

S-Club 机制在网格信息服务系统中的实现

夏先波, 富公为, 刘旭东

(北京航空航天大学计算机新技术实验室, 北京 100083)

摘要: 介绍了 S-Club 原理, 分析了已有实现方法的不足, 提出一个适合在网格信息服务系统中实现 S-Club 机制的设计方案, 包括如何创建、维护和撤销 S-Club 以及如何在 S-Club 中进行查询。利用该套设计方案, 在 CROWN 信息服务系统 RLDS 中成功实现了 S-Club 机制, 通过实验证明了 S-Club 机制确实能够提高信息服务系统的查询效率, 减少综合网络开销。

关键词: 网格; 信息服务; 科学活动环境综合试验平台; RLDS; S-Club

Implementation of S-Club in Grid Information Service System

XIA Xianbo, FU Gongwei, LIU Xudong

(Institute of Advanced Computing Technology, Beijing University of Aeronautics and Astronautics, Beijing 100083)

【Abstract】 This paper introduces the S-Club mechanism and analyzes the deficiency of the original implementation method. It proposes a framework appropriate for applying S-Club mechanism in grid information service, which includes the procedure of creating, maintaining, destroying an S-Club and the query in an S-Club. According to the framework, S-Club mechanism is successfully implemented in the RLDS, the grid information service in CROWN Grid. Experiments show that with the introduction of the S-Club, query efficiency of RLDS is promoted significantly, and the overall cost of communication is reduced as well.

【Key words】 Grid; Information service; CROWN; RLDS; S-Club

网格是一种伴随着计算机、信息和网络技术的飞速发展而产生的新技术, 其目的是为了实现在广域网环境下资源的共享与协同; 为此, 需要建立网格应用与网格资源之间的动态组织机制, 支持资源发现和运行中资源状态监控^[1], 维护网格中各类实体(包括服务、硬件等)的相关信息^[2], 网格信息服务(GIS)应运而生, 构成了网格环境的重要基础。在基于SOA的服务网格中, 信息服务作为服务提供者和服务请求者之间的中介, 具有类似于服务注册表的职能, 是网格中许多上层服务(如资源服务的发现与选取、网格应用管理等)的基础。

以网络为基础的科学活动环境综合试验平台(CROWN)是面向国内各个领域的科学家的一个综合网络科研环境。CROWN 的基本设计思路是采用面向服务的体系结构, 将服务化的资源按照其物理隶属关系组织起来, 在已有的 TCP/IP 网络互联的基础上构造可控可管的资源层叠网, 实现资源能力的描述与发布、自动的服务发现和服务交互, 在广域网络中实现对资源能力的共享和综合利用。

CROWN 的信息服务组件是 RLDS, 它采用了森林结构的信息组织方式: 将网格划分为多个 Region, 每个 Region 内, 采用树型结构组织信息; 在 Region 之间, 采用 P2P 的方式组织信息。

通过 RLDS 进行资源发现时, 必须沿着森林的拓扑结构转发用户的请求, 效率相对较低, 因此在 RLDS 中实现了 S-Club 机制^[3], 通过在信息服务树中建立 S-Club, 减少用户请求被转发的次数, 从而提高资源发现的效率, 减少综合网络开销。

1 S-Club 介绍

1.1 基本原理

S-Club^[3]的基本原理是: 在现有 GIS 构成的底层网络上, 依据每个 GIS 所注册的服务类别信息, 对不同的 GIS 进行组织

构成面向服务类别的组, 并建立 GIS 之间的逻辑链接形成层叠网。当进行服务发现时, 请求首先在层叠网内部检索, 消息转发的范围被局限在相对较小的范围内, 提高服务发现的效率, 减少网络的综合开销。每一个 GIS 所构成的组将含有一类服务的 GIS 聚合在一起, 其中每个成员具有共同的特征, 类似人类社会中“俱乐部”的结构, 因此称为 Service Club, 简称 S-Club。S-Club 示例如图 1 所示。

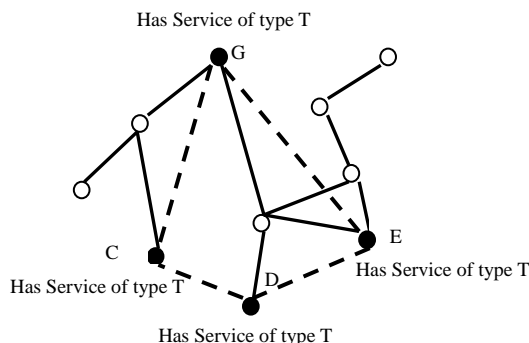


图 1 S-Club 示例

在图 1 中, 空心和实心的节点代表不同的 GIS, 其中实心的 GIS(C、D、E、G)表示其服务注册信息中含有类型为 T 的服务而构成一个层叠网(如图 1 中虚线所示)。

基金项目: 国家自然科学基金委重大研究计划“网络计算环境综合试验平台”(90412011); 国家“863”计划基金资助项目“中英网络试验平台核心技术及应用研究”(2005AA119010)

作者简介: 夏先波(1982-), 男, 硕士生, 主研方向: 网格计算, Web 服务; 富公为, 硕士生; 刘旭东, 教授

收稿日期: 2006-05-17 E-mail: xiAXB@act.buaa.edu.cn

1.2 实现方法

S-Club思想的作者提出了一种实现方法^[3]：

(1)服务的分类方法。用 portType 区分服务，具有相同 portType 的服务表示同一类服务。

(2)S-Club 的创建。每个 GIS 在本地为每个注册的服务类型维护一个使用频度，每个 GIS 周期性地对服务的使用频度进行排序，为使用频度高于某个阈值的服务类型 T 发起创建 S-Club 的请求，将请求转发给其他 GIS，每个 GIS 根据自身的统计情况决定是否同意创建 S-Club，这是一个类似投票的过程，发起者统计投票结果，如果赞成率大于某个阈值，便创建 S-Club，否则终止创建过程。发起者将同意创建 S-Club 的 GIS 当作该 Club 的成员，将成员列表发给每个 GIS，每个 GIS 从列表选取距离自己最近的若干个 GIS 作为 S-Club 的联系地址。

(3)S-Club 的维护。当 GIS 加入 S-Club 时，向 S-Club 成员发送加入请求，所有成员更新 S-Club 的联系地址。当 GIS 离开 S-Club 时，向所有 GIS 发送离开请求，GIS 收到该请求后更新 S-Club 的联系地址。

(4)S-Club 的撤销。当 GIS 发现 T 类型服务的使用频度低于某个频度阈值时，便发起撤销 S-Club 的请求，和创建 S-Club 的过程类似。

(5)基于 S-Club 的查询。当 GIS 收到查询类型为 T 的服务请求时，根据 S-Club 的联系地址，将请求转发给联系地址中的任一成员 GIS，该成员再根据自己保存的联系地址向其他成员 GIS 转发请求。

2 设计要点

从系统实现角度来看，1.2 节提出的实现方法不宜在 RLDS 中直接采用，具有如下缺陷：

(1)用 portType 区分服务不合理。portType 能区分服务，但 portType 是一段 WSDL 文档，比较两段 portType 不直观，不方便。

(2)根据服务的使用频度建立 S-Club 不合理。为记录服务的使用频度，必须要求每个服务在被调用时向其注册的 GIS 进行报告。这一过程既花费网络带宽，又不合规范，因为网格信息服务并没有规定服务提供者必须汇报服务被调用的次数。

(3)采用投票机制不合理。每个 GIS 实例都根据自身统计的使用频度发起创建和撤销 S-Club 请求，这是局部观测的结果，很显然，很多创建 S-Club 请求会被否决，然而每次请求无论成功与否都会引起多次服务调用，网络带宽浪费严重。

(4)S-Club 查询时需要转发请求不合理。在进行 S-Club 查询时，需要根据各 GIS 实例保存的联系地址进行请求转发。当用户指定结果记录的数目时，很难控制转的结束时间，引起不必要的查询，增加响应时间，浪费网络带宽。

针对以上问题，结合了 RLDS 本身的树型结构特点，本文设计出如下的解决方案：

(1)用 classID 标识一类具有相同 portType 的服务，它是一个 URN，服务提供者在注册时指明服务所属的 classID。

(2)为查询频度高的服务类型建立 S-Club。一般来说，当一类服务被查询的频度较高时，其使用频度也会较高，因此，可以用服务的查询频度取代使用频度作为建立 S-Club 的依据，以避免让服务提供者向 GIS 汇报服务被使用的次数。为查询频度高的服务建立 S-Club，其利用率就高，而在 S-Club 内进行查询的效率很高，因此整个信息服务系统的查询效率

也得到提升。

(3)通过 SClubCenter 创建和撤销 S-Club。在每个 Region 内设置一个专门负责创建和撤销 S-Club 的 SClubCenter。对于用户的查询请求，GIS 维护查询结果中各类服务的查询频度。SClubCenter 定期收集 Region 内所有 GIS 的服务查询频度并排序，为查询频度高的服务类建立 S-Club，同时，撤销查询频度低的服务类的 S-Club。通过这种集中式的方式，避免了投票机制中的无效投票引起的网络开销，因为 SClubCenter 得到的服务查询频度信息是全局的而不是局部的。

(4)每个 GIS 实例保存 S-Club 的所有成员地址，在进行 S-Club 查询的时候，采用依次转发的策略，将请求依次转发给每个成员 GIS 实例，这样能控制结果集的数目，减少无谓的请求转发。

3 具体实现

3.1 S-Club 的创建与撤销

SClubCenter 决定 S-Club 的创建与撤销，它收集并统计整个 Region 内的服务查询频度，为查询频度高的服务类建立 S-Club，同时撤销查询频度低的服务类的 S-Club。在该过程中，SClubCenter 需要与 Region 内的所有 GIS 实例交互，其创建过程如图 2 所示。

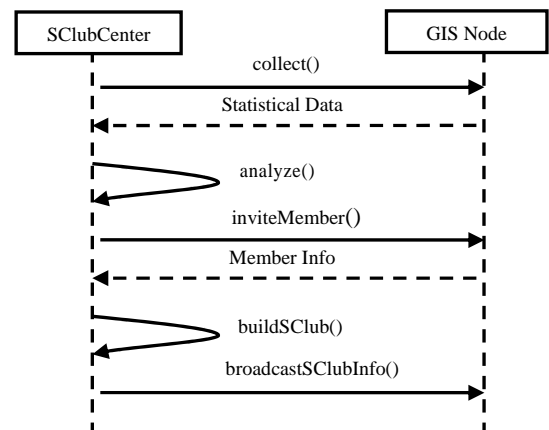


图 2 S-Club 的创建过程

该流程可以描述为：

(1)每个 GIS 在处理用户的查询请求后，将查询结果的信息记入日志。

(2)SClubCenter 每隔固定时间 t(t 可配置，默认为 2h)向所有 GIS 发送“收集近 t 内服务被查询的情况”的请求。

(3)GIS 将近 t 内记录的服务被查询的情况返回给 SClubCenter。

(4)SClubCenter 统计所有服务的查询频度并排名，分析出需要建立 S-Club 的 classID。

(5)SClubCenter 将需要建立 S-Club 的 classID 的信息发布给所有 GIS，邀请 GIS 加入 S-Club。

(6)GIS 收到邀请后，判断本地是否注册了相应的服务，如果有，便向 SClubCenter 发送同意加入 S-Club 的消息，否则，忽略该邀请。

(7)SClubCenter 根据所有 GIS 的反馈信息构造 S-Club，将同意参加 S-Club 的 GIS 添加到 S-Club 的成员列表中。

(8)SClubCenter 将构造好的 S-Club 信息广播给所有 GIS。GIS 收到消息后将 S-Club 的信息存储在本地。S-Club 创建过程结束。

S-Club 的撤销过程与创建过程类似。当 SClubCenter 统

计服务的查询频度后,分析出排名靠后的 classID,如果有 S-Club 便撤销。

3.2 S-Club 的维护

S-Club 的维护包括成员的加入和离开。当 GIS 新注册某类服务,且存在该类服务的 S-Club 时,便加入该 S-Club,向其他 GIS 发送“我加入 x S-Club”的消息,GIS 收到消息后更新本地关于 x S-Club 成员的信息。

当 GIS 上某类服务全部被反注册,且自己属于该类服务的 S-Club 时,便退出 S-Club,退出过程与加入过程类似,需要向其他 GIS 发送“我退出 x S-Club”的消息。

3.3 S-Club 查询

当 GIS 收到用户的查询请求时,首先分析请求所涉及到的服务的 classID,然后判断是否为该类服务创建了 S-Club,如果没有,进行一般的查询过程,否则进行 S-Club 查询,从内存中取出该 S-Club 的成员列表,将请求依次转发给列表中的 GIS。如果用户指明了查询的响应时间或结果记录数目,请求接收者在收到 S-Club 成员的响应消息后,判断当前的查询时间和结果记录数目,如果满足了用户要求,便停止请求的转发,直接将结果返回给用户,否则将请求转发给下一个 S-Club 成员。

4 性能评测

本文从两方面来评测 S-Club 机制对网格信息服务系统的影响,一方面评测它对信息查询效率的影响,检验引入 S-Club 是否提高了信息服务系统的查询效率;另一方面评测它对信息服务系统的综合网络开销的影响,因为无论是 S-Club 的建立、撤销还是维护都需要花费网络带宽。信息查询效率可用查询的平均响应时间来衡量,平均响应时间越长,查询效率就越低;综合网络开销可用信息服务系统中单位时间内 SOAP 报文的字节数来衡量,单位时间内 SOAP 报文字节越大,综合网络开销就越高。

实验如下:在集群 32 个节点上部署 RLDS 信息服务,组成一棵信息服务树,在其中一个节点上部署 SClubCenter 服务。在每个 RLDS 节点上注册若干条服务的信息,其中 classID 为“crown:test”的服务信息的数目服从[0, 3]的均匀分布。

在实验 1 中,不启动 SClubCenter 服务即不采用 S-Club 机制,每隔 30s 进行一次服务查询,总共查询 100 次,要求返回 10 条 classID 为“crown:test”的服务信息。在实验 2 中,启动 SClubCenter 服务,SClubCenter 设置为每 20min 统计一次服务的查询频度,然后进行与实验 1 相同的查询。在实验中记录查询的响应时间和系统中 SOAP 报文的大小,得到响应时间和 SOAP 报文字节数与时间的关系,见图 3 和图 4。

从图 3 可看出,20min 后,因为关于 classID 为“crown:test”的服务的 S-Club 已建立,查询的响应时间急剧下降,这说明采用 S-Club 机制能够提高查询的效率。

从图 4 可以看出,在第 20min 时,SOAP 报文字节数很大,是因为 SClubCenter 需要从所有 RLDS 节点获取各类服务的查询频度,当 S-Club 建立后,还需将 S-Club 信息发送给每个 RLDS 节点,所以造成 SOAP 报文增多。在第 40min 时,SOAP 报文字节数也很大,但比第 20min 时要少很多,这是因为此时 SClubCenter 虽然需要从所有 RLDS 节点获取各类服务的查询频度,但由于 classID 为“crown:test”的服

务类的 S-Club 已经建立,不需建立新 S-Club,因此不用向 RLDS 节点发送 S-Club 信息,所以 SOAP 报文字节数没有第 20min 时大。

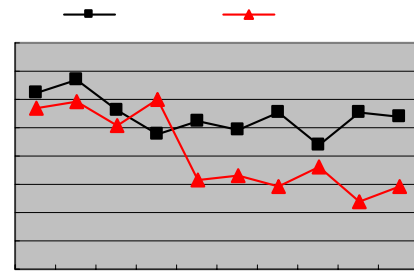


图 3 响应时间与时间的关系

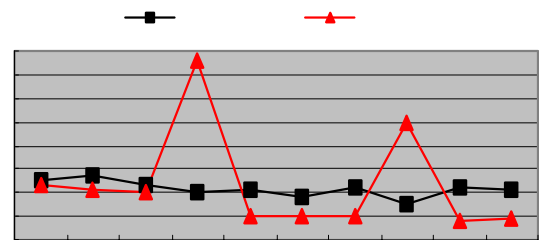


图 4 SOAP 报文大小与时间的关系

当 S-Club 建立后,SOAP 报文的字节数明显减少,这是因为进行 S-Club 查询减少了请求转发的次数。因此,只要设置适当的创建和撤销 S-Club 的时间间隔,就能降低信息服务系统的网络开销。

5 总结

本文分析了 S-Club 机制的作者给出的实现方法的不足,提出了一个适合在网格信息服务系统中实现 S-Club 机制的设计方案,并介绍了如何在信息服务系统中实现 S-Club 机制,最后通过实验证明了 S-Club 机制确实能提高信息服务系统的查询效率,减少综合网络开销。

参考文献

- 1 Czajkowski K, Fitzgerald S, Foster I, et al. Grid Information Services for Distributed Resource Sharing[C]//Proceedings of the 10th IEEE Int'l Symposium on High Performance Distributed Computing. IEEE Press, 2001: 181-184.
- 2 Plale B, Dinda P, Laszewski G. Key Concepts and Services of a Grid Information Service[C]//Proceedings of ISCA the 15th International Parallel and Distributed Computing Systems. 2002.
- 3 Hu Chunming, Zhu Yanmin, Huai Jinpeng, et al. S-Club: An Overlay-based Efficient Service Discovery Mechanism in CROWN Grid [C]// Proceedings of 2005 IEEE International Conference on e-Business Engineering. 2005: 441-448.