内容寻址网络中路径缓存定向多播路由算法

张伟哲，张宏莉，许 笑，吴太康

（哈尔滨工业大学 计算机科学与技术学院，哈尔滨 150001，zwz@pact518.hit.edu.cn）

摘 要： 为解决内容寻址网络中资源定位速度和路由效率问题，提出了一种基于路径缓存技术的定向多播路由算法.该算法结合定向路由与广播路由的优势，引入扩展系数对定向多播路由算法进行空间维度扩展，降低了集体失效概率.将路径缓存技术与定向多播路由算法相结合，提高了系统的定位效率.通过与传统的定向路由策略进行实验比对，验证了该算法的有效性.

关键词： 对等网络；内容寻址网络；路由算法；定向多播；路径缓存

中图分类号： TP393 **文献标志码：** A **文章编号：** 0367－6234(2010)11－1762－05

A directional multicast routing algorithm based on path redundancy in content addressable network

ZHANG Wei-zhe, ZHANG Hong-li, XU Xiao, WU Tai-kang

(School of Computer Science and Engineering, Harbin Institute of Technology, Harbin 150001, China, zwz@pact518.hit.edu.cn)

Abstract: To improve the resource positioning rate and routing efficiency in the content addressable network, a directional multicast routing algorithm based on path redundancy is put forward. Integrating the advantages of the directional and broadcasting routing algorithms, the new algorithm employs the extending coefficient to enlarge the dimensions of the CAN logical space. Meanwhile, the positioning efficiency of the new algorithm is increased by combining the directional multicast routing algorithm with the probability path redundancy. The effectiveness of the algorithm is proved in the simulation.

Key words: peer to peer network; content addressable network; routing algorithm; directional multicast; path redundancy

内容寻址网络(Content Addressable Network，CAN)是基于多维空间结构的P2P网络[1]，利用分布式哈希表将数据和结点映射为键值，完成多维笛卡尔空间中数据存储与查询.与基于带弦环结构Chord网络、基于异或距离度量的Kademlia和机遇跳表的SkipNet等[2－4]结构相比较，基于多维空间的CAN网络可以结合网络测量信息和地域信息，有助于解决P2P覆盖网络的拓扑不匹配问题[5].内容寻址网络上的路由策略是当前的研究热点.高效的路由机制可以快速地进行资源定位，降低资源加入和退出的带宽消耗.传统的路由方法主要包括：1)广播.当结点接收到查询消息的时候，将其转发到路由表中的每一项.这种方法缺点是网络中转发了大量无用消息，消耗了网络带宽，给系统引入了巨大的负担[6－7].2)定向路由.根据网络延迟或邻居位置来选择一个转发目的地，通过贪心的方法逐渐接近目的结点[8－9].此种方法虽然效果不错，但当被选中路径上出现结点失效时，必须通过回退的方法重新探测一条新的路径.回退再探测的过程非常消耗时间，难以做到快速路由响应和定位.此外，Wu等[10－11]提出通过增加路由表项来提高路由效率：每个结点在路由表中除包含自己的邻居外，还包含邻居的邻

收稿日期： 2009－07－28.

基金项目： 国家自然科学基金资助项目(60703014)；国家重点基础研究发展规划资助项目(G2005CB321806，2007CB11100)；国家高技术研究发展资助项目(2009AA01Z437)；高等学校博士学科点专项科研基金资助项目(20070213044)；中国博士后科学基金经费资助项目(20070410263).

作者简介： 张伟哲(1976—)，男，博士，副教授；
张宏莉(1973—)，女，教授，博士生导师.

居，每个结点负责维护的路由表项数急剧增加；而且当网络中结点状态变化频繁时，结点更新各自路由表系统开销增大.

本文结合定向路由与广播路由的优势，阐述了内容寻址网络中定向多播路由算法的基本原理. 考虑内容寻址网络的动态性与失效问题，引入扩展系数对定向多播路由算法进行空间维度扩展，降低集体失效的概率并避免路由失效发生. 此外，为提高系统的可靠性和查询效率，将路径缓存技术与定向多播路由算法结合，进一步保证系统容错性并提升寻址效率.

1　定向多播路由策略

定向多播路由策略建立在内容寻址网络之上. 其路由方法以多维空间逻辑结构为基础. 假设多维空间的维度为 d，那么逻辑空间中每个结点的路由表的项数至少有 $2d$ 项（空间边界结点除外）. 在 d 维逻辑空间中，两个结点的位置关系在 d 维的坐标上存在偏序关系. 因此，在转发消息的时候，如果目标在第 i 维度上的坐标值大于当前结点坐标的第 i 维，则向第 i 维正向邻居转发消息即可，不需要向负方向的邻居发送消息. 图 1 是在二维逻辑空间下定向多播路由方法示意图. 图 1 中的路由起始点为结点 3，路由的目标结点为结点 14. 箭头为路由消息的流向. 由图 1 可见，路由消息将流经结点3和结点14两个结点之间的所有结点. 形成一个矩形路由区域. 由于查询者和查询目标都是随机的分布在多维空间中，因此查询结点和目标结点的距离统计平均值应该接近于逻辑空间对角线长度的 1/2. 而查询结点和目标结点之间包含的区域统计均值将接近 $\sqrt[d]{Size}$（其中，d 为空间维度，$Size$ 为逻辑空间整体的大小）. 假设空间中有 N 个结点，那么查询结点和目标结点之间包含的结点数统计平均值为 $\sqrt[d]{N}$. 所以，定向多播路由给系统带来的消息负载应该是 $O(\sqrt[d]{N})$ 级别. 相对于广播模式的 $O(N)$ 有了明显的性能提升. 当结点数量巨大（N 值大）并且逻辑空间维度高（d 值大）的情况下，定向多播路由方法相比广播模式降低了系统开销.

相对于定向路由的方法，定向多播路由方法对于结点失效具备更强的容错性，能够容忍路由过程中结点失效的情况. 假设图 1 中结点 12，结点 9 均失效. 定向路由算出来的最佳路径为 3—8—12—13—14. 当按照定向路由的方法转发消息时，结点 8 转发消息到结点 12. 遇到结点 12 失效的时候，结点 8 将进行重新探测，并将消息由结点 8 转发到结点 9. 然而结点 9 也失效，因此将进行路由回退重新探测的过程. 路由路径回退到结点3，而结点3重新选择将路由消息发给结点5. 结点 5 收到消息后，计算的下一跳路由又为失效结点9，因此当结点 5 发现转发向结点 9 不可行时将再次选择将消息转发给结点 6. 最后通过路径 3—5—6—10—14 将消息送达目的. 上述过程中，需要探测等待 3 次，路径回退 1 次. 这 4 个过程将会降低系统的响应速度. 而按照定向多播的方式当结点 3 要寻找结点 14 时，由于结点 14 在 x，y 坐标上的值都大于结点 13，因此结点 13 会向 x，y 正方向上的邻居（结点 5 和结点 8）发送路由消息. 后面每个结点都通过这种坐标比较的方法来确定发送方向，并把消息转发到确定方向上的所有邻居. 形成消息的流如图 1 中箭头所示的情况. 当结点 9 和结点 12 失效的时候，仅打断了其中两条路径，而路由消息会通过路径 3—5—6—10—14 到达目的. 因此结点的失效并没有影响路由查询的效率和结果.

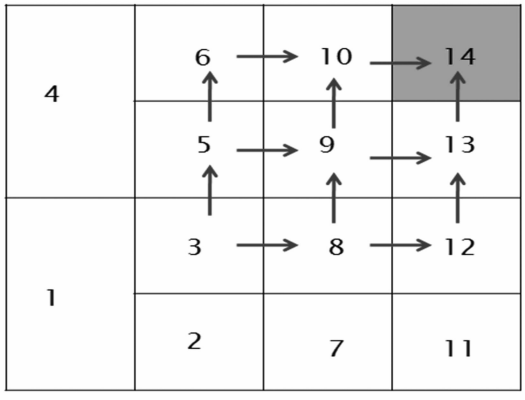

图 1　定向多播路由消息流向示意图

然而，在二维的情况下如果结点10和结点13都失效，那么定向多播路由方法同样不可行. 对此，通过增加维度 d 来提升系统可靠性. 在路由过程中比较坐标大小时，如果当前结点在第 i 维坐标已经大于目标结点，仍然向该方向转发路由消息，但需要将这样的消息进行标识. 最多允许超过坐标区域转发 k 次，k 值可以根据网络的动态情况来设定. 所形成的路由消息覆盖区域将包含原来的定向多播路由方式所覆盖的区域. 除非目标结点所有的邻居都已经失效，否则路由消息一定能到达目的区域. 图 2 中短箭头为原始的定向多播路由消息的流向图，长箭头是在 $k = 1$ 的情况下，扩展定向多播路由方法下路由消息的流向图（数据流将覆盖图 2 中所示的灰色区域）. 通过图 2 可以看出原始方法的路径流向区域包含在扩展定向

多播路由方法所覆盖的区域中. 因此,如果出现访问结点 14 失效的情况那么只能是结点 14 的邻居均已经失效. 这种情况下,结点 14 无法访问到的情况是不可避免的.

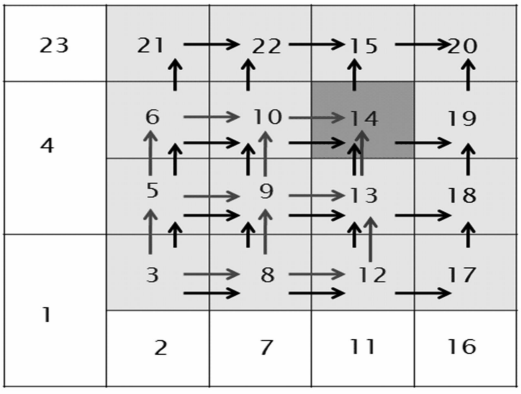

图 2 扩展定向多播路由消息流向示意图

扩展的定向多播路由算法:

输入:查询目标坐标

输出:下一跳结点列表

全局参数: 扩展系数 k; 维度 d;

```
Begin
  List nodelist;
  For (i = 0;i < d;i ++) {
  if(目标结点第 i 维坐标大于当前区域第 i 维坐标最大值)
     nodelist. add (第 i 维正方向所有邻居);
  else if(目标结点 第 i 维 坐标小于当前区域第 i 维坐标最小值)
       nodelist. add (第 i 维反方向上所有邻居);
       else{
         if (k > 0) {
            if (目标结点第 i 维坐标大于消息源结点第 i 维坐标值)
     nodelist. add (第 i 维正方向所有邻居);
       else if (目标结点第 i 维坐标小于消息源结点第 i 维坐标值)
       nodelist. add (第 i 维反方向上所有邻居);
         k --;}}}
     Return nodelist;
End
```

上述扩展定向多播路由算法中,设定扩展系数 $k = 0$ 则算法描述就是定向多播路由算法.

2 定向多播和路径冗余的结合

数据的冗余将增加数据在网络中的副本. 有助于提高数据的查询效率. 而在传统的冗余方法中,冗余的路径单一,冗余具有明显的偏向性——即对热点数据的冗余备份数较多,而对于较少访问的数据备份数量较少. 为了解决这个问题,将路径冗余方法和本文的定向多播路由方法相结合起来,形成了等概率路径缓存定向多播算法. 规定为:

1) 数据加入网络时,加入消息的转发路径按照定向多播路由策略提出的定向多播方式进行转发;

2) 数据查询消息的路由转发过程,也按照定向多播的方式进行转发;

3) 数据在加入网络的过程中,按照概率 p 进行复制备份.

按照上述规则进行数据备份,冗余数据在数据加入起始结点和数据加入目的结点这两个结点坐标所包围的一个 d 维矩形区域内都存在冗余数据. 如图3 所示,数据资源从结点5 开始,加入到网络中,并最终将数据索引存放到结点 20 处. 按照定向多播路由方法, 消息将流经结点 5—6—10—13—14—15—18—22 所围区域. 在数据经过上述区域中的每一个结点的时候,按照一定的概率 p 确定是否要进行数据冗余备份. 最后,在此区域中将会散布着结点 5 处加入的数据的副本(图2 中的结点5,结点19,结点20,结点21 所形成的方块).

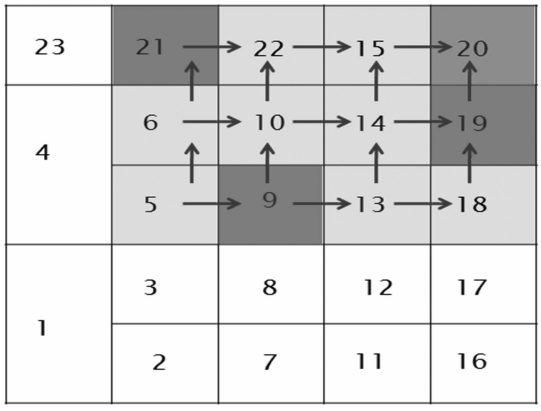

图 3 定向多播方式下路径冗余方法效果示意图

当另外的结点要访问结点 20 上的数据时,按照规则 2,查询消息也按照定向路由的方式进行发送查询消息,那么查询消息过程如图 4 所示.

结点 8 发出查询结点 20 上数据的查询消息. 结点 8—10—12—13—14—15—17—18—22 所围成区域为查询消息按照定向多播方法转发消息时覆盖的逻辑区域. 箭头为查询消息的流动情况. 由于结点 20 上的数据在加入过程中,按照路径冗余

方法在结点 9 和结点 18 上保存了数据副本. 因此当查询消息由结点 8 发送出来时，仅需要通过一次消息发送就可以到达结点 9，从而获取到结点 20保存的数据. 而正常通过结点 8—12—13—14—15—20 这条查询路径所需的路由跳数为 5 次，是具备路径冗余机制的系统中数据查询路由跳数的 5 倍.

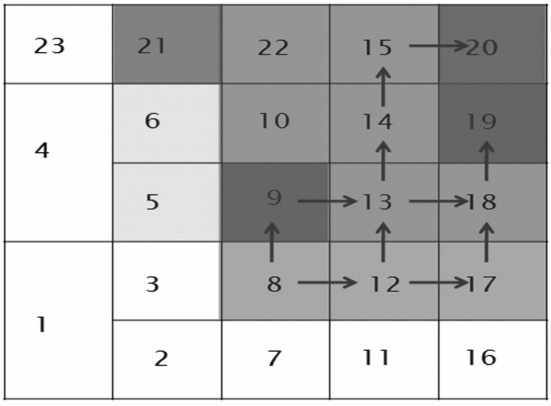

图 4　数据查询示意图

3　实验与结果分析

以 PlanetSim[12] 作为仿真实验平台，首先测试了支持路径缓存的定向多播路由策略与 Sylvia 等[13] 提出的定向路由方法间的性能差异，然后通过查询过程中路由跳数统计体现路由策略定位效率的高低.

3.1　定向多播路由与传统方法性能比较

仿真实验基于 PlanetSim 3.0，加入系统的结点数为 100 结点，向系统中添加的数据资源对象为 2 000 个资源，CAN 逻辑空间设置为 2 维，逻辑空间的区域范围为 $0 \sim 2^{20}$.

图 5 中 0% 重复率为采用传统的定向路由方法（无多播无冗余）；20%、50%、70% 和 90% 重复率分别为在资源加入路径上以概率 0.3、0.5、0.7 和 0.9 进行资源冗余（定向多播概率冗余）.

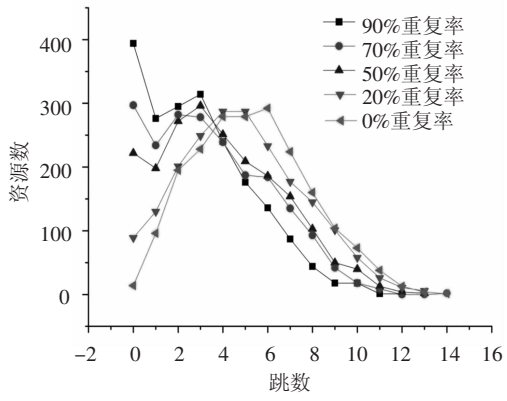

图 5　定向多播路由与传统方法查询开销关系对比图

随着资源冗余的程度加强，寻找资源所需的跳数减少，0，1，2 跳就能到达的资源数量随着冗余的概率增长迅速，而通过多跳才能到达的资源的数量相应的减少. 在资源不冗余的情况下，寻找资源所需的跳数主要集中在 3 ~ 8 跳区域. 以 0.5 的概率进行冗余时，寻找资源所需的跳数集中在 0 ~ 6 跳区域，而以 0.9 的概率进行冗余时，寻找资源所需的跳数集中在 0 ~ 4 跳区域. 说明了数据冗余的方法相比于传统的路由定位方法对于系统中的资源查询效率有很大的提高. 试验证明系统中存储多个数据备份，可以用来提高查询效率，降低查询开销. 通过与传统的无冗余方法进行比较，可以得出本文提出的定向多播和路径冗余相结合的方法在定位效率和查询负载上比其他的传统方法有更优秀的表现.

3.2　资源查询的路径长度统计

系统查询开销所需要的路由跳数可以体现系统的开销，也直接体现了定位效率的高低. 内容寻址网络中每个结点包含了 $2 \times d$ 个方向的邻居表. 从一个结点转发消息到目的过程中，转发路径的下一跳最多有 d 个选择，因此，路由跳数应该是 $O(\sqrt[d]{N})$. 其中，N 为结点数，d 为维度. 试验中采用了 2 维空间，加入的结点数为 100，因此维度 $d = 2$，结点数 $N = 100$，资源查询过程中的查询消息的转发次数应该在 $\sqrt[2]{100} = 10$ 跳左右. 试验中随机选择结点访问每一个资源，最终统计随机访问所有资源的过程中查询消息被的转发次数. 通过图 6 可以看出，绝大部分资源的访问消息转发跳数集中在 2 ~ 8 跳，具有较好的分布情况，符合理论预期值.

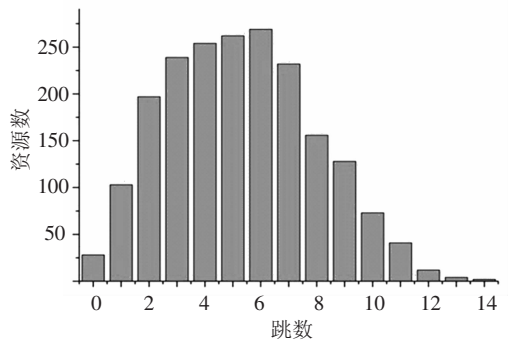

图 6　资源查询路由跳数统计图

4　结　论

1）提出基于 CAN 模型定向多播路由方法改善了传统的 P2P 系统中路由效率低下，路由开销

大的问题.

2）提出定向多播路由和路径冗余相结合的方法,明显提高了系统的查询效率.

参考文献:

[1] RATHASAMY S, FRANCIS P, HANDLEY M. A scalable content-addressable network[C]//Proceedings of the 2001 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication. New York, NY: ACM, 2001:161 – 172.

[2] STOICA I, MORRIS R, KARGER D, *et al.* Chord: A scalable peer-to-peer lookup service for internet applications[C]//Proceedings of the 2001 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications. New York, NY: ACM, 2001: 149 – 160.

[3] MAYMOUNKOV P, MAZIERES D. Kademlia: A peer-to-peer information system based on the XOR metric [C]//Revised Papers from the First International Workshop on Peer-to-Peer Systems. London, UK: Springer-Verlag, 2002: 53 – 65.

[4] NICHOLAS J A, HARVEY N, JONES M B, *et al.* SkipNet: A scalable overlay network with practical locality properties [C]//Proceedings of the 4th Conference on USENIX Symposium on Internet Technologies and Systems. Berkeley, CA: USENIX Association, 2003: 29 – 38.

[5] REN S S, GUO L, JIANG S, *et al.* SAT-match: A self-adaptive topology matching method to achieve low lookup latency in structured P2P overlay networks [C]//Proceedings of the 18th International Parallel and Distributed Processing Symposium. New York: IEEE Press, 2004: 83 – 91.

[6] RATNASAMY S, STOICA I, SHENKER S. Routing algorithms for DHTs: Some open questions [C]//Revised Papers from the First International Workshop on Peer-to-Peer Systems. London, UK: Springer-Verlag, 2002: 45 – 52.

[7] ROWSTRON A, DRUSCHEL P. Pastry: Scalable, distributed object location and routing for large scale peer-to-peer systems [C]//Proceedings of the 18th IFIP/ACM International Conference on Distributed System Platforms. New York: IEEE, 2001: 329 – 350.

[8] ZHAO B Y, KUBIATOWICZ J D, JOSE A D. Tapestry: An Infrastructure for Fault-tolerant Wide-area Location and Routing[D]. Berkeley, CA: University of California at Berkeley, 2001.

[9] WU Z D, RAO W X, MA F Y. Efficient topology-aware routing in peer-to-peer network [C]//Proceedings of the GCC2002. New York: IEEE, 2002: 172 – 185.

[10]陈贵海,李振华. 对等网络:结构、应用与设计[M]. 北京:清华大学出版社. 2007:159 – 165.

[11]齐庆虎, 李津生, 洪佩琳, 等. 内容寻址网络中内容的有效定位[J]. 电路与系统学报, 2004, 9(5): 67 – 71.

[12]Jordi Pujol Ahulló, Marc Sánchez Artigas, Pedro García López. PlanetSim User and developer tutorial [EB/OL]. [2008 – 10 – 20]. http://ants. etse. urv. es/planetsim.

[13]RATHASAMY S, FRANCIS P, HANDLEY M, *et al.* A scalable content-addressable network [C]//Proceedings of the 2001 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications. New York, NY: ACM, 2001: 161 – 172.

（编辑 张 红）