

随机步长无向环网通信延迟的研究

方木云,王 俊,王 超

(安徽工业大学 计算机科学与技术学院,安徽 马鞍山 243032)

摘要:高性能计算机(HPC)系统期望尽可能低的通信延迟和建造成本。传统固定步长拓扑已经无法降低通信延迟和建造成本,如固定步长环网无法突破 Wong 和 Coppersmith 给出的下界;高节点度拓扑能进一步降低延迟,但增加的交换机和物理链路提高了建造和运营成本。针对这两个缺点,采用随机步长构造方法来生成一种新型的无向环网,避免高节点度的同时,将减少节点间步长的长度从而降低通信延迟作为目标。通过仿真实验分别对无向环网随机步长的直径、平均距离和固定步长的直径下界、平均距离下界进行比较。结果表明:在一定节点度范围内,随机步长无向环网得到的值小于传统固定步长环网得到的值。因此,随机步长拓扑可成为下一代高性能计算机潜在的拓扑结构。

关键词:无向环网;固定步长;随机步长;通信延迟

中图分类号:TP393

文献标识码:A

文章编号:1673-629X(2016)10-0027-05

doi:10.3969/j.issn.1673-629X.2016.10.006

Research on Communication Delay of Random-step Undirected Loop Networks

FANG Mu-yun, WANG Jun, WANG Chao

(School of Computer Science and Technology, Anhui University of Technology, Ma' anshan 243032, China)

Abstract: High-performance computer system expects the lowest possible communication delays and construction costs. Traditional fixed-step topology has been unable to reduce communication latency and construction costs. For example, fixed-step loop networks cannot break through the lower bound given by Wong and Coppersmith, and high degree of topological node can further reduce latency, but increased switch and physical links improve the construction and operating costs. In view of two shortcomings, random-step is used to construct a new type of undirected loop networks which can reduce the length of steps between nodes to reduce the communication delay as the destination, at the same time can avoid high degree of node. Respectively compares diameter and average distance of random-step undirected loop networks as well as lower diameter and average lower distance of fixed-step through simulation experiments. The results show that within a certain range of the node, the value of the random-step to the ring obtained less than the value of traditional fixed-step ring obtained. Thus, random-step topology may be the next generation of high-performance computer underlying topology.

Key words: undirected loop networks; fixed-step; random-step; communication delay

1 概述

目前,高性能计算机技术已成为世界各国竞相争夺的战略制高点,是衡量一个国家综合国力的重要标志,以服务国家经济建设和改善民生为最高目的,并广泛应用于国家经济和人民生活相关领域^[1]。无向多环网络即 m ($m \geq 2$) 环网络是计算机互连网络或通讯系统的一类重要拓扑结构,广泛用于计算机局域网和各种并行处理结构^[2-4],其图论模型是指这样一个无向图 $G(N; \pm r, \pm m, \pm s)$,节点记为 $0, 1, \dots, N-1$,每个

节点 i 向相邻节点发出 6 条有向边: $i \rightarrow i + r \pmod{N}$ 、 $i \rightarrow i + m \pmod{N}$ 、 $i \rightarrow i + s \pmod{N}$ 、 $i \rightarrow i + N - r \pmod{N}$ 、 $i \rightarrow i + N - m \pmod{N}$ 和 $i \rightarrow i + N - s \pmod{N}$,分别记为 $[+r]$ 边、 $[+m]$ 边、 $[+s]$ 边、 $[-r]$ 边、 $[-m]$ 边和 $[-s]$ 边,其中 S 为自然数, $1 \leq r < m < s < N$ 且 $\gcd(N, r, m, s) = 1$ ^[5-7]。传统的固定步长无向环网通信延迟遭遇了瓶颈,例如:无向双环网和无向多环网都无法突破 Wong 和 Coppersmith 给出的通信延迟下界^[8]。

收稿日期:2016-01-08

修回日期:2016-04-13

网络出版时间:2016-09-19

基金项目:国家自然科学基金资助项目(61003311);安徽省教育厅重大项目(ZD2008005-1)

作者简介:方木云(1968-),男,教授,博士,研究方向为计算机网络、网络拓扑结构等;王 俊(1990-),男,通信作者,硕士研究生,研究方向为计算机网络与网络拓扑结构。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20160919.0843.062.html>

2012 年,日本国立情报学研究所 Michihiro Koibuchi 等提出在环网中采取随机连线的方式来构造环网,在同节点数和网络环数(即节点度)的情况下,发现网络的通信延迟比一种用在 on-chip 系统上的固定步长环网的通信延迟可下降 300 倍,同时发现同网络环数同节点数的随机步长环网与现在高性能计算机(HPC)中采用的传统环网(TORUS)等拓扑结构相比,其直径和平均距离小、扩展性和容错性好^[9-10]。

文献[11-12]分别指出了度量环网优越性的两个重要参数:直径、平均距离。直径和平均距离越小,无向环网的节点步长长度越小,则环网中通信延迟越小。但根据 Michihiro Koibuchi 等的研究,仍存在不完善的地方^[13-14]:

(1)在同节点数和网络环数的情况下,没有选择直径和平均距离达到 Wong 和 Coppersmith 给出的下界的紧优网络^[15]来与随机步长网络进行对比。选作对比的非随机步长(non-random shortcut)网络是固定步长环网中的最差情况之一,以此网络与随机步长网络进行对比,导致随机步长大幅降低通信延迟的这一结论可信度不高。

(2)没有分析随机步长构造无向 m 环网的直径、平均距离分布特性。因为随机步长的直径是变化的,所以固定步长也是随机步长的一个样本,Michihiro Koibuchi 等并没有具体对比两者在直径和平均距离上的联系与差异,以及没有给出在构造无向 m 环网中随机步长优于固定步长直径和平均距离的分布特性。

(3)在同环数的情况下,随着网络节点数的增多,即网络规模不断增大,随机步长网络通信延迟优于固定步长网络。对该结论没有给出分析研究。

Michihiro Koibuchi 等工作是针对任意度的环网,为了做好横向对比,文中针对同节点数情况下,节点度为 4/6/7/8/10(即环数为 2/4/5/6/8)的无向环网,同网络环数情况下,不同节点数的无向环网网络,以及将网络环数和节点数两者变化相结合的无向环网,重点分析这三个问题。

2 随机步长无向环网的生成

随机步长无向环网是在传统固定步长无向环网的基础上构造的,在传统固定步长环网生成时,将固定步长改为随机步长,同时改变网络环数。因此需依次遍历网络中各个节点。以节点数 $N=32$ 和环数 $L=4$ 为例,随机步长无向四环网络的生成如下:

步骤 1:生成一个节点数为 32 的环网,如图 1(a)所示。

步骤 2:从节点 1 出发,在节点度小于 6 且与节点 1 不相连的节点集合 $G_1 = \{3, 4, \dots, 31\}$ 中随机选出 4

个不相同的节点 a, b, c 和 d ,这里假设 $a=7, b=8, c=9, d=19$,连接节点 1 和 7,1 和 8,1 和 9,再连接节点 1 和 19,如图 1(b)所示。

步骤 3:逆时针依次遍历剩下的节点,若当前节点 i 的度数为 i_k ,则当前节点 i 可发出的随机边数 $i_{k'} = 6 - i_k$,在节点度小于 6 且与 i 节点不相连的节点集合 G_i 中,若 G_i 的元素个数小于 $i_{k'}$,则说明图 G 中已不存在足够的可选节点供节点 i 连接,返回步骤 1,重新生成环网。

步骤 4:重复执行步骤 3,直到所有节点满足条件。节点 2 的生成如图 1(c)所示,节点 3 的生成如图 1(d)所示。

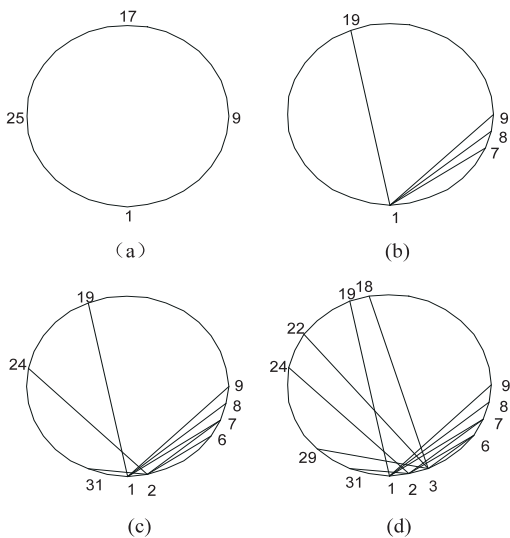


图 1 节点 1、2、3 生成示意图

由上述步骤可知,此次随机步长无向四环网络生成后的拓扑结构如图 2 所示。

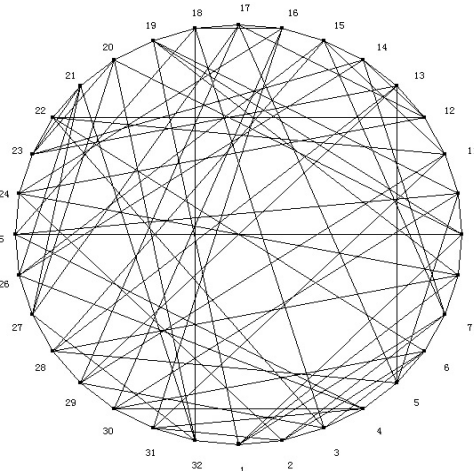


图 2 随机步长无向四环网络拓扑结构

3 仿真算法

3.1 随机步长无向环网的生成算法

Step1:构建一个 $N \times N$ 的邻接矩阵 $A[N-1, N-1]$,矩阵内的每个元素赋初值 0;构建一个长度为 N

的队列 $C[N-1]$, 队列中每个元素赋初值 0, 并建立一个空链表 L , 同时令网络的环数为 H , 这里 $H \geq 2$ 。

Step2: r 从 0 到 $N-1$ 循环: 置 $A[r, (r+N-1) \% N] = 1$ 和 $A[r, (r+N+1) \% N] = 1$ 。

Step3: s 从 0 到 $N-1$ 循环(对每个 s, k 从 $s+1$ 到 $N-1$ 循环):

(1) 记 $T = C[k]$, 若 $T = 0$, 则节点 r 上不需要再加入随机边, 开始下一次 r 循环。

(2) 清空 L , 若 $A[r, k] \neq 1$ 并且 $C[k] < H$, 则将 k 加入 L 中。

(3) 记 L 中元素的个数为 L_{Length} , 若 $L_{\text{Length}} < T$, 则图中已无足够的 r 点连接的节点数, 返回(1), 重新生成。

(4) 若 $T = 1$, 则在 L 中随机选择节点 d_1 , 置 $A[r, d_1] = A[d_1, r] = 1, C[r] = 2, C[d_1] = C[d_1] + 1$ 。

(5) 若 $T = 2$, 则在 L 中随机选择两个互不相同节点 d_1, d_2 , 置 $A[r, d_1] = A[d_1, r] = A[r, d_2] = A[d_2, r] = 1, C[r] = 2, C[d_1] = C[d_1] + 1, C[d_2] = C[d_2] + 1$ 。

(6) 若 $T = 3$, 则在 L 中随机选择三个互不相同节点 d_1, d_2, d_3 , 置 $A[r, d_1] = A[d_1, r] = A[r, d_2] = A[d_2, r] = A[r, d_3] = A[d_3, r] = 1, C[r] = 2, C[d_1] = C[d_1] + 1, C[d_2] = C[d_2] + 1, C[d_3] = C[d_3] + 1$ 。

(7) 若 $T = H$, 则在 L 中随机选择 H 个不相同节点 d_1, d_2, \dots, d_H , 置 $A[r, d_1] = A[d_1, r] = A[r, d_2] = A[d_2, r] = A[r, d_3] = A[d_3, r] = \dots = A[r, d_H] = A[d_H, r] = 1, C[r] = 2, C[d_1] = C[d_1] + 1, C[d_2] = C[d_2] + 1, C[d_3] = C[d_3] + 1, \dots, C[d_H] = C[d_H] + 1$ 。

3.2 随机步长无向环网的直径和平均距离求解算法

Step1: 对随机步长无向环网 $G = (V, E)$ 的邻接矩阵 A 进行如下变换:

$$D[i, j] = \begin{cases} 0, & i = j \\ 1, & A[i, j] = 1 \\ N, & A[i, j] = 0 \text{ 且 } i \neq j \end{cases}$$

其中, A 是具有如下性质的 N 阶方阵:

$$A[i, j] = \begin{cases} 1, & \text{若 } (V_i, V_j) \text{ 是 } E(RG) \text{ 中的边} \\ 0, & \text{若 } (V_i, V_j) \text{ 不是 } E(RG) \text{ 中的边} \end{cases}$$

Step2: r 从 0 到 $N-1$ 循环: 对每个 i, j 从 0 到 $N-1$ 循环, 对每个 j, k 从 0 到 $N-1$ 循环, 若 $D[j, i] + D[i, k] < D[j, k]$, 则 $D[j, k] = D[j, i] + D[i, k]$; 若 $D[j, k] < D[k, j]$, 则 $D[k, j] = D[j, k]$ 。

Step3: 对 Step2 中得到的距离矩阵 D , 求出直径^[12,16]:

$$\text{Dia} = \text{Max}(D[i, j]), 0 \leq i, j \leq N$$

Step4: 对 Step2 中得到的距离矩阵 D , 求平均距离^[12]: 万方数据

$$\text{AvgDia} = \frac{1}{N} \sum_{i=0}^{N-1} \left[\frac{1}{(N-1)} \sum_{j=0}^{N-1} D[i, j] \right], 0 \leq i, j \leq N$$

4 仿真实验及结果分析

由文献[12]可知传统固定步长环网的直径下界为:

$$D(N; \pm s_1, \pm s_2) \geq \frac{1}{2}[(l! N)^{\frac{1}{l}} - (2l+1)] \sim \frac{1}{2}(l! N)^{\frac{1}{l}}$$

其平均距离下界为:

$$M(N; \pm s_1, \pm s_2) \geq \frac{1}{2(l+1)! N}[(l! N)^{\frac{1}{l}} - (3l+1)]^{l+1} \sim \frac{l}{2(l+1)}(l! N)^{\frac{1}{l}}$$

文中选择与传统固定非单位步长无向环网的下界作对比实验。选用 Visio Studio 2013 中 C# 语言编制程序进行仿真, 计算直径和平均距离, 同时将网络拓扑结构和得到的结果数据输入到 Matlab 中进行分析。

下面给出仿真实例:

(1) 当环数 $L = 4$ 不变, 改变节点数时, 进行 100 次对随机步长无向环网的仿真实验, 对其产生的直径的平均值和平均距离的平均值与固定步长环网的直径下界和平均距离下界进行对比并做出相应的分析, 如图 3 和图 4 所示, 其数据对比如表 1 所示。

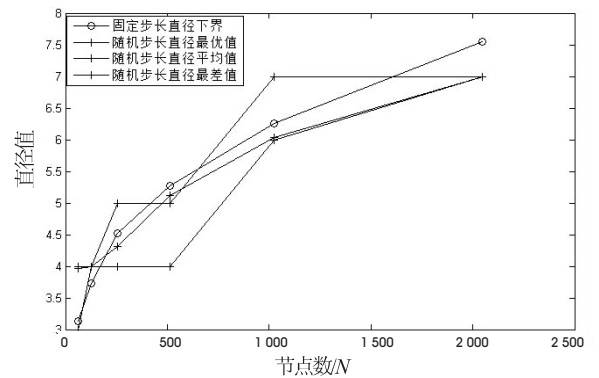


图 3 $L = 4$ 时直径对比图

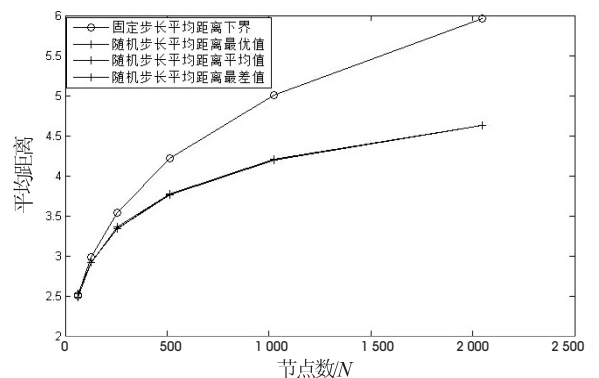


图 4 $L = 4$ 时平均距离对比图

表 1 环数 $L=4$ 时随机步长与固定步长的
直径、平均距离数据对比表

节点数 N	固定步长直 径/平均距离	随机步长(直径/平均距离)		
		最优生成	最差生成	平均生成
64	3.130/2.504	3/2.486	4/2.533	3.970/2.488
128	3.723/2.979	4/2.915	4/2.915	4.000/2.915
256	4.527/3.542	4/3.334	5/3.356	4.315/3.341
512	5.264/4.212	5/3.765	6/3.771	5.120/3.768
1 024	6.260/5.008	6/4.198	7/4.203	6.030/4.201
2 048	7.545/5.956	7/4.625	7/4.625	7.000/4.625

由图 3、4,表 1 得出如下结论(固定步长无向环网简称固定步长,随机步长无向环网简称随机步长):

①随机步长的平均生成直径和平均生成的平均距离均小于固定步长的直径下界和平均距离下界。

②随着节点数 N 不断增大,固定步长直径下界与平均距离下界的值增加速度明显快于随机步长平均生成直径和平均生成的平均距离的值增加速度。

由上述结论可知,在保持无向环网中其他因素相同时,当环数为定值,随机步长的通信延迟优于传统固定步长。

(2)当节点数 $N=1\,024$ 不变,改变网络环数时,同样进行相应的仿真实验,并对其结果做出分析,如图 5 和图 6 所示,其数据如表 2 所示。

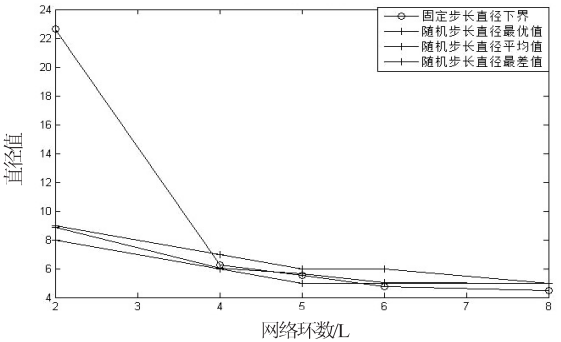


图 5 $N=1\,024$ 时直径对比图

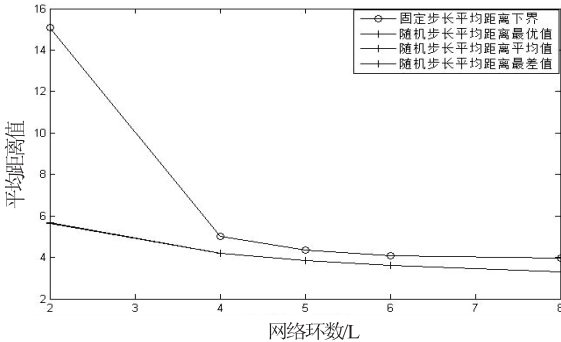


图 6 $N=1\,024$ 时平均距离对比图

由图 5、6 和表 2 表明:

①随机步长平均生成的平均距离均小于固定步长平均距离的下界。

表 2 $N=1\,024$ 时随机步长与固定步长的
直径、平均距离数据对比表

环数 L	固定步长直 径/平均距离	随机步长(直径/平均距离)		
		最优生成	最差生成	平均生成
2	22.628/15.085	8/5.638	9/5.683	8.860/5.660
4	6.260/5.008	6/4.198	7/4.203	6.030/4.201
5	5.521/4.342	5/3.856	6/3.858	5.640/3.857
6	4.753/4.074	5/3.610	6/3.614	5.001/3.612
8	4.477/3.979	5/3.303	5/3.303	5.000/3.303

②随着网络环数 L 不断增大,随机步长的直径和平均距离均小于固定步长直径下界和平均距离下界的趋势越来越不明显。

由上述两点,在保持无向环网中其他因素相同时,当网络环数(即节点度)在一定范围内,随机步长的通信延迟优于传统固定步长,这一点也避免了高节点度拓扑所带来的建造和运营成本的问题。

(3)结合实验(1)、(2),同时改变节点数 N 和网络环数 L ,进行 100 次直径和平均距离对比实验,并对结果做出分析。实验结果如表 3 所示。

表 3 固定步长直径/平均距离
(随机步长直径/平均距离)

	2	4	5	6	8
64	5.7/3.77 (5.5/3.18)	3.3/2.51 (3.2/2.49)	3.0/2.50 (4.0/2.32)	3.0/2.57 (3.5/2.20)	3.2/2.82 (3.0/2.00)
128	8.0/5.34 (6.2/3.79)	4.0/2.98 (3.9/2.91)	3.5/2.87 (4.1/2.69)	3.4/2.88 (4.0/2.54)	3.5/3.07 (3.9/2.35)
256	11.3/7.54 (7.1/4.41)	4.6/3.55 (4.3/3.34)	4.0/3.29 (5.0/3.09)	3.8/3.24 (4.2/2.90)	3.8/3.35 (4.0/2.65)
512	16.0/10.68 (8.0/5.04)	5.3/4.22 (5.1/3.77)	4.6/3.78 (5.1/3.47)	4.3/3.63 (4.9/3.27)	4.1/3.65 (4.1/2.96)
1 024	22.6/15.09 (8.9/5.66)	6.3/5.01 (6.0/4.20)	5.3/4.34 (5.6/3.85)	4.7/4.08 (5.0/3.61)	4.5/3.98 (5.0/3.30)
2 048	32.0/21.33 (9.1/6.29)	7.6/5.96 (7.0/4.62)	6.0/4.99 (6.0/4.25)	5.8/4.58 (6.0/3.97)	4.9/4.34 (5.0/3.60)

分析其总体趋势如下:

①当 $L \leq 4$ (即节点度 ≤ 6) 不同节点数时,随机步长直径、平均距离均小于固定步长直径下界和平均距离下界;当 $L \geq 5$ 时,随机步长直径 \geq 固定步长直径下界,但随机步长平均距离 \leq 固定步长平均距离下界。

②当 L 一定节点数不断增加时,随机步长直径和平均距离的值增加速度明显小于固定步长直径下界和平均距离下界的值增加速度;当 N 一定环数不断增加时,虽然随机步长直径 \geq 固定步长,但随机步长直径的值增加速度相比固定步长直径下界的值增加速度较慢,且随机步长平均距离均小于固定步长平均距离下界。

综上,在保持无向环网中其他因素相同时,当网络

环数(即节点度)在一定范围内,随机步长得到的直径、平均距离均小于传统固定步长直径下界、平均距离下界,则随机步长网络节点步长长度较小,从而得出结论:在无向环网中,随机步长较传统固定步长通信延迟占有优势。

5 结束语

文中采用随机步长来构造无向环网,分别对随机步长无向环网直径、平均距离和固定步长无向环网直径下界、平均距离下界进行仿真实验对比。结果表明,在保持无向环网中其他因素相同时,当节点度在一定范围内,随机步长无向环网的网络通信延迟和网络性能均优于固定步长无向环网,同时随着网络节点数不断增加,即网络规模的不断增大,随机步长降低通信延迟的优势更加明显。因此,与当前主流的 HPC 采用的高节点度和固定步长拓扑结构相比,随机步长拓扑可成为下一代高性能计算机的潜在拓扑结构。

参考文献:

[1] 周兴铭. 高性能计算技术发展[J]. 自然杂志,2011,33(5): 249-254.

[2] Chen C Y, Hwang F K. Equivalent L-shapes of double-loop networks for the degenerate case[J]. Journal of Interconnection Networks,2000,1(1):47-60.

[3] 陈协彬. 步长有限制的双环网络的最优路由算法[J]. 计算机学报,2004,27(5):596-603.

[4] Hwang F K. A survey on multi-loop networks[J]. Theoretical Computer Science,2003,299(1):107-121.

[5] 方木云,屈玉贵,赵保华. 双环网络的[+h]边优先寻径策略[J]. 计算机学报,2008,31(3):536-542.

[6] 苏小虎,方木云,邵伟鹏,等. 双环网络 $G(N;1,s)$ 的 L 形瓦仿真算法改进[J]. 小型微型计算机系统,2012,33(9):

2053-2055.

[7] 方木云,赵保华,屈玉贵. 非单位步长双环网络 $G(N; r, s)$ 的 L 形瓦仿真算法[J]. 系统仿真学报,2006,18(10):2963-2965.

[8] Wong C K, Coppersmith D. A combinatorial problem related to multi-module memory organizations[J]. Journal of ACM, 1974,21(3):392-402.

[9] Koibuch M, Matsutani H, Amano H, et al. A case for random shortcut topologies for HPC interconnects[C]//Proc of 39th annual international symposium on computer architecture. [s. l.]:IEEE,2012:177-188.

[10] Kim J, Balfour J, Dally W. Flattened butterfly topology for on-chip networks[C]//Proceedings of the 40th annual IEEE/ACM international symposium on micro-architecture. [s. l.]: IEEE Computer Society,2007:172-182.

[11] 方木云,侯海金,吴爱清,等. 双环网络直径点和宽直径点的分布特性[J]. 小型微型计算机系统,2013,34(4):749-752.

[12] 陈业斌,李颖,郑啸,等. 关于有向环网平均直径的研究[J]. 通信学报,2013,34(2):138-146.

[13] Fujiwara I, Koibuchi M, Casanova H. Cabinet layout optimization of supercomputer topologies for shorter cable length[C]//Proc of 13th international conference on parallel and distributed computing, applications and technologies. [s. l.]: IEEE,2012:227-232.

[14] Koibuchi M, Fujiwara I, Matsutani H, et al. Layout-conscious random topologies for HPC off-chip interconnects[C]//Proc of 19th international symposium on high performance computer architecture. [s. l.]:IEEE,2013:484-495.

[15] 方木云,赵保华,屈玉贵. 基于圈的紧优双环网络 $G(N;1,s)$ 求解算法[J]. 华中科技大学学报:自然科学版,2005,33(6):17-19.

[16] 方木云,赵保华. 新的无向双环网络 $G(N;\pm 1,\pm s)$ 直径求解方法[J]. 通信学报,2007,28(2):124-129.

(上接第 26 页)

IEEE Transactions on Engineering Management,2005,52(4): 509-521.

[8] Ba E H, Hu W C, Yo O W S. Document configuration control processes captured in a workflow[J]. Computers in Industry, 2004,53(2):117-131.

[9] Meziane F, Rezgui Y. A document management methodology based on similarity contents[J]. Information Sciences,2004, 158(1):15-36.

[10] 徐立臻,徐宏炳. 面向对象数据库系统中的版本管理[J]. 东南大学学报:自然科学版,1999,29(3):34-38.

[11] 张维,何卫平,刘平,等. PDM 实施中的版本管理研究与应用[J]. 组合机床与自动化加工技术,2001(6):11-12.

[12] 刘方鑫,李毅. 变量化图形 CAD 系统中的版本管理[J]. 计算机应用,2000,20(S):111-112.

[13] 陆嘉恒. 大数据挑战与 NoSQL 数据库技术[M]. 北京:电子工业出版社,2013:7-8.

[14] 冯向兵,莫蓉,桂元坤. PDM 中版本管理的图模型表达方法与实现技术[J]. 航空制造技术,2008(12):89-92.

[15] 夏秀峰,富钢,丛丽晖,等. 一种基于 SQL 的层次查询方法[J]. 微处理机,2001(1):42-44.