

基于移动代理的网络拓扑发现技术的研究^{*})

邓 勇¹ 王汝传^{1,2} 黄海平¹ 徐喜春¹

(南京邮电大学计算机学院 南京 210003)¹

(南京大学计算机软件新技术国家重点实验室 南京 210093)²

摘要 随着 Internet/Intranet 的迅猛发展,网络结构日趋复杂,各种应用服务不断增多,网络规模也随之不断扩大,得到一个完整准确的网络拓扑结构图对于网络管理、网络优化、定位故障等应用越来越重要。国内在网络拓扑发现技术方面的研究仍处于初级阶段,因而必要性和紧迫性不容忽视。将移动代理这一分布式计算技术引入到网络拓扑发现领域,提出了基于移动代理网络拓扑发现算法,这种算法以 TCP/IP 网络管理体系为依托,将各种可以利用的协议和移动代理的特性有机地结合起来,旨在发现速度和发现准确度上有更高的要求。与传统的面向协议的方法不同,它是一种面向过程的方法,分层次地获得网络拓扑,该方法与网络分层结构相一致,能更好地体现网络层次。基于移动代理的网络拓扑发现方案不同于以往的拓扑发现算法,它体现了集中与分布的统一,适合于当今在地理上越来越分布的网络环境,具有很好的实用和研究价值。

关键词 移动代理,拓扑发现,网络管理,简单网络管理协议

Research on the Technology of Network Topology Discovery Based on Mobile Agent

DENG Yong¹ WANG Ru-chuan^{1,2} HUANG Hai-ping¹ XU Xi-chun¹

(Dept. of Computer Science and Technology, NJUPT, Nanjing 210003, China)¹

(State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China)²

Abstract With the rapid development of Internet/Intranet and the wide applications of the network, the network structure is becoming complicated and the network scale is extended rapidly. Therefore, it's more and more important to get a complete and accurate topology structure map, which is used in these fields such as network management, network optimization and trouble shooting. The research on topology discovery technique is in its infancy in our country, so it's very necessary to consider its necessity and urgency. This paper introduces mobile agent technology into the field of network topology discovery and puts forward an algorithm of network topology discovery based on mobile agent. The algorithm integrates various of available protocols with characteristics of mobile agent, which can enhance the speed and accuracy of network topology discovery. Different from the traditional methods which are protocol oriented, the algorithm based on mobile agent is process oriented. It is the same as the network architecture. Therefore it can help incarnate network layers. This scheme is different from the traditional topology discovery algorithms, which is unify of centralization and distributing. It is also suitable for network environment which is more and more distributed geographically, from these it can be seen that the scheme has practicality and research value very much.

Keywords Mobile agent, Topology discovery, Network management, SNMP

1 引言

网络拓扑发现是网络管理功能的基本组成部分之一,也是衡量商业网络管理系统成败的尺度。拓扑发现在搜索网络瓶颈,解决交通堵塞,排查潜在的致命点,增强网络连接性和网络升级方面都是至关重要的。另外,拓扑发现还可以用于实时网络仿真和帮助用户决定从何处切入网络以获得最大带宽,对开发新型网络协议和算法也是很有利的。

目前,拓扑发现的算法都是通过 SNMP(Simple Network Management Protocol,简单网络管理协议)来读取路由器中

的信息,现在用作网关的网络设备都支持 SNMP 代理,网络拓扑信息主要包含在 MIB-II 中,通过对 MIB-II 中的拓扑信息的获取^[1],由于路由表中的下一站地址项所标识的必然具有路由功能的网络节点,因此从管理工作站的缺省路由器开始,通过读取路由器的路由表,可逐步向下发现网络中的所有具有路由功能的网络节点。在子网发现时,利用了子网地址和子网掩码得出 IP 地址范围,而后向特定的 IP 地址范围发送 ping 消息,接受响应来建立网络视图。这种方法发现效率较高,系统和网络开销小,易于实现。但是由于这种方法是对 IP 的辨别,不管是否实际连接了一个主机都要 ping 每一个

^{*}) 本课题得到国家自然科学基金(60573141 和 60773041),江苏省自然科学基金(BK2005146),江苏省高新技术研究计划(BG2004004, BG2005038, BG2006001),国家高科技 863 项目(2006AA01Z201, 2006AA01Z219),南京市高科技项目(2007 软资 127),现代通信国家重点实验室基金(9140C1101010603),江苏省计算机信息处理技术重点实验室基金(kjs050001, kjs06006)和江苏省高校自然科学基金研究计划(07KJB520083)和南京邮电大学科研项目(NY206041)资助。邓 勇 博士研究生,研究方向为计算机网络和网络技术、信息安全和移动代理技术等;王汝传 教授,博士生导师,主要研究方向为计算机软件、计算机网络、信息安全、移动代理和虚拟现实技术等;黄海平 博士生,主要研究方向为计算机网络、计算机软件在通信中的应用和信息安全等;徐喜春 硕士研究生,主要研究方向为计算机网络及其在通信中的应用。

IP 地址,浪费时间^[3],而且由于路由器中的信息是每隔一段时间才更新一次,因此读取的信息并不准确^[4]。

随着网络复杂性的增加,利用传统的 C/S 方法很难满足要求。因此可以考虑通过产生一定数量的移动代理注入到网络中并以预定的迁移算法在网络中移动。本文在研究了当前主要的拓扑发现方法的基础上,给出了基于移动代理的网络拓扑发现算法,并阐述了拓扑发现系统的具体设计与实现,验证了此方案的可行性。

2 当前主要拓扑发现方法研究及评价

2.1 基于 ARP 协议的拓扑发现方法^[2]

ARP 协议的广泛应用源于以太网的普遍存在,任何有以太网接口的网络设备都必须支持地址解析协议(ARP),并在本机维护着一张 ARP 表,用于 IP 地址与以太网地址之间的地址解析和转换^[3]。

根据任何一台路由器或者交换机的 ARP 表,可以发现与其各以太网端口相连的以太局域网中的所有网络设备,再判定网络中的路由器和交换机,并继续根据 ARP 表进行发现,从而得到整个以太网络的拓扑结构关系。

这种方法适合于局域网发现,发现效率很高,但不适用于网络过大、不能发现那些不支持 ARP 协议的网络连接和设备。

2.2 基于 SNMP 的路由表的拓扑发现算法

SNMP 是一种基于 UDP 协议的协议,利用它可以从网管设备中提取用来支持网络管理(包括路由器、主机等)的 MIB 信息:MIB-II 包含路由表(如表 1 所示),利用路由表中的数据可以提取出许多重要的拓扑信息^[4]。但是,如果要访问路由器,必须通过认证。现在用作网关的网络设备都支持 SNMP 代理,网络拓扑信息主要包含在 MIB-II 中,通过对 MIB-II 中的拓扑信息的获取,就可以分析出网络的拓扑连接情况^[1]。

表 1 路由表

IpRouteDest	IpRouteIfIndex	IpRouteNextHop	IpRouteMask
INTEGER	INTEGER	INTEGER	IpAddress

IpRouteDest 路由的目的地址,如为 0.0.0.0 表示该路由为缺省路由。

IpRouteIfIndex 路由的当地接口索引,唯一地指明通往下一路由器的当地接口。

IpRouteMask 路由目的地址子网掩码,IpAddress 类型。

从 IpRouteNextHop 可以得到与该设备直接连接的所有设备,但是不能得到哪个端口连接到哪个设备,不能实现端口级的网络拓扑的显示。

2.3 基于 ICMP 的方法

ICMP 允许主机或路由器报告差错情况和提供有关异常情况的报告。但 ICMP 不是高层协议,它仍是 IP 层中的协议,ICMP 报文作为互连网层数据报的数据,加上数据报的首部,组成 IP 数据报发送出去^[5]。ICMP 报文的类型很多,总体上可分成两种类型,即 ICMP 差错报文和 ICMP 询问报文。一般在拓扑发现中,仅用到 2 种报文——回应(ECHO)请求或回应应答。回应请求/应答被用于测试信宿机的可达性或连通性。

基于 ICMP 协议的工具是一个主要的提取路由信息的工

具,主要是利用 2 种基于 ICMP 协议的工具:PING 来检测主机是否可达;TRACEROUTER 来发现给定主机的路由信息。

3 基于移动代理的网络拓扑发现算法

3.1 问题建模

针对不同的预测和改善 Internet 性能的目的,建立合适的 Internet 拓扑模型是非常重要的,可以将整个 Internet 抽象为一个图,将一个子网想象成一个图的顶点,并以它的边界路由器为代表体现在整个 Internet 网络图中,如图 1 所示。如关心子网内拓扑情况,同样可以将子网内的设备细化成与路由节点相关联的一系列节点。本算法将以本地子网为出发点,首先进行一级拓扑发现,然后针对子网进行二级拓扑发现。

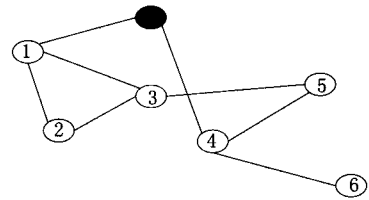


图 1 Internet 抽象图

3.2 算法描述

本算法采用基于分层的树形结构组织各个子图。网络节点之间的连接问题可以转换为图的方式加以考虑,对路由器的发现采用图的深度优先遍历算法的思想。一级拓扑发现的同时,利用移动代理的移动性,智能性和自治性在子网内进行二级拓扑的发现。

3.2.1 一级拓扑发现算法

一级拓扑发现将获得被拓扑网络的主干结构,一级拓扑发现就是对网络中具有 IP 路由能力的网络设备进行拓扑发现,是一种中粒度的发现。这些设备包括路由器,有路由能力的主机(启用路由协议的服务器和代理服务器等),通过在目标网络中注入路由发现代理发起洪泛实现,结果由报告代理传回。在路由发现代理中,设定了节点类(Node)、子网队列和当前节点的邻居节点列表,分别说明如下:

(1) Node:封装有关的节点信息,如节点名称,节点类型,特别是记录节点的多个接口的 IP 地址及其对应的子网地址。

(2) 子网队列:记录和本地相连的子网信息,如子网地址和子网掩码。

(3) 当前节点的邻居列表:记录路由发现代理遍历过程中与本地节点相邻的节点列表。

对路由发现代理行为算法的描述如下:

1. 如果路由发现代理进行首次发现时,路由发现代理将所在子网的缺省路由器节点定义为类 Node 的一个实例,即作为“Home Node”,TTL 计时开始,当获得邻接的节点信息后,将其添加到当前节点的邻居列表中。路由发现代理复制自身,将邻居列表中的每一个元素作为每个子代理的目的地址,将它们派遣出去。

2. 路由发现代理携带 TTL 和代理最近搜集的拓扑信息到达该节点后,将该节点数据区加锁;如果该节点未被访问过,发现代理将该节点标志为“visited”,将该节点到达链路状态字标识为“upstream”,转 3;如果代理发现当前节点已经被访问过,那么代理注销;如果 TTL 减到 0,那么代理并变异出报告代理,把搜集的拓扑数据遗留给报告代理,报告代理(Re-

port Agent)将信息传给管理站,然后路由发现代理注销自身。

3. 采用预置的 SNMP 会话读取当前路由器的 ipAddrTable 表,取出 ipAddrEntAddr 的最大值来标志当前节点。

4. 访问当前路由器的路由表,对每个路由表项,如果 ipRouteType 等于 indirect,将路由表中 ipRouteNextHop 的内容不重复地加入邻居列表中,如果 ipRouteType 等于 direct,把 ipRouteDest 和 ipRouteMask 不重复地放到本地相连子网队列中。

5. 如果路由发现代理对当前路由器的路由表的访问未结束,转 4。

6. 路由发现代理创建报告代理,将本地拓扑信息传回,然后从邻居列表中取出邻接节点,设节点个数为 d ,说明当前节点周围有 d 个链路可供迁移选择,如果 d 为 0,说明发现当前节点已经是叶子节点,那么代理并变异出报告代理,把搜集的拓扑数据遗留给报告代理,报告代理携带新发现节点以及相邻节点的 ID 号等拓扑信息返回管理站点,报告代理沿着上游链路移动回到管理站点,交由管理站点存储各节点拓扑信息,并计算是否已经取回所有站点的拓扑信息,而最终得到全网络模型。如果 d 大于等于 1,不考虑自己迁移进来所使用的链路和其他所有标志为“upstream”的链路(设个数为 k),然后发现代理复制 $(d-k-1)$ 个发现代理和 1 个核心节点代理,后将数据区解锁,复制的 $(d-k-1)$ 个发现代理沿当前节点空闲链路迁移,转第 2 步。

7. 管理站节点获取子网所有拓扑信息后,算法结束。

报告代理的行为算法:

1. 被变异出来之后,继承发现代理的数据内容,然后按照标有“upstream”的链路向管理站迁移。

2. 到管理站后,将携带的拓扑信息交付给管理站驻留代理,随即注销自己。

管理站驻留代理(归属代理)的行为算法:

1. 发放代理:以所在节点为起点,根据需要生成路由发现代理,并将其派遣到网络中,发起洪泛。

2. 实现接收报告代理和核心节点代理的消息处理和交互机制。

3. 计算信息提取:来自节点发现代理的消息可以分成两类:网络拓扑信息和控制信息,对消息内容进行计算过滤。管理站驻留代理要能进行信息的区分、提取和计算,存入拓扑信息数据库。

4. 更新数据:与管理站相连的是拓扑信息数据库,是一大型数据库(如 Sybase),用于存储各子网管理者的日报数据以及收集到的拓扑信息,考虑到网络状态的动态变化以及路由发现代理传回的拓扑信息,信息数据的更新操作由管理站代理来进行。

3.2.2 子网发现算法

子网拓扑发现的任务是确定子网的类型,发现子网内活动的主机并获得该子网内主机的信息。这一过程由核心节点发现代理和巡回代理完成。核心节点代理由路由发现代理创建,充当了子网管理站的角色,负责探测和收集与之直连的子网拓扑结构信息。以该路由节点为起点,对直连的子网再进行二级拓扑发现,即子网-主机发现,拓扑发现的粒度更细。子网巡回代理和子网发现代理是为了进一步完成子网内的拓扑探测,它们是一些更轻型的代理,也叫信使(Messenger)。这些信使是由驻留在路由节点上的核心节点代理派遣的,它们

的业务逻辑更简单,仅仅探测子网内的一些活动终端的拓扑信息,并通过代理之间的通讯,完成子网信息的收集。

核心节点代理的行为算法描述如下:

1. 将从发现代理继承下来的子网队列取出一个元素为选定子网。

2. 核心节点代理利用 SNMP 协议的 GetNextRequest 操作取表 ipNetToMediaTable 的一行中 ipNetToMediaNetAddress, ipNetToMediaPhysAddress 以及 ipNetToMediaIflIndex 的相应的值。分析取回的 ipNetToMediaPhysAddress 的值是否为“FF:FF:FF:FF:FF:FF”,如果是,表示相应的 ipNetToMediaNetAddress 是子网地址或者是广播地址,不是主机地址,跳转到 5 执行;如果不是,则认为取回的值是活动主机 IP 地址、物理地址和路由器接口索引,执行 3 进行进一步判断。

3. 将取回的 ipNetToMediaNetAddress 的值与选定子网的掩码进行“与”操作,结果与选定子网地址比较。如果相同,则主机是选定子网内的主机,取回的 ipNetToMediaPhysAddress 和 ipNetToMediaNetAddress 分别是该主机的物理地址和 IP 地址,并将该主机地址加入到子网主机列表中。

4. 分析取回的 ipNetToMediaIflIndex 的值,在 Interfaces 组接口表 iftable 中查找 iflIndex 与 ipNetToMediaIflIndex 对应的表项,获得 iftype 的值,从而确定子网的类型。

5. 使用 GetNextRequest 操作读取 ipNetToMediaTable 表中的下一行的 ipNetToMediaPhysAddress 和 ipNetToMediaNetAddress 以及 ipNetToMediaIflIndex 的值。重复 2-5 直到 ipNetToMediaTable 表访问完毕。然后变异出等于主机数目的子网发现代理和一个子网巡回代理,将核心节点获得的子网主机队列作为子网巡回代理的初始巡回列表。

6. 重复 1,直到子网队列为空。

7. 按照一定的时间间隔,进行 1-6 的操作并周期性向管理站报告所管辖子网的拓扑信息。

子网巡回代理的行为算法描述如下:

1. 巡回代理按照初始巡回列表中记录的地址逐一进行迁移。

2. 每迁移到一处,巡回代理继续资源发现,如果发现的节点不在巡回列表里,则对新节点作好记录,丰富巡回列表,继续迁移。

3. 如果目标节点可迁移,并且是作过记录的新节点,则在被管理节点采集路由信息和地址信息,并按照事先约定的消息格式,通过代理间发送消息的方式将信息发送给核心节点代理;如果向目标节点迁移失败,则要对失败节点进行连通性测试,如果连通性测试成功,说明该节点未启动移动代理执行环境;如果连通性测试失败,说明主机可能宕机,子网巡回代理对此应作必要的记录,将失败节点的关键信息发送给核心节点代理,请求更新。然后向目标节点的下一个节点继续迁移。直到子网内所有活动主机都被访问过为止,然后巡回代理注销自身,注销前给核心节点代理发送死亡消息。

子网巡回代理是为了提高拓扑发现的完备性而设计,子网发现代理则是为了提高子网拓扑发现的效率而设计的。当核心节点代理获得了选定子网内的主机列表时,创建等于主机数目的子网发现代理,同时派至与之相连的各个主机中去,是子网层次上的一次洪泛。子网发现代理的行为非常简单,迁移至目标节点上就开始获取本地拓扑信息,然后将信息发送给核心节点代理,获得应答后注销自身,旨在使获取子网主

机信息的时间达到最优。

3.3 代价分析

网络模型表示为由点 v 和边 E 组成的图 $G = \{V, E\}$, $V = \{V_0, V_1, \dots, V_{n-1}\}$, 表示网络中的节点集合, 节点数为 n , $E = \{e_1, e_2, \dots, e_m\}$, 表示网络节点间的链路集合, 边数为 m 。显然, 顶点 V_0 是移动代理出发的根节点, 即管理站驻留节点。定义节点 v_i 中存储的邻接节点的信息为拓扑信息。设图中节点总数为 n , 路由节点数为 k , 发现代理将在树 T 的链路上迁移一次, 在非树 T 的链路上, 发现代理将到达它的两端节点, 并为该链路派发移动代理, 即: 在非树 T 的链路将有两个发现代理旅行。令 h 是一个分组交换包头的尺寸, A_a 是发现代理的代码尺寸, c_i 是在链路 e_i 上传送一个字节信息的开销。由于子网巡回代理是一个轻型代理, 本身除携带一张路由表不携带任何拓扑信息, 只是对拓扑发现起到查漏补缺的辅助作用, 因此它在网络中的巡回开销较小, 可不作统计。

路由发现代理和子网发现代理在网络中漫游造成的开销为:

$$\begin{aligned} \text{Discovery_cost} &= (h + A_a) \left[\sum_{i=1}^{n-1} C_i + 2 \left(\sum_{i=1}^m C_i - \sum_{i=1}^{n-1} C_i \right) \right] \\ &= (h + A_a) \left(2 \sum_{i=1}^m C_i - \sum_{i=1}^{n-1} C_i \right) \end{aligned}$$

每个节点信息伴随发现代理在网络中传送的开销为:

$$\begin{aligned} \text{Information_cost} &= d_1 c_1 + d_2 (c_1 + c_2) + \dots + d_{n-1} (c_1 + c_2 + \dots + c_{n-1}) \\ &= \sum_{j=1}^{n-1} d_j \left(\sum_{i=1}^j C_i \right) \end{aligned}$$

设: 报告代理的代码尺寸为 A_b , 报告代理的网络开销为:

$$\text{report_cost} = \sum_{j=1}^{k-1} (h + A_b) \left(\sum_{i=1}^j C_i \right)$$

则该算法总的网络开销为:

$$\begin{aligned} \text{Total_cost} &= (h + A_a) \left(2 \sum_{i=1}^m C_i - \sum_{i=1}^{n-1} C_i \right) + \sum_{j=1}^{n-1} d_j \left(\sum_{i=1}^j C_i \right) \\ &\quad + \sum_{j=1}^{k-1} (h + A_b) \left(\sum_{i=1}^j C_i \right) \end{aligned}$$

算法的时间开销等于发现代理到达最远的叶子节点返回报告代理的计算和网络传输的时间。

对该算法的分析评估得出:

优点是算法的运算时间大大减少, 降低了节点拓扑信息在网络上的传输。

缺点是随着网络规模的扩大, 报告代理在网络上传输的开销会显著增加, 因为代理程序码往往比拓扑信息大得多, 网络开销很大。为了降低这种开销, 可以通过将报告代理改良成采用叶子节点的报告代理来实现, 方法是仅仅当发现代理到达叶子节点, 且是被未被任何其他发现代理访问过的叶子节点时, 才派生发送一个报告代理返回管理站点。报告代理沿上游链路返回时, 取出各节点未报告过的节点拓扑信息, 一同回到管理站点。

报告代理的网络开销减少为:

$$\text{report_cost} = (h + A_b) \sum_{i=1}^{k-1} C_i$$

算法总的开销为:

$$\begin{aligned} \text{Total_cost} &= (h + A_a) \left(2 \sum_{i=1}^m C_i - \sum_{i=1}^{n-1} C_i \right) + \sum_{j=1}^{n-1} d_j \left(\sum_{i=1}^j C_i \right) \\ &\quad + (h + A_b) \sum_{i=1}^{k-1} C_i \end{aligned}$$

4 原型系统的设计与实现

4.1 系统设计

该系统总体结构如图 2 所示, 基于移动代理的网络拓扑

发现系统主要由移动代理产生器、拓扑发现模块和拓扑图绘制模块组成。

移动代理产生器模块用来产生系统执行所需要的两种类型的移动代理: 移动代理和静态代理。归属代理属于静态智能代理, 负责代理的生成和发放以及主动轮询被管节点。节点发现代理可以在网络中漫游, 收集网络拓扑信息, 并能复制自身, 在网络中泛洪, 发现新的节点, 采用这种方式不但能减轻管理站负担, 而且通过给代理定制迁移策略, 适时终止代码的执行, 不会占用太多的网络带宽, 可以提高网络拓扑发现的效率。

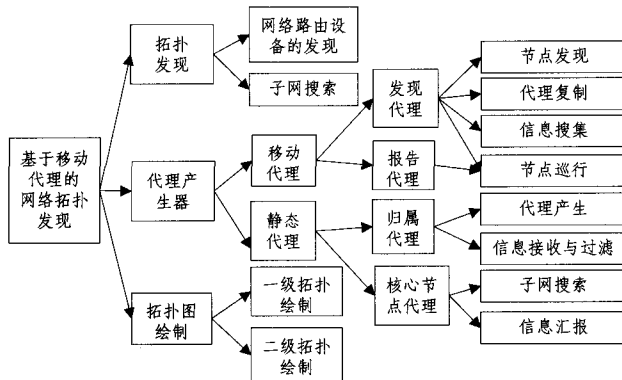


图 2 基于移动代理的网络拓扑发现系统的总体结构图

拓扑发现模块是系统的核心, 是整个系统的引擎, 主要实现网络路由设备的发现和子网的搜索。需要完成的功能指标有:

1. 网络路由设备的发现, 交换机、路由器等设备类型的识别和拓扑信息获取;
2. 子网节点的搜索, 子网内存在哪些活动节点并获取其基本信息;
3. 定期进行拓扑信息的搜集, 并尽可能对拓扑变化的原因进行分析定位。

拓扑图的绘制模块根据收集到的网络拓扑信息绘制网络拓扑图。主要包括一级拓扑图的绘制和二级拓扑图的绘制, 模块功能包括:

1. 用图形化的方式直观形象地表示出网络设备之间的互联关系;
2. 可在故障节点上进行注释或者做标记;
3. 界面友好, 发现正确。

结合上面对各功能模块的分析, 开发此系统需要进行各种代理的设计, 后台拓扑数据库的设计以及拓扑显示模块的设计等。系统的模型示意图如图 3 所示。

4.2 详细设计

4.2.1 采用的工具和环境

经过对系统设计需求的分析, 我们确定了系统的开发工具及平台:

操作系统: Windows2003/NT
 移动代理平台: MAP 2.0
 开发环境: visual studio .net 2003
 开发语言: C#

为了便于构建功能更加完善的网络管理原型系统, 我们增加了新的设备, 构建了较完备的 Ethernet 式的局域网络实验环境。采用了一台配备了双网卡的 PC, 通过设置, 使之成为一台具有路由转发功能的模拟路由器, 从而使我们的实验网络划分为多个(目前 2 个)子网的实验环境, 如图 3 所示。

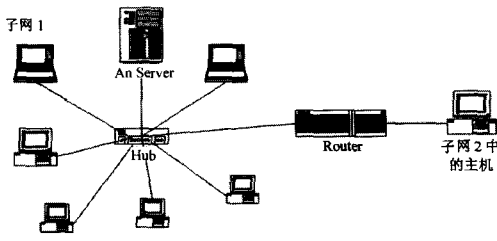


图 3 实验环境示意图

理,核心节点代理和报告代理。路由发现代理和报告代理主要负责一级拓扑发现,核心节点代理及其子代理:子网巡回代理和子网发现代理,负责探测子网的类型和子网内的活动主机。

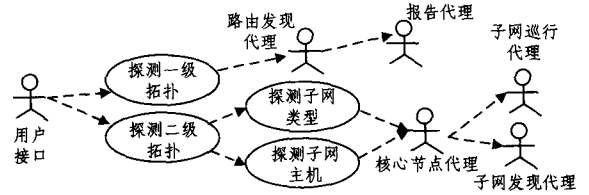


图 4 模块用例图

4.2.2 代理设计

如图 4 所示,系统中的代理主要有归属代理,路由发现代

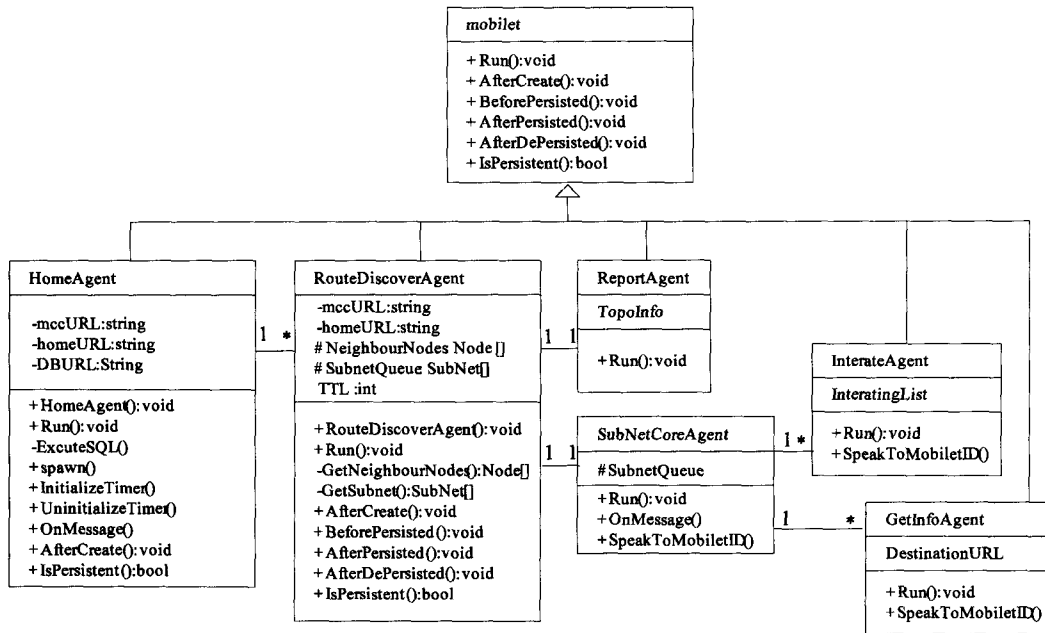


图 5 代理对象类图

根据对各个代理的功能描述,代理之间的对象类图如图 5 所示。

4.2.3 拓扑数据库的设计

采用数据库存放拓扑信息的好处是,可以提供标准的统一接口,与拓扑发现算法独立开来;由于数据库驱动内置了支持多线程访问的处理,因此可以很方便地在拓扑检测的同时读取数据信息来构造拓扑图,实现拓扑检测和拓扑显示的并行实时处理。拓扑数据库中存放的表有路由表和主机表,路由表中的信息包括路由节点标志 IP,直连子网号,相邻路由节点等,主机表主要设有节点 IP 地址,该地址的子网号,主机名称,是否安装 MAP 平台标记,是否运行正常,故障信息描述等字段。

4.2.4 拓扑图的显示模块设计

拓扑图的显示模块是网络管理软件的门户,友好的界面,清晰、准确地表达网络拓扑意义尤其重要。一个合理的网络拓扑图应该是层次型的,即一级拓扑图显示整个网络系统中各子网及网关间的互连结构,二级拓扑图显示所选子网内部各网络设备间的互连关系。在系统中还给每个图标设置了右键菜单,提供了丰富的选项,供用户查阅,例如节点的基本信息,性能和故障信息等。因为网络的拓扑结构极有可能变动,所以拓扑图还要经常刷新。可利用 C# 的多线程技术,设计两个线程,使它们同时运行,一个线程为应用程序的主线程,

负责当前的拓扑显示和事件处理,另一个线程是一个定时器,每隔一段时间后,刷新拓扑图。两个线程交叉运行,完成该功能。我们利用图形开发库 GDI+ 绘出最后的拓扑结构图。

4.3 实验结果

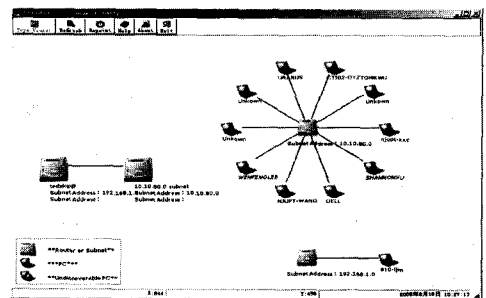


图 6 点击 Topo Viewer 显示的拓扑图

实验结果如图 6 所示。该拓扑图与图 4 所示的实验环境拓扑结构一致,证明了系统拓扑发现的正确性。系统每隔一定时间会自动刷新拓扑图,如果发现失败节点,将在图上做出标记。

结束语 本文提出的基于多移动代理的网络拓扑发现算法以 TCP/IP 网络管理体系为依托,将各种可以利用的协

(下转第 93 页)

最后根据评价得分降序对输出文档排序。

5 试验结果

以通用的搜索引擎 Google、百度(baidu)、雅虎(Yahoo)、搜狐(Sohu)为成员搜索引擎建立了一个中文元搜索原型 MWS,进行了 20 次各类查询主题的实验。由于研究表明用户浏览的平均页面为 2.35,因此本实验中每次查询从各个搜索引擎中选取返回的前 30 条记录作为合成的输入文档。

表 1 搜索引擎的相关性比较(MWS 为基准)

rank	MWS	baidu	google	sohu	yahoo
5	0.7667	0.6667 (-13.0%)	0.8000 (+4.3%)	0.7333 (-4.3%)	0.6667 (-13.0%)
10	0.7633	0.6167 (-19.2%)	0.6833 (-10.5%)	0.6000 (-21.4%)	0.5667 (-25.8%)
15	0.7557	0.6133 (-18.8%)	0.6777 (-10.3%)	0.5083 (-32.7%)	0.5433 (-28.1%)
20	0.7333	0.5757 (-21.5%)	0.6083 (-17.0%)	0.4787 (-34.7%)	0.4833 (-34.1%)
25	0.7267	0.5667 (-22.0%)	0.5500 (-24.3%)	0.4733 (-34.9%)	0.4533 (-37.6%)
30	0.6723	0.5800 (-13.7%)	0.5643 (-16.1%)	0.4847 (-27.9%)	0.4447 (-33.9%)

对搜索结果的评价采用了文献[7]中提出的相关性评价方法:

$$relevancy = \frac{2 * r + u}{2 * n} \quad (9)$$

其中, r 是相关文档, u 是不确定文档, n 为检索到的文档总数。

实验测试了元搜索原型 MWS 的平均相关性,并与各成员搜索引擎进行比较,其实验结果见表 1 所示。

表 1 为元搜索原型 MWS 与成员搜索引擎的搜索结果平均相关性比较。从表中可以看出 MWS 的检索结果相关性比成员搜索引擎均有较大的提高。对于中文搜索而言,百度和 Google 是两个比较好的搜索引擎,针对不同的搜索主题,其性能各有优劣。在本试验的结果中,Google 的相关性略高于百度。而本文的搜索原型与最好的 Google 相比,除前 5 个文档的相关性略有降低外,平均相关性提高了 10%~20%,比最差的成员搜索引擎提高了 10%~35%。

(上接第 80 页)

议和发展技术有机地整合起来,高健壮性地、高完备性地、高效率地、高准确度地快速发现网络拓扑信息。它不同于传统的面向协议的发现算法,是一种面向过程的算法,即,对发现过程分步骤,每一个阶段确定一个执行目标,是一种分级搜索的策略,将网络拓扑发现分成两级进行:一级拓扑发现主要发现路由器(网关,有路由功能的主机)设备和子网;二级拓扑发现主要发现子网内的主机以及子网类型等一些信息。该方法与网络的分层结构相一致,能更好地体现网络的层次。

不过移动代理技术刚兴起不久,发展还不够成熟,想要投入实际使用还有很多问题需要解决,例如管理者怎样动态影响移动代理的生命周期,或代理如何自主预测、避免和减少由于自身移动所可能出现的网络拥塞、系统崩溃、安全侵害等问题,研究更多的用于网络管理的移动代理,同时考虑更好的代

结束语 本文提出了一个基于 Muti-agent 的元搜索引擎系统。采用了多 Agent 技术和移动 Agent 技术,可以减轻网络负载,且对网络连接的要求不高,提高了元搜索引擎的效率,并具有良好的可扩展性。数据库的分类有利于元搜索引擎中的资源选择算法选择与用户查询更相关的数据库,从而提高检索效率。在搜索引擎的调度策略中,在对搜索引擎数据库进行概念类划分的基础之上,采用虚拟语言模型完成数据库的选择。该方法不需要数据库文档的统计信息或对文档进行训练,因此简单易于实现。对于搜索结果的合并,充分利用搜索结果隐含的信息,可有效减小搜索引擎之间的差异,试验结果表明查询效率明显高于 Web 搜索引擎。

参考文献

- [1] Wang W, Meng W, Yu C. Concept hierarchy based text database categorization in a metasearch engine environment//Proceedings of the First International Conference on Web Information Systems Engineering (WISE'00). 2000;283-290
- [2] Ponte J M, Croft W B. A language modeling approach to information retrieval//Proceedings of the ACM Conference on Special Interest Group on Information Retrieval. New York, N. Y., USA: ACM Press, 1998;275-281
- [3] 张卫丰,徐宝文,周晓宇,等.元搜索引擎结果生成技术研究.小型微型计算机系统,2003,24(1):34-37
- [4] Hoon G K, Tan S S, Tang E K, et al. Rank Aggregation model for meta search- an approach using text and rank analysis measures//Proceedings of the International Conference on Intelligent Information Processing (ICIIP 2004). London: Springer-Verlag, 2004;325-339
- [5] Bordogna G. Soft fusion of information accesses//Proceedings of the 2002 IEEE International Conference on Fuzzy Systems. 2002 (2):1466-1471
- [6] Wu S L, McClean S. Result merging methods in distributed information retrieval with overlapping databases. Information Retrieval, 2007, 10(3):297-319
- [7] Keyhanipour A H, Moshiri B, Piroozmand M, et al. WebFusion: fundamentals and principals of a novel meta search engine//Proceedings of the 2006 International Joint Conference on Neural Networks. 2006:4126-4131

理受控算法或代理安全策略,也是我们今后的研究方向。

参考文献

- [1] 杨家海,任宪坤,王沛瑜.网络管理与实现技术[M].北京:清华大学出版社,2000
- [2] 王志刚,王汝传,王绍棣,等.网络拓扑发现算法的研究[J].通信学报,2004,25(8)
- [3] 岑贤道,安常青.网络管理协议与应用开发[M].北京:清华大学出版社,2002
- [4] Dah M C, Ram S, Chiu D M. Network Monitoring Explained: Design And Application [M]. Massachusetts: Prentice Hall, 1992;50-180
- [5] Ramadas S. Spial Edition Using TCP/IP[M].北京:电子工业出版社,2003