的 GMA 接口。文[14]所描述的 DMF 是美国国防部国家联合实验室项目,提供数据密集网格

环境下性能分析和检测。文[15]论述 NWS 分布式环境各种资源的监控预测技术。这些理论技术可以用于网络大规模搜索系统的控制与监测。

目前基于网格机制的网络数据搜索系统还不太多,本文给出一个基于网格机制和 OGSA^[16]体系结构的、通过对 internet 区域覆盖,面向网络数据挖掘的、分布式、大规模的半结构化网络数据搜索模型。该模型的控制结构是基于网格结构的,其具体搜索机制是非结构化的。该模型可以解决分布式的、大规模数据搜索机的管理问题,适合互联网上的数据搜索、数据挖掘应用。

1 DSG 体系结构

按照 OGSA 规范,一个网格就是通过若干个网格服务相互作用的虚拟计算环境。而环境的构成部分是动态加入、撤出。网格资源、规模、处理能力、任务调度是自动完成。它们在一定规则集的约束下有机地生存,网格约束机制是根据能力(Ability)、可信度(Reliability)、服务代价(Price)等因素制定的。

定义 1(数据搜索节点) 数据搜索节点(Data Searching Node)是一个六元组 $DSN(D_i, D_a, D_r, D_p, D_{func}, D_{ub})$,其中 D_i 为其标识符; D_a 为其能力,可根据 DSN 的计算能力、存储能力、I/O 能力、通信能力、搜索算法而定; D_r 为其可信度,可

^{*}国家自然科学基金(60573108)、上海自然科学基金(04ZR14100)。陈庆奎 教授,主要研究领域为网格、计算机机群、并行数据库;那丽春 副教授,主要研究领域为网格技术、并行计算、数据挖掘。

以根据该挖掘节点的稳定生存时间、成功完成挖掘任务的比率、网络可达性等运用统计机制计算而得; D_p 为其服务代价, 即其作为网络成员应该获得的回报; D_{func} 为该数据搜索节点的搜索算法集合; D_{lib} 为其局部数据库, 用来缓存数据字典、网络数据搜索的临时结果。

定义 2(数据搜索服务器) 数据搜索服务器(Data Searching Server)是一个九元组 $DSS(S_{id}, S_{level}, S_a, S_r, S_p, S_{father}, S_{sm}, S_{func}, S_{lib})$, 其中 S_{id} 为其标识符; S_{level} 为该服务器的级别; S_a 为其能力, 可根据 DSS 的计算能力、存储能力、I/O 能力、通信能力、搜索算法而定; S_r 为其可信度, 可以根据该挖掘节点的稳定生存时间、成功完成挖掘任务的比率、网络可达性等运用统计机制计算而得; S_p 为其服务代价, 即其作为网络成员应该获得的回报; S_{father} 该服务器父服务器的集合(本模型中该域为网格主控制器); S_{sm} 为该服务器所辖的数据搜索节点的集合; S_{func} 为其搜索功能部件集合; S_{lib} 为其服务器数据库, 用来缓存数据字典、网络数据搜索的临时结果。

定义 3(数据搜索网格 DSG) 一个数据搜索网格为四元组 $DSG(Master, S_{ds}, S_{dn}, S_{rule})$, 其中 Master 为网格主控制器; S_{ds} 为其所辖数据搜索服务器的集合; S_{dn} 为其所辖数据挖掘节点的集合; S_{rule} 为其逻辑拓扑结构规则。

定义 4(网络跳机制距离) 我们把以网络跳数为度量单位的距离称作网络跳机制距离。

定义 5(圆形数据搜索区域) 常数 r 为网络跳机制距离, 在 Internet 上以任意一个计算节点为圆心, 以 r 为半径做一个圆, 所得的圆形面积区域为一个圆形数据搜索区域; 圆形数据搜索区域内的所有被搜索的数据源(如 Web)称为该区域的数据搜索对象; r 是圆形区域搜索半径。

定义 6(数据服务器区域) 一个数据搜索服务器 DSS 所管辖的多个相邻的(网络相邻)圆形数据搜索区域构成的复合区域, 称该服务器的数据服务器区域。其规模由所辖圆形数据搜索区域集合的数据业务综合确定。每个圆形数据搜索区域一定隶属一个 DSS。

为了简化描述, 定义其规模为该 DSS 所辖的圆形数据搜索区域的个数, 并记常数 NS。

定义 7(数据搜索节点密度) DSG 的一个搜索区域 A , s 为 A 中的 DSN 的个数, w 为 A 中的数据搜索对象的数目, 则 $d=s/w$ 为区域 A 的数据搜索节点密度。

定义 8(计算节点的特征向量) 为 internet 上的每个计算节点都设有一个记载其能力、可信度和服务代价的向量 $v=(v_a, v_r, v_p)$, 称该计算节点的特征向量。

定义 9(区域服务代价值) 在一个圆形数据搜索区域内的所有数据搜索节点都会得到回报 Repay, 用一个数值来表示, 称区域服务代价值。

定义 10(计算节点市场) 计算节点市场是一个计算节点集合, 每个计算节点都有自己的特征向量。

定义 11(数据搜索代理 DA) 在互联网的各个区域上, 介于数据搜索用户和 DSG 之间、进行业务关联的网格服务实体被称为数据搜索代理。

DSG 覆盖 Internet 的结构如图 1 所示。图中的小圆为圆形数据搜索区域, 圆内按着数据搜索节点密度 d 分布着许多 DSN 和数据搜索对象; 大虚线圆为同心圆, 是 DSG 区域扩展线; 矩形是被搜索的 Internet; 虚六边形所联接圆形数据区域集合(4, 15, 14)是一个 DSS 数据服务器区域。图 1 的 DSG 由 5 个不相交的 DSS 区域的集合构成。

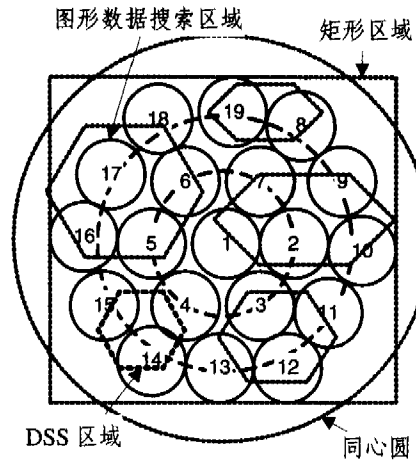


图 1 DSG 示意图

2 网格机制数据搜索模型算法

根据 OGSA 规范, 数据搜索网格的构造应该是由若干个分布在 Internet 各个区域的、有一定计算能力、可信度、有限价格的计算节点, 通过第 1 节所定义的部件自动构建、自动完善、自动搜索。本节分两个部分描述: 2.1 节讲解 DSG 的构建、动态调整机制; 2.2 节介绍 DSG 的网络数据搜索过程。

2.1 DSG 的构建与调整

算法 1(DSG 初始化构建) 设在 Internet 的各个区域有 p 个计算节点 $C_1, C_2, \dots, C_p$, 它们的特征向量分别为 $v_1, v_2, \dots, v_p$; r 为初始数据搜索区域半径; d 为 DSG 的数据搜索节点密度阈值。

算法描述如下:

(1) 根据网络跳为距离单位, 计算 p 个计算节点的所在区域边界, 确定该计算节点集合的矩形区域, 设该区域长为 L , 宽为 W ; 取矩形区域左下角为坐标原点。

(2) 计算矩形区域的中心 (x, y) , 计算如下:

$x=(L+2r)/2$; $y=(W+2r)/2$; /* 在矩形的四周个加入一个宽度为 r 的带条 */

(3) 以 (x, y) 为圆心、以 r 为半径作圆, 形成一个区域为 A_1 ;

(4) 以 (x, y) 为圆心, 以 $2r$ 为半径作一个虚线圆 $Circle_1$; 置 $Circle = Circle_1$;

(5) 然后在 $Circle$ 的圆弧上任取一点, 以 r 为半径作圆, 得到区域 A_2 ; 同样重复在 $Circle$ 的弧线上作圆, 得到一组区域 $A_3, A_4, \dots, A_k$, 使 $A_2, A_3, A_4, \dots, A_k$ 恰好完全覆盖 $Circle$ 的圆弧, 并且各个区域互不相交(如图 1 中的小圆 2, 3, 4, 5, 6, 7 为 6 个圆形数据搜索区域, 覆盖了半径为 $2r$ 的同心圆扩展线);

(6) 同样以 (x, y) 为圆心, 以 $4r, 6r, 8r, \dots$ 为半径做同心圆 $Circle_1, Circle_2, \dots$; 取 $Circle = Circle_2, Circle_3, \dots$, 重复执行(5), 直到某个虚线同心圆覆盖了矩形区域为止;

得到圆形数据搜索区域序列 $A_1, A_2, A_3, \dots, A_n$ (如图 1 中的 19 个小圆为 DSG 的圆形数据搜索区域集合);

(7) 确定计算节点 $C_1, C_2, \dots, C_p$ 所在的圆形数据搜索区域, 并把它们的特征向量的各个分量 v_a, v_r, v_p 分别添加到 D_a, D_r, D_p 域;

计算每个圆形数据搜索区域的数据搜索节点密度;

根据阈值确定各个圆形数据搜索区域的状态, 设区域 A_i

的实际数据搜索节点密度为 d_i , 圆形区域 A_i 的状态计算方法如下:

```

IF  $d_i/d=0$  then
   $A_i$  为无效状态;
Else IF  $0 < d_i/d < 1$  Then
   $A_i$  为薄弱状态;
Else IF  $d_i/d > 1$  Then
   $A_i$  为过强状态;
Else IF  $d_i/d=1$  Then
   $A_i$  为饱满状态;
End If;

```

(8) 对每个圆形数据搜索区域 A_i 构建数据搜索节点, 定义其区域服务代价, 形成 n 个数据搜索区域;

(9) 对无效和薄弱状态区域, 增加服务代价值, 以便吸引更多该区域内的闲散的节点成为该区域的数据搜索节点;

(10) 对饱满状态的区域, 适当减少服务代价值, 迫使一些节点退出数据搜索工作;

(11) 通过网络相邻的多个数据搜索区域组合构成数据服务器区域, 使得每个数据服务器区域的规模不大于 NS, 这样 $\{A_1, A_2, A_3, \dots, A_n\}$ 被划分成多个互补相交的子集合, 每个子集合为 DSG 的一个 DSS 区域 (图 1 的虚六边形涉及的圆形区域为一个 DSS);

在每个数据服务器区域的所有数据搜索节点中, 取可信度最高、能力最大的数据搜索节点 DSN 为该数据服务器区域的数据搜索服务器 DSS ($S_{id}, S_{level}, S_a, S_r, S_p, S_{father}, S_{son}, S_{func}, S_{lib}$), 从 DSG 的 Master 获取 S_{func}, S_{lib} , 并填写所有参数, 并使 DSN 的 D_a, D_r, D_p 赋值给 DSS 的 S_a, S_r, S_p ;

(12) DSG 的 Master 获取所有数据服务器区域的信息, 并存入数据库, DSG 构建完毕。

算法 2 (DSG 动态调整)

1) DSG 的所有 DSS 定期地并行执行如下规则:

(1) 监控其所属的 DSN 的特征状态值 D_a, D_r, D_p ; 如果 DSN 的能力、可信度发生变化, 则动态修正 D_a, D_r ;

(2) 如果某 DSN 的 D_a, D_r 值低于阈值, 则收回其 DSN 资格, 遭送到计算节点市场;

(3) 如果计算节点市场有新的该圆形区域的数据搜索节点 C_i 出现, 并且其所需代价符合区域服务代价, 则

{ 如果该圆形数据搜索区域的状态为饱满状态或过强状态, 则用 C_i 替代其中能力最差、可信度最低的 DSN;

如果该圆形数据搜索区域的状态为薄弱状态或无效状态, 则把 C_i 作为 DSN 加入该区域;

}

(4) 如果某一圆形数据搜索区域为过强状态, 则通过提降低区域服务代价值的方法来迫使一些 DSN 回到计算节点市场。

(5) 如果某一圆形区域为薄弱状态, 则通过提高区域服务代价值的方法来吸引计算节点市场的节点作为新的 DSN。

(6) 如果一个没有隶属数据服务器区域的无效状态的圆形区域, 变成饱满、过强、薄弱的任一状态, 则选其中能力最强的 DSN 为一个新数据服务器区域的 DSS; 并按 NS 为规模限制吸收其网络相邻的无效圆形数据搜索区域为新的数据服务器区域成员。

2) DSG 的所有 DSS 定期地对其所有 DSS 执行同本算法

第 1) 部分所描述的类似的规则, 来调整所有 DSS (略)

算法 3 (DSG 扩建)

设 DSG 的当前最大半径为 $2mr+r$, 并且有 n 个圆形数据搜索区域。区域则如果想扩充到半径为 $(2m+2)r+r=2(m+1)r+r$, 则算法如下:

(1) 以矩形区域中心 (x, y) 为圆心, 以 $2(m+1)r$ 为半径作一虚圆 Circle;

(2) 在 Circle 上, 执行算法 1 的第 (5) 步, 得到 k 个圆形数据搜索区域 $A_{n+1}, A_{n+2}, \dots, A_{n+k}$, 并置这些圆形区域为无效状态;

(3) 向计算节点市场公布新增加的圆形数据搜索区域 $A_{n+1}, A_{n+2}, \dots, A_{n+k}$;

(4) 启动算法 2, 进行 DSG 调整。

2.2 DSG 的执行过程

算法 4 (DSG 驱动过程)

(1) 数据搜索代理 DA, 接收来之用户的查询关键字集合 KS;

(2) DA 把 KS 传给 DSG;

(3) DSG 驱动其所有 DSS 对 KS 的搜索;

(4) 所有 DSS 并发地驱动其所有 DSN 完成在各自数据搜索区域内对 KS 的搜索 (执行算法 5);

(5) 所有 DSS 汇集其 DSN 的搜索结果, 并返回给 DSG 的 Master; Master 汇总搜索结果, 并传给 DA; DA 传给用户; 搜索结束。

算法 5 (数据搜索区域 A_i 的 DSN 搜索过程)

已知数据搜索区域 A_i 有 m 个 DSN: $DSN_1, DSN_2, \dots, DSN_m$; n 个被搜索的数据源计算机 $\{DS_1, DS_2, \dots, DS_n\}$; 每个 DS_i ($1 \leq i \leq n$) 有一个搜索标志量 SM_i , 其初值为 0; 如果本次搜索过程中 DS_i 被任何一个 DSN_j ($1 \leq j \leq m$) 第一次搜索到, 则标志 SM_i 被置为 1, 其它 DSN_k ($k \neq j, 1 \leq k \leq m$) 将不再对 DS_i 进行搜索, 当所有 SM_i 都为 1 时, A_i 的本次搜索结束。该过程是一个半结构化的过程。算法描述如下:

(1) A_i 对所有 DS_i ($1 \leq i \leq n$) 的搜索标志 SM_i 置为 0;

(2) 所有 DSN_j ($1 \leq j \leq m$) 并发地做 (3)、(4);

(3) 置距离变量 Distance=1;

(4) DSN_j 向距离其为 Distance 的并且 SM_i 为 0 的 DS_i 发出搜索请求;

置 SM_i 为 1; /* 第一次搜索到 */

对返回信息进行搜索;

存储搜索结果;

Distance++; /* 距离加 1 */

如果 Distance 未超过 A_i 的边界, 则重复 (4);

如果所有 DS_i ($1 \leq i \leq n$) 的搜索标志 SM_i 都为 1, 则执行 (5);

(5) 所有 DSN_j ($1 \leq j \leq m$) 并发地把搜索结果传给它们的 DSS; 本次搜索结束。

2.3 理论分析

定理 1 DSG 能够完成对 Internet 的数据搜索。

证明: 根据算法 1 的第 (1)(2) 步, 通过网络跳距离机制, 把被搜索的广域网络形成一个逻辑矩形区域。然后以该矩形区域的中心为圆心, 分别以 $r, 2r, 4r, \dots, 2mr, \dots$ 为半径作同心圆, 最大的同心圆将覆盖矩形区域 (算法 1 的 (3)⑥步); 在算法 1 的第 (5) 步, 以这些同心圆的弧线上的点为圆心, 以 r 为半径构造若干个互不相交, 并相邻互外切的圆形数据搜

索区域,这些圆形数据搜索区域覆盖了矩形区域(可以对每个圆形搜索区域适当扩大搜索半径 $r+\epsilon$ (ϵ 为一个很小的数值)形成微小重叠)。所以这种区域划分方法可以完成 DSG 对被搜索区域的数据搜索。根据算法 3 的扩充描述,随着同心圆区域扩展线(图 1 虚圆线所示)的无限扩大,扩充的圆形搜索区域将覆盖 Internet,所以说 DSG 可以完成对 Internet 的网络数据搜索。

定理 2 DSG 具有良好的可扩展性。

证明:根据定理 1 可知,DSG 在基于网络跳距离机制下,通过扩展可以覆盖 Internet。对每个圆形数据搜索区域内,通过四个状态(无效、薄弱、饱满、过强)和网络搜索节点密度来控制圆形数据搜索区域的规模;通过市场代价机制,来调节网络搜索节点的密度,使每个数据搜索区域的挖掘性能与本区域的被搜索的数据量成有效比例,所以不管 Internet 多大,圆形数据搜索区域确是本地的、性能均衡的;即圆形数据搜索区域仅与自身网络临近的网络信息状况有关,而与其在大同心圆的位置无关。

另一方面,通过构建数据服务器区域机制(算法 1 的第(11)步和算法 2 的第⑥步),构造二级管理区域,使 DSG 具有更好的区域自治性,而数据服务器区域可以根据规模动态调整。

综上所述,DSG 的挖掘性能仅和圆形数据搜索区域的性能有关,而与它们的个数无关,所以 DSG 具有良好的可扩展性。

定理 3 DSG 具有良好的挖掘效率。

证明:根据算法 4 所述,对一组关键字集合 KS 的搜索任务被并发地发送到 DSG 的不同 DSS 区域。各个 DSS 又把 KS 并发地发送到其所属的圆形数据搜索区域,最终是所有覆盖 Internet 的圆形数据搜索区域并行的进行数据搜索。所以在总体机制上是完全并行的。

在一个圆形数据搜索区域 A 内,其所有数据搜索节点并行地对相邻网络资源节点进行搜索。据算法 5 所述,通过搜索标志 SM_i 来保证每个被搜索站点不被重复搜索,即 A 内的所有资源站点仅被 A 的一个 DSN 搜索且必须搜索一次。

所以,DSG 具有良好的挖掘效率。

2.4 模拟实验环境简介

构建了由 10 台 PC 机和 3 个物理网络 A、B、C 构成的 3 个 DSS 区域来模拟一个实验 DSG。A 物理网络由 4 台 pc 构成,其中一台作 Master,B、C 物理网由 3 台 pc 构成。A 物理网络由固定 IP 地址的局域网络联入 Internet,B、C 网络由 FTTB 拨号联入 Internet。A、B、C 分别处于 Intranet 的 3 个不同的区域,即 A、B、C 由 WAN 连接构成。WAN 的通信数度在 128k/s~512k/s 范围内变化。每个 DSS 中的 3 台计算机都作为 DSN 运行。

Web Service 技术已经成为开发下一代互联网络的主要技术之一,其基于服务的思想恰好与 OGSA 规范相吻合。利用 JAVA(JAXML 1.2.) 和 Web Service 框架初步实现了本文提供的模型框架。

应用一个软件 WebTime 模拟了全局时钟系统 GTS,在

DSG 的每个 DSS 和 DSN 上安装 WebTime 完成时钟同步软件。

试验对 DSG 覆盖的网络资源进行了搜索,该模型的并行机制和覆盖机制得以验证。一旦市场机制成熟,将扩充到 Internet 进行广域搜索研究。

结束语 对不断日益扩大的 Internet 进行数据搜索,需要有效地管理、自动地调整数据搜索节点的分布。本模型所描述的基于网格的二级区域划分机制,具有良好的扩展性、较高的效率。通过市场代价机制,使大规模搜索资源的获取有了一个自动模式。本模型的研究证明了这一点。

参考文献

- 1 Napster Home Page. <http://www.mapster.com>
- 2 Gnutella Home page. <http://www.gnutella.com>
- 3 Baeza-Yates R, Ribeiro-neto B. Modern Information Retrieval. Boston: Addison Wesley, 1999
- 4 Stoica I, Morris R, Karger D, et al. Chord: A scalable peer-to-peer lookup service for internet applications. ACM SIGCOMM Computer Communication Review, 2001, 31(4): 149~160
- 5 Ripeanu C. Peer-to-Peer architecture case study: Gnutella network; [Technical Report]. TR-2001-26, University of Chicago, 2001
- 6 Freenet. <http://freenet.sourceforge.net>
- 7 Foster I, Kesselman C. The Grid: Blueprint for Future Computing Infrastructure [M]. San Francisco, USA: Morgan Kaufmann Publishers, 1999
- 8 Felten E W, Zahorjan J. Issues in the Implementation of a Remote Memory Paging System; [Technical Report]. 91-03-09. University of Washington, 1991
- 9 Markatos E P, Dramitinos G. Implementation of Reliable Remote Memory Pager. In: Proceedings of the 1996 Usenix Technical Conference [C], 1996. 177~190
- 10 Acharya A, Setia S. Using Idle Memory for Data-Intensive Computations. In: Proceedings of the 1998 ACM SIGMETRICS Conference on Measurement and Modeling of Computer Systems [C], June 1998, 278~279
- 11 DeFanti T, Foster I, Papka M E, et al. Overview of the I-WAY: Wide Area Visual Supercomputing. International Journal of Supercomputing Applications, 1996, 10(2)
- 12 <http://www.didc.lbl.gov/GGF-PERF/GMA-WG/papars/GWD-GP-16-2.pdf>
- 13 <http://www.didc.lbl.gov/pyGMA-text.html>
- 14 http://www.didc.lbl.gov/MDF/MDF-proposal_funded_public.pdf
- 15 Gaidioz B, Wolski R, Tourancheau B. Synchronizing Network Probes to Avoid Measurement Intrusiveness with the Network Weather Service. In: Proceedings of 9th IEEE High-performance Distributed Computing Conference, August 2000. 147~154
- 16 Foster I, Kesselman C, Nick J, et al. The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration. January 2002. <http://www.globus.org/research/papers/ogsa.pdf>