

基于 P2P 网络的 Web Services 发现机制

李 娟, 明德廷, 邱晓红

(江西农业大学 计算机与信息工程学院, 江西 南昌 330045)

摘 要: 提出一种改进的两层混合式 P2P 网络模型, 用该模型作为 Web Services 发现技术的底层网络. 该网络模型主要在两方面进行改进: 在网络结构方面提出 CP 节点分类策略和 SP 节点分组策略; 在网络组织方面提出一种基于 Web Services 兴趣因子空间的节点分布自调整策略.

关键词: P2P; Web Services; 聚类; 兴趣因子空间

中图分类号: TP311.3

文献标识码: A

文章编号: 1000-7180(2010)02-0152-04

A Web Services Discovery Mechanism Based on Peer-to-Peer Network

LI Juan, MING De-ting, QIU Xiao-hong

(Department of Computer and Information, Jiangxi Agricultural University, Nanchang 330045, China)

Abstract: An improved two-tier hybrid P2P network model is proposed on the basis of carefully analyzing the development, advantages and disadvantages of the existing P2P network, and used as the underlying-network foundation of the discovery techniques of Web Services. It mainly improve two aspects of the traditional method: firstly, the CP nodes classification strategy and SP nodes classification strategy are proposed in the net work structure; secondly, a self-adjusting strategy based on user-interested spaces is proposed in the network organization and corresponding searching algorithms of the Web Services.

Key words: peer to peer; web services; clustering; interesting gene space

1 引言

由于用户在使用 Web 页面和使用 Web Services 上具有极高的相似性, 文中提出将兴趣空间的理论和其聚类方法应用在 Web Services 发现机制中. 与文献[1]不同的是文中使用 Web Services 兴趣空间, 该兴趣空间以 Web Services 兴趣为坐标轴, 每个用户的坐标是其在各个 Web Services 兴趣上的投影. 用这个坐标系来表示所有用户, 使得每个用户是 Web Services 兴趣空间中的一个点. 由此, 对该兴趣空间中的行向量聚类就是对用户的聚类. 基于文献[2]中的试验分析, 在构造 Web Services 兴趣空间的方法上文中用因子分析理论构造 Web Services 兴趣

空间(简称“Web Services 兴趣因子空间”).

2 双层混合式 P2P 网络

通过对现有的 P2P 网络测量, 已经发现网络上节点的处理能力和行为方式差别很大^[1,3]. P2P^[4]网络上大部分节点在线时间非常短, 带宽很小, 也有一小部分节点在线时间很长, 有很大的存储能力和较多的带宽资源. 这种节点的异质性(heterogeneity)表明, 需要将它们赋予不同的角色, 让能力强的节点担当更多任务. 混合式 P2P 结构就是基于这种考虑的设计.

为了能够适应高度动态性, 同时又能进行高效搜索, 文中的发现机制采用两层混合式 P2P 网络结

构,如图1所示.一层是由超级对等体(Super Peer, SP)组成的有结构 P2P 网络,称这层网络为上层网络;另一层是由普通对等体(Common Peer, CP)组成的 P2P 网络,称为下层网络.

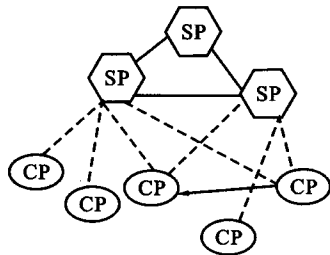


图1 两层混合式 P2P 网络结构

超级对等体(SP):由 P2P 网络上可靠性高、在线时间长、变动少的节点组成.所有的 SP 节点形成封闭环状结构.由于 SP 节点和 CP 节点存在索引关系,所以 SP 节点在加入建立邻居关系时需要遵循一定的规则,SP 节点也不能随意离开,离开时需要进行修复操作.也正因为要在离开时要修复,所以 SP 节点要选择由在线时间长、变动少的节点组成^[5].

普通对等体(CP):可以是任意的网络节点,对普通节点没有任何要求. CP 节点只可以和一个 SP 节点连接建立邻居关系,CP 节点和哪个 SP 节点建立邻居关系是由节点分布自调整策略决定的.

上层网络是由 SP 节点组成的网络,每个 SP 节点都保存有全局所有 SP 节点的地址.由此,每一个 SP 节点都能迅速的定位到其他 SP 节点.此种网络中 SP 节点的加入和离开都会影响上层网络的结构,需要对上层网络作适当的修复操作.

这种两层混合式 P2P 网络利用了节点异质性的特点,发挥了系统的节点性能,但缺点是 CP 节点不能任意的加入此网络,因为每次加入都需要进行网络的修复工作,从而影响了网络的动态性.为解决此问题,文中提出一种基于 CP 节点加入网络时间的 CP 节点分类策略.

3 CP 节点分类策略

文中提出的双层混合式 P2P 网络中的 CP 节点在网络中的位置是由全部节点构成的兴趣因子空间以及其自身的兴趣向量决定的.其中的兴趣因子空间是根据网络内所有 CP 节点的兴趣集(即节点抽取的服务特征项)计算而得,所以 CP 节点的加入是会影响网络的整体结构.然而 P2P 网络正是一个高动态的网络,其节点进出很频繁.这就产生了矛盾,不可能在每加入或者退出一个节点时都重新进

行 Web Services 兴趣因子空间计算,因为那样的计算量是无法估计的.针对这个问题,文中提出按照节点加入网络的时间将节点进行分类处理^[6].

文中提出的基于 Web Services 兴趣因子空间的节点分布自调整策略正是要在一段时间内保持整体网络的网络拓扑、兴趣空间和兴趣划分不变.其大致思路为,当节点加入或者离开时,网络的拓扑结构、兴趣空间和兴趣划分保持不变,经过一定时间周期后,系统选择一个系统负载较轻的时刻对全部节点进行兴趣因子空间计算并进行节点的调整.基于这种策略,将 CP 节点按照加入网络的时间分为新节点 NCP 和稳定节点 OCP.

为了对两种节点分别进行处理,这里将两层混合式 P2P 网络进行改进.从上层网络中选出部分 SP 节点作为整个网络的入口节点 EP,组成入口网络.入口网络结构如图2所示.

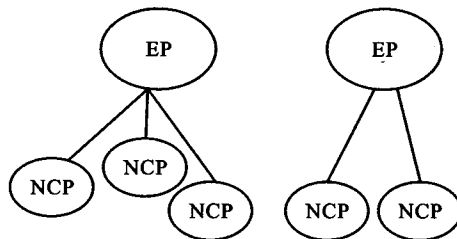


图2 入口网络结构图

当新的 CP 节点加入时,系统将根据网络连接速度和入口节点的负载能力来决定将 CP 节点作为 NCP 节点注册在哪一个 EP 节点上,并在该 EP 节点上记录该 NCP 节点加入网络的时间.当节点分布自调整策略在某一个时间点对整个系统进行自调整时,系统将根据 EP 节点上记录的 NCP 节点的加入时间来从这些 NCP 节点中选取那些加入时间大于某一阈值(该阈值可由经验给出)的点,这样的 NCP 节点是较为稳定的节点.将这些选取出的 NCP 节点加入到稳定网络,并重新进行兴趣因子空间计算,之后对网络进行调整.

4 SP 节点分组策略

Web Services 兴趣因子空间的用户聚类按用户对 Web Services 兴趣的相似性将用户分组,使得同组的用户有高度的相似性,不同组的用户相似性很小.由于 Web Services 特征项的随机性和热点性决定了这些分组不可能做到分布均衡,以至有些分组拥有较多 Web Services 用户节点(大聚类),而有些分组则拥有相对少的 Web Services 用户节点(小聚类).如果让 SP

节点与聚类一一对应势必会造成负责大聚类的 SP 节点具有过高的负载,而那些负责小聚类的 SP 节点出现大量的空闲时间.这样会使得网络的负载极度不平衡,也使资源出现了大量的浪费.

为解决这个问题,文中将 SP 节点根据计算出的聚类进行分组,用分组和聚类一一对应.这样与大聚类所对应的分组便可以分配较多 SP 节点,与小聚类对应的分组则使用较少的 SP 节点组成.并在分组后的 SP 节点中选出一个性能最好的节点作为该分组的中心节点 CSP,每个 CSP 节点存储全局 CSP 节点.这样原来的上层环状网络就变成了一个以分组为单位的全连通网络.SP 节点的分组方法将在基于用户兴趣因子空间的节点分布自调整策略中进行描述.为了有更直观的理解,图 3 描绘了这种分组策略.

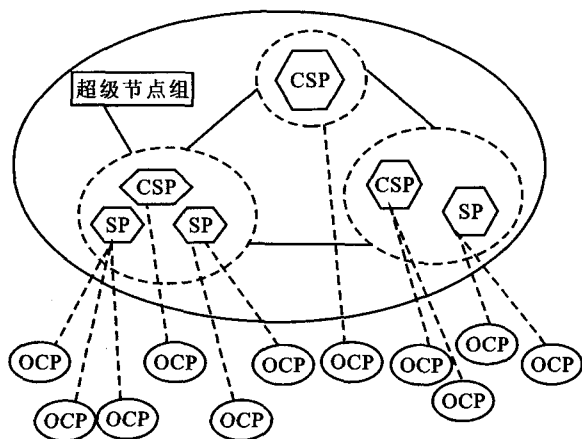


图 3 SP 分组策略结构图

5 基于 Web Services 兴趣因子空间的节点分布自调整策略

用 Web Services 兴趣因子空间中用户节点聚类的算法^[7]可以得到 m 个用户聚类.文中提出的分布自调整策略就是基于这些聚类进行的.

由于本系统的节点数量庞大并且网络节点是高度动态性的,所以每当有节点加入或者退出网络时都重新进行向量空间的计算是不现实的,那会造成整个系统的低性能甚至是瘫痪.故文中提出将 CP 节点进行分类,并采用一段周期时间后对向量空间进行计算并对节点逻辑分布进行调整的策略.该策略具体过程如下:

(1) 经过一个指定的时间周期(如一天)后,稳定网络中的所有 CSP 节点在特定时间(如午夜)相互协商并推举一个 CSP 节点收集所有符合条件节

点的 Web Services 特征项并构造 Web Services 兴趣因子空间,然后按用户节点聚类算法将所有符合条件的节点进行聚类.其中符合条件的节点是指稳定网络中的所有 OCP 节点和加入时间大于阈值 T 的 NCP 节点.

(2) 将所有聚类拥有的节点数目相加得到参加聚类的节点数目 N_a .

(3) 计算出每个 SP 节点应承担的平均节点数 $N_{avg} = \lceil N_a / N_s \rceil$,其中 N_s 为 SP 节点的数目,本系统中 SP 节点由人工指定.

(4) 分别计算每一个聚类对应 SP 节点组所应包含的 SP 节点数目,计算公式为 $\lfloor \text{节点组包含节点数} / N_{avg} \rfloor$.

(5) 按步骤(4)计算出的节点数目为每一个节点组分配 SP 节点,同一分组的 SP 节点选取遵循物理位置接近的原则.将同一聚类中的节点均分给 SP 节点,并将同组 SP 节点列表和下属节点列表发送至相应的 SP 节点.

(6) 从每一组 SP 节点中选取出一个作为 CSP 节点,向其发送全局 CSP 节点列表,并向所有 SP 节点发送其所在分组的 CSP 节点地址.

(7) 当 SP 和 CSP 节点收到列表后清除原有列表,按新列表重新完成 OCP 节点注册.

(8) 将已经移至稳定网络结构的 NCP 节点在 EP 节点上的注册进行注销.

(9) 全部自调整过程结束.

这种策略的好处是将兴趣较为相似的节点放置在逻辑位置比较接近的地方,以便在节点进行 Web Services 查找时在逻辑上较快的发现具有最大可能提供服务的提供者.这种做法不仅可以提高 Web Services 的查找效率,也有效减少了网络流量的压力,并保证了系统的稳定性.

6 结束语

对用户节点进行聚类使得在每一个聚类中的用户都具有较高的相似度,文中将兴趣空间理论引入到 Web Services 发现机制的设计中正是要利用这个特性.对于同属于一个聚类中的节点,将它们放在逻辑结构比较近的节点位置,使网络结构中小范围具有高度的相似性以便于在服务查找时具有较高的效率和性能.

参考文献:

- [1] Ratnasamy S, Francis P, Handley M, et al. A scalable con-

- tent - addressable network [C]// Proc. ACM Sigcomm USA: San Diego, 2001.
- [2] Zhao B Y, ubiatowicz J K, Joseph A. Tapestry: an infrastructure for fault - tolerant wide - area location and routing [R]. University of California at Berkeley, Computer Science Department: Technical Report UCB/CSD - 01 - 1141, 2001.
- [3] 廖祝华, 刘建勋, 易爱平. 基于用户兴趣的Web服务发现[J]. 微电子学与计算机, 2006, 23(21): 23 - 25.
- [4] Verma K, Sivashanmugam K, Sheth A, et al. Meteor - s Wsdi: a scalable infrastructure of registries for semantic publication and discovery of Web Services[J]. Special Issue on Universal Global Integration, 2005, 6(1): 7 - 39.
- [5] Serge Abiteboul, Ioana Manolescu, Nicoleta Preda. Construction and querying peer - to - peer warehouses of XML resources[C]// icde: 21st International Conference on Data Engineering (ICDE'05). Japan: Tokyo, 2005: 1122 - 1123.
- [6] Nizamuddin Channa, Shanping Li, Wei Shi, et al. A CAN-Based P2P infrastructure for semantic web services publishing and discovery[C]// The First IEEE and IFIP International Conference in Central. Asia, 2005: 26 - 28.
- [7] 袁冠, 夏士雄, 张磊, 等. 基于兴趣度的Web用户聚类方法[J]. 微电子学与计算机, 2008, 25(9): 112 - 114.

作者简介:

李娟 女, (1971 -), 讲师. 研究方向为分布式系统、P2P网络.

明德廷 男, (1972 -), 讲师. 研究方向为计算机网络与安全、Web services数据挖掘.

邱晓红 男, (1967 -), 教授. 研究方向为网络与信息安全、计算复杂性理论.

(上接第151页)

- [3] 张守娟, 周詮. 基于加权小波分形的卫星遥感图像飞机类型识别[J]. 微电子学与计算机, 2007, 24(6): 1703 - 175.
- [4] 王庆华, 张兴彪, 张洪朋, 等. 分形理论在液压泵故障诊断中的应用[J]. 大连海事大学学报, 2004, 30(2): 40 - 43.
- [5] 叶菲, 罗景青, 海磊. 基于分形维数的雷达信号脉内调制方式识别[J]. 计算机工程与应用, 2008, 44(15): 155 - 157.
- [6] 陈冬冰, 吴平东, 毕路拯, 等. EEG非线性特征参数的研究[J]. 北京生物医学工程, 2006, 25(6): 611 - 614.
- [7] Grassberger P, Procaccia I. Characterization of strange attractor[J]. Phys Rev Lett, 1983, 50(5): 346 - 349.
- [8] Maria Hansson, Tomas Gansler, Goran Salomonsson. A system for tracking changes in the mid-latency evoked potentials during anesthesia [J]. IEEE transactions on biomedical engineering, 1998, 45(3): 323 - 334.

作者简介:

张烈平 男, (1971 -), 博士研究生, 副教授. 研究方向为计算机应用技术.

莫玮 男, (1956 -), 博士, 教授, 博士生导师. 研究方向为信号与信息处理、自动测试技术与系统.