

BuNet: 一种基于 Cayley 图的覆盖网络^{*}

李德苑¹, 魏文红², 王高才³

(1. 华南理工大学 计算机科学与工程学院, 广州 510640; 2. 东莞理工学院 计算机学院, 广东 东莞 523808; 3. 广西大学 计算机与电子信息学院, 南宁 530004)

摘要: 为了构建一个具有路由表小、查询路径长度短、鲁棒性强且支持文件组浏览服务的 P2P 覆盖网络, 通过对扩展蝶网理论的分析与设计, 提出了一个新的基于 Cayley 图的结构化 P2P 网络 BuNet。通过模拟实验, 证明了 BuNet 相比其他结构化 P2P 网络在查询路径短和鲁棒性等方面具有更好的性能。

关键词: 对等网络; 结构化; 蝶网; 凯勒图; 组浏览服务

中图分类号: TP393.02

文献标志码: A

文章编号: 1001-3695(2010)02-0638-04

doi:10.3969/j.issn.1001-3695.2010.02.065

BuNet: overlay network based on Cayley graph

LI De-yuan¹, WEI Wen-hong², WANG Gao-cai³

(1. School of Computer Science & Engineering, South China University of Technology, Guangzhou 510640, China; 2. School of Computer, Dongguan University of Technology, Dongguan Guangdong 523808, China; 3. School of Computer & Electronic Information, Guangxi University, Nanning 530004, China)

Abstract: This paper first introduced a extend butterfly method to design and analyzed a novel structured P2P network named BuNet based Cayley graph. BuNet had many excellent properties such as small routing table, short query path and high robustness, and it proposed file group browsing service. Finally, it discussed and evaluated the performance compared to other DHT-based schemes.

Key words: P2P; structured; butterfly; Cayley graph; group browsing service

近年来, 一些 P2P (peer-to-peer) 系统相继被提出来, 如 Chord^[1]、CAN^[2] 等, 它们为大规模 P2P 应用提供了自组织的基础设施。这类结构化的 P2P 系统是以分布式散列表 DHT 为基础的。它们使用一致性哈希函数将数据和节点都均匀地映射到一个键值空间。然而, 每个节点都被随机映射到一个节点标号, 这样的映射过程丢失了很多节点和资源的性质, 同时网络缺乏对用户使用 P2P 网络方式的考虑, 无法提供文件组浏览 (group browser) 功能, 因而无法更加有效地平衡网络负载。

为了解决上述问题, 笔者引入了基于 Cayley 图^[3,4] 的 P2P 结构化覆盖网络, 而基于 Cayley 网络模型可以有效地解决网络负载均衡问题^[5]。Cayley 图是使用代数群论建立的一类图, 它的最大优势是其对称性^[5], 其顶点具有传递性, 即图中的每个顶点的地位相同。这使得对图中顶点间路由的研究可以转换为对任意一个顶点到某个特殊顶点路由的研究^[5]。因此可以使用代数方法来分析和设计基于 Cayley 图的 P2P 网络路由算法; 同时, 可以采用统一的方法来分析真实 P2P 网络中各个对等点的负载情况。在基于 Cayley 图的 P2P 网络中, 负载都被均衡地分配到各个对等点中。

本文设计了一种新颖的基于 Cayley 图的 P2P 结构化覆盖网络——BuNet (butterfly network)。该拓扑图的节点度为 $O(\log n)$, 图直径为 $O(\log n / \log \log n)$ 。BuNet 可以支持文件浏览以及基于兴趣的分组等模仿人类社会行为, 具有更高的可用性。模拟仿真实验表明, BuNet 还能达到文献^[6] 所提到的

DHTs 最佳网络性能以及更优的容错性。

1 静态蝶网协议

BuNet 的静态拓扑结构是以蝶网 (butterfly) 结构^[7] 为基础的, 静态蝶网结构具有节点度小、直径小的性质。

1.1 Butterfly 的定义

k 维有向 butterfly 图可以表示为 B_2^k , 总共有 $n = k \times 2^k$ 个节点。其定义如下^[5]: B_2^k 是一个有向图, 顶点 $V = V(B_2^k) = Z_2^k \times Z_k$; 而从 $a = (v_0, v_1, \dots, v_{k-1}; i) \in V$ 到 $b = (w_0, w_1, \dots, w_{k-1}; j) \in V$ 有边当且仅当 $i \in \{0, 1, \dots, k-1\}$, $j = i+1$ 且 $v_l = w_l$ 。其中: $l \in \{0, 1, \dots, k-1\} - \{i\}$ 。而 k 维无向 butterfly 图可以表示为 B_2^k , 其定义和 B_2^k 类似; 不同的是连边方式, 即在 B_2^k 中两个连着有向边的点, 在 B_2^k 中用无向边连接起来。 $k \geq 3$ 时, B_2^k 是 4 正则图。

无向图 B_2^k 是 Cayley 图^[3,4], 其定义为 $\text{Cay}(G, S)$ 。其中: 群 $G = (Z_2^k \times Z_k, \cdot)$, G 上的操作 $\cdot$ 为 $\forall (v, i), (w, j) \in V, (v, i) \cdot (w, j) = (v_0 \oplus w_{-1}, v_1 \oplus w_{-l+i}, \dots, v_{d-1} \oplus w_{-l+d-1}; i \oplus j)$, $\oplus$ 是取模加法。显然 G 的单位元是 $(0^k; 0)$, $S = \{(1, 0), (1, (1, 0, \dots, 00 \dots 0))\}$, 其相关数学细节可参见文献^[8]。

1.2 BuNet 的静态拓扑定义

根据 butterfly 无向图的定义, 推广其顶点集为 $V = Z_2^k \times Z_k$, 则得到 BuNet 的静态拓扑定义为

收稿日期: 2009-05-18; **修回日期:** 2009-06-29 **基金项目:** 国家自然科学基金资助项目 (60763013); 广东省科技计划资助项目 (2006B11301001); 广东省工业科技攻关计划项目 (2006B80407001)

作者简介: 李德苑 (1985-), 男, 广东汕头人, 硕士研究生, 主要研究方向为 P2P 拓扑结构与并行算法 (deyuan.li@mail.scut.edu.cn); 魏文红 (1977-), 男, 讲师, 博士, 主要研究方向为 P2P 拓扑结构与并行算法; 王高才 (1976-), 男, 教授, 博士, 主要研究方向为计算机网络、容错计算等。

$$\begin{aligned} V &= \{ (v; i) \mid (v; i) \in Z_r^k \times Z_k, \text{ where } v = (v_0, v_1, \dots, v_{k-1}), \\ &\quad v_i \in Z_r, i \in Z_k \}; \\ E &= \{ \langle (v; i), (v; i \oplus r) \rangle \mid r \in \{1, \dots, k-1\} \} \cup \\ &\quad \{ \langle (v; i), (w; i \oplus 1) \rangle, \langle (v; i), (w; i \oplus (r-1)) \rangle \} \\ w &= (v_0, v_1, \dots, w_i, \dots, v_{k-1}), w_i \in \{0, 1, \dots, r-1\} \end{aligned}$$

BuNet 的静态拓扑就是在 butterfly 的基础上增加进制的同时,增加了不同层间的完全图连边。可以证明 BuNet(k, r) 是 Cayley 图^[3,4],其定义为 $\text{Cay}(G, S)$ 。其中:群 $G = (Z_r^k \times Z_k, \cdot)$, G 上的操作 $\cdot$ 为 $\forall (v; i), (w; j) \in V, (v; i) \cdot (w; j) = (v \oplus \sigma^i(w); i+j)$ 。其中: σ 是循环右移操作, $\oplus$ 是模 r 加法, $+$ 是模 k 加法。显然群 G 的单位元是 $(0^k; 0)$, $S = S_k \cup S_r$, where $S_k = \{ (0^k, i) \mid i \in Z_k \setminus \{0\} \}$, $S_r = \{ (c0^{k-1}, 1) \mid c \in Z_k \}$ 。因为 BuNet 有两种生成集 S_r 和 S_k ,因此边集可以分成两种类型,即由生成集 S_k 构成的层与层之间的连接 k -link 和由生成集 S_r 构成的不同组之间的连接 r -link。

1.3 BuNet 静态拓扑图的相关性质

性质 1 BuNet 静态拓扑图的度为 $(k-2+2r)$ 。

Cayley 图的度等于 S 集合的势,所以 BuNet 静态拓扑图节点的度为 $|S| = |S_k| + |S_r| = k-2+2r$,而该静态图为 $(k-2+2r)$ 正则图。

性质 2 BuNet 静态拓扑图的度可以达到 $O(\log n)$,直径可以达到 $O(\log n / \log \log n)$ 。

BuNet 静态拓扑图的节点个数为 $n = kr^k$,假设 $k = \log n / \log \log n$,那么 $r = (n / (\log n / \log \log n))^{1 / (\log n / \log \log n)} < \log n$ 。根据性质 1,节点的度为 $k-2+2r = \log n / \log \log n - 2 + 2r < \log n / \log \log n - 2 + 2 \log n$ 。BuNet 静态拓扑图的度和直径的下界分别是 $\Omega(\log n)$ 和 $\Omega(\log n / \log \log n)$,因此将该静态拓扑图作为 P2P 覆盖网络的理论原型有可能获得较好的网络性能^[9]。

2 BuNet 协议设计

Chord 和 CAN 等的网络拓扑结构并不是 Cayley 图,它们不具备很强的对称性。本章将介绍把 P2P 对等点嵌入到其对应的基于 Cayley 图的静态拓扑图中,并形成 BuNet 的 P2P DHT 覆盖网络拓扑算法。

2.1 BuNet 的标志符空间

在 BuNet 中,每个对等点的标志符是由一个二元组 (p, l) 惟一确定。其中, $p \in \{ (c_0, c_1, \dots, c_s) \mid s = \text{binarylen}(r) \times k, c_i \in Z_2 \cup \{*\} \}$, $l \in Z_k$ (binarylen 是将十进制转换为二进制,去掉前导 0 后的长度)。从定义可以看出,对等点标志符和静态拓扑图的顶点标志符明显的区别在于 p 是二进制的字符串,由“0”“1”和“*”组成。其中,“*”代表可以匹配“0”或“1”。因为 p 是定长的,所以在实际应用中,为了使标志符空间足够大, p 采用 160 bit 的大二进制数,可以取 $k=8, r=2^{19}$ 。而对等点标志符 $(p; l)$ 和静态拓扑图顶点标志符 $(v_0, v_1, \dots, v_{k-1}; i)$ 可以互相转换: $l=i$,而 p 中的 $i, i+k, i+2k \dots$ 位就是 v_i 的二进制表示。将 v_i 称为 v 的第 i 维; p 中各 $i, i+k, i+2k \dots$ 位称为 p 的第 i 维。例如对等点标志符为 $(1011, 0)$,则其等价的静态拓扑图顶点标志符为 $(31, 0)$ 。每个对等点根据“*”匹配负责一小片区域,例如对等点标志符 $(0*1*, 1)$ 负责的顶点区域为 $(0010, 1)$ 、 $(0011, 1)$ 、 $(0110, 1)$ 和 $(0111, 1)$ 。

2.2 BuNet 的分布式散列表

在 BuNet 的 DHT 中的数据文件名和对象索引都会被映射到一个二元组 (α, β) 。其中, α 是定长的二进制字符串 ($|\alpha| \geq |p|, \beta \in Z_k$)。标志符为 (p, l) 的对等点负责所有满足 p 是 α 的

前缀,并且 $l=\beta$ 的散列表表项。例如对等点 $(0111, 2)$ 负责键值 $(011100, 2)$ 、 $(011101, 2)$ 等表项。把 α 称为组标志符, β 称为层标志符,把所有满足 p 是 α 前缀的 α 组成的标志符空间称为对等点 p 负责的组区域。

2.3 BuNet 的拓扑

BuNet 拓扑是 BuNet 静态拓扑图在动态网络的映射。在 BuNet 的拓扑中, l 层中的对等点 (p, l) 将连向 $l+1$ 层和 $l-1$ 层的节点 $(p', l+1)$ 、 $(p', l-1)$ 。其中, p 和 p' 在除第 i 维外,其他维上都匹配,第 i 维可匹配可不匹配;同时 (p, l) 还连向 $(p, l+1)$ 、 $(p, l+2), \dots, (p, l+k-1)$ 。也可以说两个对等点 (p, l) 、 (p', l') 相邻,当且仅当它们满足下面两个条件之一: $l'=l \pm 1$ 且 p 和 p' 在 $z_k - \{l\}$ 维上相匹配,它们之间的连接对应于 r -link; $l' \neq l$,但 p 和 p' 在所有维上都匹配,它们之间的连接对应于 k -link。

图 1 是 BuNet 静态拓扑图与动态拓扑图的映射关系。其中静态拓扑图中灰色区域对应由一个动态对等点负责,如 $P0(00*, 0)$ 负责 $(000, 0)$ 和 $(001, 0)$ 的区域,而 $P6(1*0, 2)$ 负责 $(100, 2)$ 和 $(110, 2)$ 的区域等。为了计算对等点所负责的区域,只需要将对等点标志符中“*”替换成“0”或“1”。若标志符中“*”的个数为 m ,则对等点负责区域顶点个数为 2^m 个。而对等点的邻居连接关系和静态拓扑定义一样,只是在标志符匹配上面要多考虑一个“*”,而“*”可以匹配“0”“1”和“*”。如 $P_4(1***, 1)$ 在第零维和第二维上都与 $P_6(1*0, 2)$ 匹配,因而 P_4 和 P_5 是邻居;但是 P_4 在第二维上与 $P_5(0***, 2)$ 不匹配,因而 P_4 和 P_5 没有邻居关系。对等点邻居关系也可以从另外一个方面来得到,即如果对等点负责区域中有顶点和另一个对等点负责区域的顶点有边连,则这两个对等点有边连。

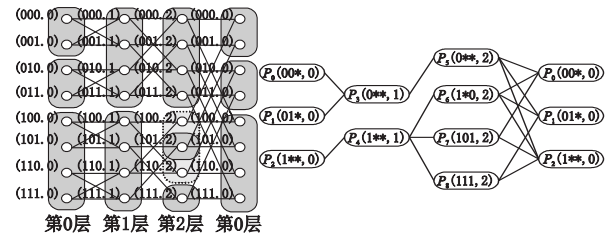


图1 BuNet静态拓扑图与具有12个对等点的动态拓扑映射关系

2.4 BuNet 的路由

BuNet 中路由包括查找对等节点和数据对象索引,对象索引由包含其前缀的对等点负责,所以数据对象索引的查找可归结于目标节点的查找。BuNet 的路由算法框架是通过每一跳来纠正当前标志符与目标标志符的差别,同时由于网络动态性,算法引入了容错性。算法 1 给出了在 BuNet 中路由的方法。例如,对于图 1 的 BuNet 拓扑结构,假设当前路由算法的初始入口节点为 P_1 ,目标节点为 P_7 。则基于每一跳匹配一位的原则,路径为 $P_1(01*, 0)$ 、 $P_3(0***, 1)$ 、 $P_5(0***, 2)$ 、 $P_2(1***, 0)$ 和 $P_7(101, 2)$ 。但是对等点可能因为各种原因而失效,那么可通过 13~15 行的容错算法,选择下一跳以避免查询失败。比如之前查询中如果 P_3 失败,将在其他邻居中任选一个作为下一跳。

算法 1 BuNet 路由算法

```
请求节点 (P, i) 查找目标节点 (α, j)
输入: 目标节点 (α, j)
输出: 下一跳
begin
if (P 和 α 在所有维度上都匹配)
if (i == j) 则到达目标节点;
else 通过 k-link 可以到达目标, 即 j 层;
```

```

else
    if(P 和  $\alpha$  在第  $i$  维上匹配)
        for( $c = 1, k - 1$ )
            if(P 和  $\alpha$  在第  $i + c$  维上不匹配)通过  $k$ -link 到达下一跳( $P', i + c$ );
            else 通过  $r$ -link 修正,使得  $P'$  和  $\alpha$  在第  $i$  维上匹配,下一跳为( $P', i + 1$ );
        if(下一跳为失效节点)
            if(P 和  $\alpha$  在所有维度上都匹配)目标节点失效,失败退出
            else 在  $r$ -link 和  $k$ -link 中随机选择一个邻居作为下一跳;
    end
end

```

2.5 加入 BuNet

一个新的对等点 P_1 的加入包括寻找网络位置、更新标志符、更新路由表以及更新分布式散列表四个步骤。寻找网络位置是为了确定新的对等点在现有 BuNet 中的位置,它首先使用既定的规则生成一个先验标志符(PriP, i),但此标志符并不是它最后所使用的标志符;然后在 BuNet 中随机选一个初始节点 P_2 作为入口节点,通过路由算法查找到负责(PriP, i)标志符的对等点 P_3 ,把 P_1 和 P_3 称为伙伴对等点。更新标志符最终生成 P_1 和 P_3 的更新标志符。标志符的更新方法为:从右到左查找 P_3 标志符中第一个为“*”的位置, P_1 标志符保持该位置的字符不变,将其他位上的字符从 P_3 上拷贝; P_3 将该位置的“*”改为不同于 P_1 该位置字符的“0”或“1”,即 P_1 和 P_3 标志符只有一位不相同。这样原来由 P_1 负责区域分裂成两半,一半留给自己继续负责,另一半分配给 P_3 负责。例如在 BuNet(4, 2)中,分裂前 $P_1 = (0 * * 1, 0)$,它负责的区域为(0001, 0)、(0011, 0)、(0101, 0)和(0111, 0)四个点。分裂后, $P_1 = (0 * 11, 0)$,负责的区域变为(0011, 0)和(0111, 0);而剩下两个点由 $P_3 = (0 * 01, 0)$ 负责。加入 BuNet 的最后两步是更新路由表和分布式散列表。更新路由表是为了维持 BuNet 拓扑,保证路由正确;而更新分布式散列表是将那些适合 P_3 标志符的散列表项从 P_1 迁移到 P_3 。由于 P_1 和 P_3 所负责的区域都分别是 P_1 所负责区域的子集, P_1 和 P_3 的邻居集以及散列表都分别是原来 P_1 的邻居集和散列表的子集。于是,循环检查 P_1 原来的路由表和散列表即可构造 P_1 和 P_3 的新路由和散列表。最后由于 P_1 和 P_3 有可能相邻,如果相邻则对方标志符加入到自身路由表中。算法 2 为 BuNet 的节点加入算法,图 2 为 BuNet 对等点逐步加入的构建过程。其中网络的初始状态要求每一层至少部署一个对等点,它们层标志符各不相同。

算法 2 BuNet 的节点加入算法

```

输入:随机选取的入口节点  $P_2$  的标志符( $p_2, i_2$ );
输出:待加入 BuNet 的对等点  $P_1$  的标志符( $p_1, i_1$ )、路由表和分布式散列表
更新后的  $P_3$  标志符、路由表和分布式散列表。
begin
     $\text{PriP}_1 =$  长度为  $s$ ,通过既定规则生成的组标志符; $i_1 = Z_k$  中随机选取的整数;
    ( $p_3, i_3$ ) 为以( $p_2, i_2$ )为入口节点, ( $\text{PriP}_1, i_1$ ) 为查询目标节点,调用路由算法所返回的目标对等点  $P_3$ ;
    Diff 为  $p_3$  和  $\text{PriP}_1$  从右到左第一个不同位的下标;
     $p_3[\text{Diff}] = ! \text{PriP}_1[\text{Diff}], p_1 = p_3, p_1[\text{Diff}] = \text{PriP}_1[\text{Diff}];$ 
    for(( $p_n, i_n$ ) in  $P_3$  的路由表)
        if(( $p_n, i_n$ ) 是 ( $p_1, i_1$ ) 的邻居)
            把( $p_n, i_n$ ) 添加到  $P_1$  的路由表,把  $P_1$  添加到( $p_n, i_n$ ) 的路由表;
        if(( $p_n, i_n$ ) 不再是 ( $p_3, i_3$ ) 的邻居)
            把( $p_n, i_n$ ) 从  $P_1$  的路由表中删除,把  $P_1$  从( $p_n, i_n$ ) 的路由表中删除;
        if(( $p_1, i_1$ ) 是 ( $p_3, i_3$ ) 的邻居)
            把( $p_1, i_1$ ) 添加到( $p_3, i_3$ ) 的路由表,把( $p_3, i_3$ ) 添加到( $p_1, i_1$ ) 的路由表;

```

```

for(hashItem in  $P_3$  的分布式散列表)
    if(hashItem 不再匹配  $P_3$  的新标志符)
        将 hashItem 从  $P_3$  的分布式散列表中删除;
        把 hashItem 添加到  $P_1$  的分布式散列表中;
end

```

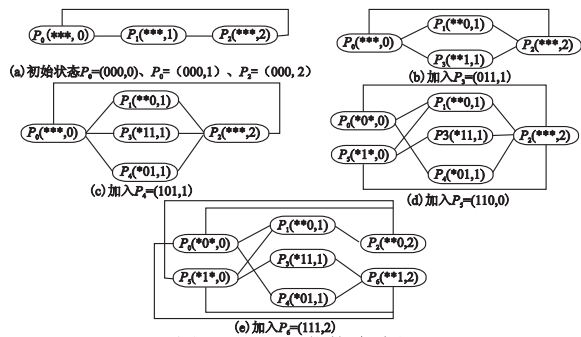


图2 BuNet网络构建过程

2.6 退出 BuNet

BuNet 网络中的节点可以在任何时刻退出。在理想情况下,节点在退出网络之前通知其邻居,并更新路由表、分布式散列表等,称此次退出为正常退出,否则称为非正常退出。若一个节点 P_1 要退出网络时,它首先需要查询它的伙伴对等点是否还没有分裂,即两个对等点负责的区域是一样大的。如果还没有分裂,则 P_1 将自己负责的区域交给其伙伴对等点负责,而伙伴对等点将标志符中和 P_1 标志符惟一不同的那一位修改为“*”,然后伙伴对等点执行加入算法的逆操作,对路由表和分布式散列表进行更新,同时通知邻居对等点更新路由表和分布式散列表。而如果 P_1 的伙伴对等点所在的区域已经进行了多次分裂,则 P_1 将在伙伴对等区域中用深度优先搜索算法(DFS)搜索一个负责区域最小的对等点 P_i ,让 P_i 负责 P_1 的区域;而 P_i 将其原来负责区域交给其伙伴对等点(一定是未分裂对等点)负责。最后各个对等点执行路由表、分布式散列表更新操作。假如图 2(d)中 $P_1(* * 0, 1)$ 节点要退出网络,而它的伙伴对等点已经分裂;则算法将选择一个最小对等点 $P_3(* 11, 1)$ 来负责 P_1 的区域,而 P_3 的标志符修改为($* * 0, 1$); P_4 将接管 P_3 原来的区域,同时 P_4 的标志符修改为($* * 1, 1$)。最后 P_3 和 P_4 更新路由表和分布式散列表。虽然此操作比较繁琐,但是这种情况出现的概率比较低。

节点非正常退出需要多一个失败检测节点:失败检测策略包括向邻居定期发送 Keep-alive 包以及异步检测机制。检测到错误后,该节点帮助非正常退出节点完成清理更新工作。而注意到非正常退出节点的分布式散列表无法通过其邻居节点进行恢复,因此需要采取其他措施如邻居节点定期备份或资源拥有者定期发送对象索引等保证健壮性。

3 性能评估

通过模拟仿真实验来评估 BuNet 的各项性能指标。本文选择 Chord^[1] 作为比较对象,是因为 Chord 代表了基于 DHT 协议的覆盖网络算法,它们的网络直径为 $O(\log n)$,路由表大小为 $O(\log n)$,而且 BuNet 也是基于 DHT 协议的。所有的实验都是用 C/C++ 编写的,评估测试都在单进程内完成,不涉及任何真正意义上的网络通信。网络中节点的个数为 256 ~ 4M。Chord 协议是参照文献[1]中算法实现的。最后所有模拟实验都是多次实验结果的平均值。

3.1 路由长度

图 3 显示了不同网络规模下,BuNet、Chord 和 CAN 的最大路由长度和平均路由长度。其中路由的源对等点和目标对等

点都是在标志符范围内随机选取的。实验评估了各种网络规模的路由路径特性。从中可以看出,BuNet、Chord 和 CAN 的平均路由长度增长缓慢,且始终没有超过它们各自的网络直径。而 BuNet 的最大路由长度约等于 $\log n/\log\log n$,Chord 和 CAN 的最大路由长度则约等于 $\log n$ (其中, n 为网络节点个数)。在相同的网络规模下,BuNet 的路由长度都比 Chord 和 CAN 要小,这就缩短了路由查找的时间。图 4 给出了由 4M 个对等点组成的网络的路由长度分布图,由图中可以看出路由长度都集中在一定的范围内(BuNet 路由长度主要集中在 5~6,Chord 的路由长度集中在 9~15,CAN 的路由长度集中在 15~23)。路由长度的方差与基于 P2P 覆盖网络的网络抖动有很大的关联:小的方差可能得到低的网络抖动,而低的网络抖动对实时应用具有决定性的意义。在网络抖动方面:BuNet < Chord < CAN。

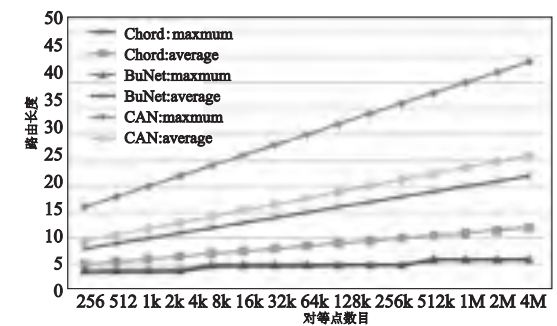


图3 最大、平均路由长度与网络大小关系图

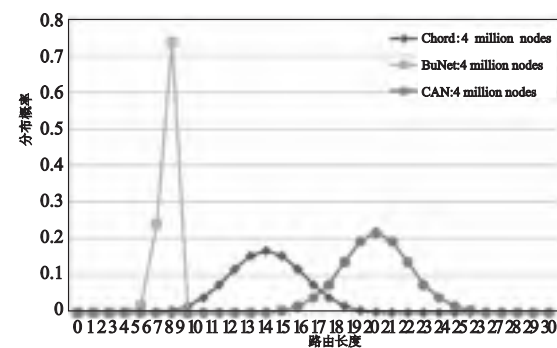


图4 含4M个对等点网络下的网络路由长度分布图

3.2 鲁棒性

鲁棒性是测试当网络中对等点以不同的概率失效时,路由失败的比率(包括入口节点失效、目标节点失效和查找过程的失败等)与相应情况下成功路由的长度。实验中测试的网络规模均为 4M 个对等点。路由测试开始前,先按照一定的比例随机生成失效节点,然后在该含有失效节点的覆盖网络上运行路由算法,计算路由失败率以及路由成功情况下的平均路由长度。对等点的失败率为 0~20%,对于每一次路由测试都随机选取源和目标,每个失败率下都运行 40M 次路由查找。图 5 给出 BuNet、Chord 和 CAN 在不同对等点失效概率下路由查询最终失败的概率。在相同的对等点失效概率下,BuNet 的路由

失败率比 Chord 和 CAN 要低。路由过程中发生错误时会增加该路由的长度,图 6 显示出路由容错对平均路由长度的影响。BuNet、Chord 和 CAN 的平均路由长度增幅类似,但是 BuNet 的路由长度总要比 Chord 和 CAN 的要小。

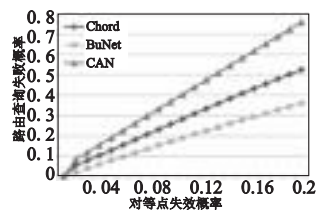


图5 路由失败率和对等点失效概率关系图

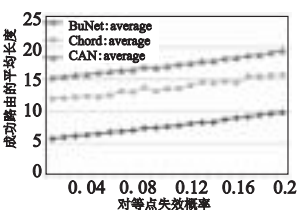


图6 平均路由长度和对等点失效概率关系图

4 结束语

针对传统 P2P 覆盖网络缺乏分组浏览功能的问题,本文提出了一种新的基于 Cayley 图的扩展 butterfly 结构拓扑 BuNet,并在上面设计了 P2P 拓扑查询路由算法和加入退出算法等。仿真实验结果表明,与 Chord 相比,BuNet 能在同等路由表大小前提下可获得更短查询路径,而且它的路由容错性和鲁棒性也更加优秀。下一步,笔者将在大规模网络下模拟测试 BuNet 的动态路由算法,并进一步分析网络拓扑在大规模网络环境下的特性;同时研究 BuNet 在组播、负载均衡方面的性能。

参考文献:

[1] STOICA I, MORRIS R, LIBEN-NOWELL L, et al. Chord: a scalable peer-to-peer lookup protocol for Internet applications [J]. IEEE/ACM Trans on Networking, 2003, 11(1): 17-32.

[2] RATNASAMY S, FRANCIS P, HANDLEY M, et al. A scalable content addressable network [C]//Proc of Annual Conference ACM Special Interest Group on Data Communications (ACM SIGCOMM 2001). 2001.

[3] AKERS S B, KRISHNAMURTHY B. A group-theoretic model for symmetric interconnection networks [J]. IEEE Trans on Computers, 1989, 38(4): 555-566.

[4] BIGGS N. Algebraic graph theory [M]. Cambridge: Cambridge University Press, 1993.

[5] QU C, NEJDL W, KRIESEL M. Cayley DHTs: a group-theoretic framework for analyzing DHTs based on Cayley graphs [C]//Proc of the 2nd ISPA. 2004: 914-925.

[6] XU J. On the fundamental tradeoffs between routing table size and network diameter in peer-to-peer networks [C]//Proc of IEEE Infocom. 2003: 2177-2187.

[7] SIEGEL H J. Interconnection networks for SIMD machines [J]. IEEE Computer, 1979, 12(6): 57-64.

[8] HEYDEMANN M C, DUCOURTHIAL B. Cayley graphs and interconnection networks [M]//Graph Symmetry, Algebraic Methods and Applications. [S. l.]: Kluwer Academic Publishers, 1997: 167-224.

[9] ABERER K, ALIMA L O, GHODSI A, et al. The essence of P2P: a reference architecture for overlay networks [C]//Proc of the 5th IEEE International Conference on Peer-to-Peer Computing. 2005.

(上接第 637 页)

[5] YOO M, QIAO C, DIXIT S. Optical burst switching for service differentiation in the next-generation optical Internet [J]. IEEE Communications Magazine, 2001, 29(2): 98-104.

[6] VU H L, ZUKERMAN M. Blocking probability for priority classes in optical burst switching networks [J]. IEEE Communications Letters, 2002, 6(5): 214-216.

[7] KASHYAP K M, JOHN E M, ALAN E W, et al. Analysis of an opti-

cal burst switching router with tunable multiwavelength recirculating buffers [J]. Journal of Lightwave Technology, 2005, 23(10): 3302-3312.

[8] YOO M, QIAO C, DIXIT S. QoS performance of optical burst switching in IP-over-WDM networks [J]. IEEE Journal on Selected Areas Communications, 2002, 18(10): 2062-2071.

[9] GROSS D, HARRIS C. Fundamentals of queueing theory [M]. New York: Wiley-Interscience, 1997: 11-23.