

基于多核单向共享树的选播路由算法的研究

张园园, 余镇危

(中国矿业大学(北京), 北京 100083)

摘 要: 为了更好地在应用层实现选播通信, 在对已有选播路由算法 PIA-SM 和 CBT 的优缺点分析的基础上, 提出了一种基于多核单向共享树的选播路由算法, 并给出了核节点和选播组成员的选择策略, 实验表明, 本算法能有效平衡网络和服务负载, 并在网络负载较重的情况下, 有效地降低数据包的传送时延。

关键词: PIA-SM CBT; 多核单向共享树; 时延; 网络负载

中图分类号: TP393

文献标识码: A

文章编号: 1000-7180(2012)06-0007-03

Research on Anycast Routing Algorithm Based on Multi-Cores Uni-Directional Shared Trees

ZHANG Yuan-yuan, YU Zhen-wei

(China University of Mining & Technology, Beijing 100083, China)

Abstract: For the purpose of realizing the anycast communicate more efficiently, the article is proposed an anycast routing algorithm based on Multi-Cores Uni-Directional Shared Trees based on the analysis of advantages and disadvantages about the existing anycast routing algorithm PIA-SM and CBT, and presented the choice strategy of the core nodes and the anycast group members. The experiments show that the algorithm can balance the network and server load, and reduce the transmission delay efficiently under the heavy network load.

Key words: PIA-SM CBT; MCUST; delay; network load

1 引言

选播服务是 IPv6 提供的一种特殊网络服务, 它允许服务申请者访问、共享同一 Anycast 地址所标识的一组接口中最近的一个(这里的最近是按路由协议的距离量度来计算)^[1]。目前对选播服务的研究方向可以分为: 应用层选播和网络层选播。应用层选播需要收集服务器主机是否在线和该主机提供何种服务类型, 以及每个潜在客户和各个不同服务器间的以用户指定的标准度量的距离信息。这需要 Anycast 应用程序不得不频繁地探测服务器的状态, 或者服务器频繁地向 Anycast 应用程序报告本机状态。当网络中的业务十分繁忙并且服务器数量很多时, 网络可能不堪重负而导致整个网络瘫痪。因此本文的选播服务是在网络层上进行研究的。

选播和组播(multicast)的相似之处在于多个成

员节点共享一个服务地址^[2]。这就意味着可以利用与组播相似的方式来管理选播成员。因此, 本文设计的路由算法的思想是利用组播来实现选播成员的管理, 用单播来实现数据包的传送和转发。在现有的研究中, 文献[1]中采用 CBT(Core-Based-Tree)树并规定了路由的次序来实现选播。文献[4]中改进了组播路由协议 PIM-SM, 提出了 PIA-SM 并通过 RP 来实现选播服务。但是缺乏可行的组成员管理机制, 如对于发送源的接纳控制和认证。PIA-SM 虽然可以提供类似的机制, 但此时多播业务流大量集中在 RP 处, 容易导致网络拥塞, 并且频繁的认证很容易使 RP 处理器过载, PIA-SM 还是单点故障协议, RP 的失效将导致整个多播树失效。本文提出一种新路由算法, 通过多点(核)接入与单向传送机制, 有效地解决了选播的发送方管理、定量可靠性和复杂性以及负载均衡问题。

2 协议描述

本文提出了多核单向共享树(Multiple Cores Unidirectional Shared Trees, MCUST)路由算法来完成选播服务. 它的主要设计思想是在网络中选择多个路由器节点作为 RP, 也就是核节点(Core). 在客户端申请选播服务时, 首先根据一定的选择策略选出一个对此客户端最优位置的核节点, 再通过这个核节点为根构造的选播树^[3-5], 得到最优的选播服务器, 如图 1 所示.

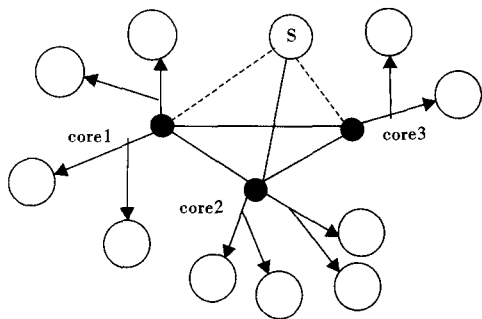


图 1 MCUST 选播树

2.1 核节点的分布与设计

在网络中选取多个核节点比选取一个核节点问题要复杂的多. 核节点的分布原则要受到一定原则的制约, 另一方面要分布均匀, 以便达到负载均衡的目的. 一种方法是给予一个节点到核节点距离的最大值 d , 设法找到满足这个条件的最小核数量, 这是最小 d 域集问题; 另一种方法是给定核节点的数量 k , 设法使得任一个节点到与它最近的核节点的最大距离最小化问题, 这是最小 k 中心选址问题, 这两类问题被称为 NP 难问题. 文献[6]中给出了一种核节点部署方法.

2.2 核节点的选择策略

要加入的选播成员向每个核节点发送加入请求, 因此当用户请求到来时, 首先要将服务请求路由到核节点并在核节点之间进行选择. 对于此用户节点最好的节点, 这里主要考虑的是时延问题, 即通过选择时延最小的核节点, 使最终的选播服务节点的选择能够达到用户要求, 保证服务质量^[7]. 因此, 我们以源节点为根, 以核节点为叶子建立一棵有源树, 通过 Dijkstra 算法可算出从源节点到各个叶子的最短路径. 距离值是反映链路的数据运输及处理能力的一个度量值. 距离一般与带宽、负载、可靠性、开销等

各网络参数中的一项或多项有关, 距离越小, 说明该链路数据运输处理能力越好.

2.3 选播生成树的构造

这里介绍以核节点为根, 选播成员为叶子节点的单向共享树的构造. 首先, 给出以下定义:

定义 1 (路由器次序) 设 $G(A)$ 为一个选播服务器组, CBT 为相应的共享树, 对于在树 CBT 上的两个相邻的路由器 X 与 Y , 记 $X >_{\text{CBT}} Y$, 如果 X 节点是 Y 的父亲节点.

有定义可知, 如果 X 和 Y 是从核节点到选播组成员 $G(A)$ 的最短路径中相邻的两个路由节点, 并且 X 比 Y 要接近核节点, 那么我们记作 $X >_{\text{CBT}} Y$, 这也就定义了单向共享树的构造. 图 2 中给出了一个有核树的构造过程. 其中 R_7 为选定的核节点.

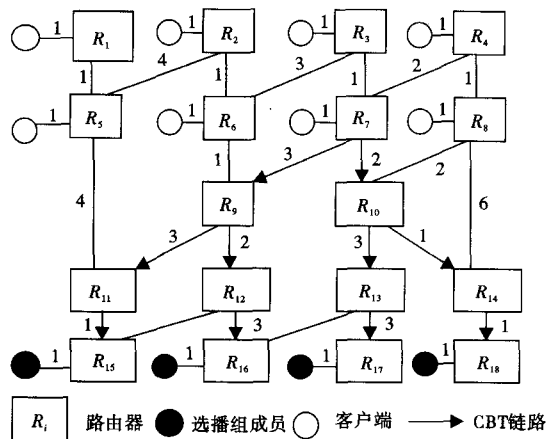


图 2 CBT 单向共享树的生成

2.4 数据包的转发

在选定了核节点后, 就要在以这个核节点为中心的 CBT 单向树中把数据包选择并转发到某个选播成员. 这个过程需要完成三个子任务: (1) 在选播路由表中进行查找. 具体的实现过程为: 从选播数据包头中找到目的地、源地址、数据流标记等相关的信息, 首先在缓存中进行查找, 如果结果中没有再查找相应的路由表. 由于所有选播成员在路由表中是连续存储的, 因此结果会返回一组记录; (2) 根据步骤 (1) 中的信息, 利用文献[2]中的随机权重法随机选择其中一个 $G(A)$ 与客户端进行通信; (3) 发送数据包. 当发送数据流的第一个数据包时, 将其路由记录通过数据流标记复制到缓存中, 以便后续数据包的传送.

3 模拟与分析

3.1 算法的可靠性分析

从理论上讲,假设网络中每个节点的失效概率为 p ,且互相独立,服从 Bernoulli 分布,MCUST 选播树有 C 个核节点,则单核树 PIA-SM 和单向 CBT 树的可靠性为: $K=1-p$;MCUST 树的可靠性为 $K=1-P^c$,即当 $P=0.1$,单核树的核节点容易发生瓶颈,它的可靠性为 $K=0.9$,MCUST 树的可靠性为 $K=1-0.1^c$,由于 C 是整数,所以本文提出的 MCUTST 算法的可靠性要有明显的提高。

3.2 实验仿真分析

这里通过实验对本文的算法与 PIA-SM 和 CBT 有向单核树进行性能比较和分析,采用 NS2 工具进行仿真,在 C++ 和 Tcl 层次上进行扩展,编写一个 treecalc 的模拟器进行实验^[8]。利用 GT-ITM 网络生成工具根据 Waxman 模型随机生成节点为 100 的拓扑图,每个节点的度设置为 0.3(30% 的叶子节点),带宽为 64 kb/s,10 个随机分布的选播组成员 $G(A)$ 节点,4 个发送源节点,发送速率为 16 kb/s,假设网络中仅有选播会话,并且核节点的放置选取随机放置(Thale 证明随着选播组成员的动态变化,最佳节点的位置放置最终会变为随机放置)。我们从核节点数与时延、负载的关系来进行对比实验,并对结果进行了分析。

(1) 核节点数与时延

当核节点 $core=1$ 时,也就是文献[1]和文献[2]中所提出的单核 PIA-SM 和单向 CBT 树的 anycast 包的传输时延,随着核节点数目的增加,时延比率大概降低了 15.8%(图 3)。因为发送源有更大的几率选择与它位置最近或者通过最短路径到达的那个核节点作为单向共享树的根进行选播路由。

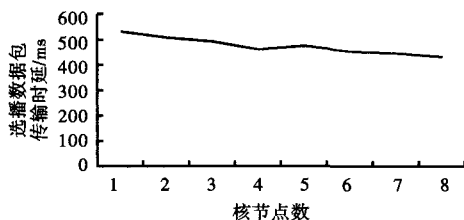


图3 核节点数与时延

(2) 网络负载

这里验证随着网络负载的不断变化,三种算法的链路平均时延对比。通过不断改变节点的平均请求频率的方式来改变网络负载,并计算 Anycast 路由中各条链路的平均总时延。由图 4 可见,在网络

负载不太大时,PIM-SM 的时延是最好的,CBT 其次。这是因为 PIM-SM 算法的 RP 节点提供了很好的成员管理机制,可以使发送端的请求能够传送的相对最优的选播成员,因而有较小的平均时延与开销。但是随着网络负载的不断增加,单核树中所有业务都要到达核节点,增加了传输的延时,这是数据传输的瓶颈。本算法在随着网络负载的增加的过程中体现了很好的可靠性和减少了平均时延。

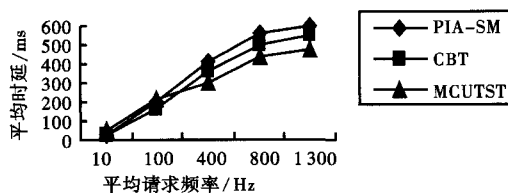


图4 三种算法时延与网络负载的对比

4 结束语

本文提出了基于多核单向共享树的选播路由协议,利用最短路径树来选择离发送源最近的核节点,并根据以它为根的单向共享树的基于随机权重的路由算法来得到最优的选播组成员。定量的分析了可靠性,从仿真实验结果可以得出,相对于单核结构,本算法降低了网络时延,平衡了网络负载,尤其是在大量发送源的情况下,本算法的优势相当明显。

参考文献:

- [1] 丘小兰,李陶深,葛志辉. 无线 Mesh 网络多网关选播 QoS 路由模型研究[J]. 微电子学与计算机, 2010, 27(8): 116-117.
- [2] Doi S, Ata S, Kitamura H, et al. IPv6 anycast for simple and effective communications[J]. IEEE Communications Magazine, 2004, 42(5): 163-171.
- [3] Xuan D, Jia W J, Zhao W, et al. A routing protocol for anycast messages [J]. IEEE Transactions on Parallel and Distributed Systems, 2000, 11(6): 571-588.
- [4] Matsunaga S, Ata S, Kitamura H, et al. Design and implementation of IPv6 anycast routing protocol: PIA-SM[C]// Proc. AINA. Tai Wan, 2005: 839-884.
- [5] 马钰璐,王重钢,程时端. 多核单向共享树多播路由协议[J]. 计算机学报, 2001, 24(7): 711-713.
- [6] Dykes S G, Jeffery C L, Robbins K A. An empirical evaluation of client-side server selection algorithms[C]// IEEE INFOCOM. Israel, 2000.
- [7] 周治国. 任播技术的研究[D]. 吉林: 吉林大学, 2008.
- [8] 李永胜,李熹. 基于自适应节点选择蚁群算法的 QoS 选播路由算法[J]. 微电子学与计算机, 2011, 28(8): 113-115.

(下转第 13 页)

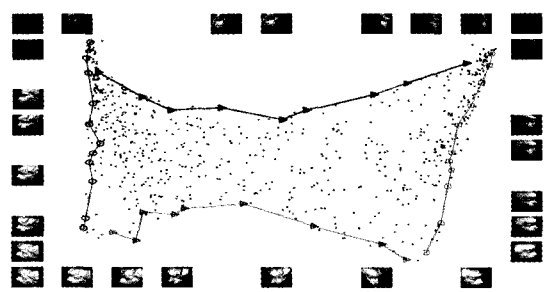


图 3 映射到 2 维空间的 Isomap 人脸图像

参考文献:

[1] Roweis S T, Saul L K. An introduction to locally linear embedding[EB/OL]. [2006-07-06]. <http://isomap.stanford.edu/data/sets.html>.

[2] 高小方. 流形学习方法中的若干问题分析[J]. 计算机科学, 2009, 36(4): 25-29.

[3] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding[J]. Science, 2000, 290(5500): 2323-2326.

[4] Tenenbaum J B, Silva V, Langford J C. A global geometric framework for nonlinear dimensionality reduc-

tion[J]. Science, 2000, 290(5500): 2319-2323.

[5] Zhang Z Y, Zha H Y. Principal manifolds and nonlinear dimensionality reduction via local tangent space alignment[J]. SIAM Journal of Scientific Computing, 2004, 26(1): 313-338.

[6] Belkin M, Niyogi P. Laplacian eigenmaps and spectral techniques for embedding and clustering[J]. Neural Information Processing Systems, 2002(14): 585-591.

[7] Mordohai P, Medioni G. Dimensionality estimation, manifold learning and function approximation using tensor voting[J]. Journal of Machine Learning Research, 2010, 10(1): 411-450.

作者简介:



黄淑萍 女,(1977—),博士研究生.研究方向为数据降维、数据挖掘以及国际投资与金融等.

(上接第 9 页)

作者简介:



张园园 女,(1983—),博士研究生.研究方向为网络体系结构、路由协议、选播.

余镇危 男,(1942—),博士生导师.研究方向为网络体系结构、覆盖网、下一代互联网.