一种基于超立方体网络的高效故障诊断并行算法

郭 杨[1,2] 梁家荣[1,2] 刘 峰[1,2] 谢 敏[1]
(广西大学计算机与电子信息学院 南宁 530004)[1]
(广西多媒体通信与网络技术重点实验室 南宁 530004)[2]

摘 要 超立方网络是一种重要的网络拓扑结构。针对现有的超立方网络故障诊断算法复杂度高的问题,引入故障扇的概念,采用并行深度优先搜索策略设计算法,通过算法寻找超立方体网络中的故障扇,确定该网络的故障节点,以便替换或修复,为增强网络的可靠性提供了一条重要的新途径。最后对所提算法的复杂性进行了分析,证明了该算法的时间复杂度不超过 $O(N)$,远优于现有复杂度超过平方级的算法。

关键词 超立方网络,故障诊断,故障扇,系统级诊断
中图法分类号 TP373 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2019.05.011

Novel Fault Diagnosis Parallel Algorithm for Hypercube Networks

GUO Yang[1,2] LIANG Jia-rong[1,2] LIU Feng[1,2] XIE Min[1]
(School of Computer and Electronics Information, Nanning 530004, China)[1]
(Guangxi Key Laboratory of Multimedia Communications and Network Technology, Nanning 530004, China)[2]

Abstract Hypercube is one of valuable interconnection networks. Aiming at the problem of high complexity of existing fault diagnosis algorithm in hypercube network, this paper proposed a concept of fault fan. The parallel depth-first search strategy algorithm is used to find the fault fan in hypercube networks, and the fault node of network is determined in order to replace or repair it, which provides a significant way for enhancing the reliability of network. In the end, the complexity of the proposed algorithm was analyzed. It is proved that the time complexity of the algorithm does not exceed $O(N)$, which is far better than the algorithm with more than square complexity.

Keywords Hypercube network, Fault diagnosis, Fault fans, System level diagnosis

1 引言

超级计算机在大规模科学计算、工程计算和云计算等领域发挥着巨大的作用。超级计算机以集群的方式把几万颗芯片构架到同一个系统,芯片间用专用链路互连以实现通信。在超级计算机互连网络设计的研究中,人们对环形(Ring)网络、树形(Tree)网络、网格(Mesh)网络、星形(Star)网络和环绕(Torus)网络等进行了广泛的研究[1-5]。然而,这些互连网络大多由于其固有的缺陷而无法满足现今巨量芯片超级计算机互连网络的设计需求。例如,环状网络虽接口数量很少,但存在通信延迟的缺陷;树形网络虽便于路由算法的设计,但缺乏容错能力。为了克服一般网络的诸多缺陷,人们提出了许多性能优越的网络拓扑结构来模拟超级计算机网络,其中超立方网络因具有优异的特性成为互连网络研究中的热点[6-7]。超立方网络具有优异的对称性、正则性、高容错性和递归生成性,可以方便地扩展网络上的芯片规模。超立方网络还具有良好的可嵌入性,使得其他成熟网络下的算法可以方便地移植并嵌入超立方网络中。最新的国际前沿研究成果表明,超立方网络将成为纳米计算机的基本构架[8],在各种互连网络中展现出独特的魅力。超立方网络由于结构明了、软件友好性等诸多优点被广泛应用于高性能计算机[9],例如 intel iPSC、Cray Origin2000、nCUBE 等并高性能计算机。

随着超级计算机芯片数量的扩增和内部结构复杂度的攀升,芯片出现故障的概率明显增大。实验研究表明,TFLOPS级超级计算机芯片的平均无故障时间为几十小时,而PFLOPS级超级计算机芯片的平均无故障时间为几小时。因此,互连网络的可靠性研究成为超级计算机领域的研究热点之一[10-14]。系统中芯片出现故障在所难免,并且一旦某些芯片发生故障将会给计算任务带来不可挽回的灾难,因此系统级故障诊断成为超级计算机研制的关键技术。系统级故障诊断完成了在系统中自动定位故障芯片的工作,使系统具有自诊断能力。诊断后,可以通过及时更换故障芯片或重新分配计算任务来保障超级计算机的可用性和易维护性。系统级故障诊断已被应用在许多机型上,例如著名的 APEmile 并行计算系统[15]。

在故障诊断问题的探究过程中,Preparata 等[16] 提出了

到稿日期:2018-03-29 返修日期:2018-06-04 本文受国家自然科学基金项目(61363002),广西自然科学基金项目(2016GXNSFAA380134)资助。
郭 杨(1992-),男,硕士生,主要研究方向为网络与并行计算、系统级故障诊断;梁家荣(1966-),博士,教授,CCF 会员,主要研究方向为网络与并行计算,E-mail:gxuliangjr@163.com(通信作者);刘 峰(1991-),男,硕士生,主要研究方向为网络与并行计算、系统级故障诊断;谢 敏(1974-),女,博士生,副教授,主要研究方向为网络与并行计算、网络控制。

PMC 诊断模型,其思想是系统内的处理器芯片相互测试,然后分析测试结果来自动定位故障芯片的位置。由于其测试规则简单、可操作性强,PMC 模型成为当今应用最广泛的诊断模型之一。PMC 诊断模型中用网络图 $G=(V,E)$ 描述一个多处理器系统,V 代表多处理器系统中的处理器集合,E 代表多处理器系统中的测试关系,有向边 $(u,v)\in E$ 代表节点 u 测试节点 v。在 PMC 诊断模型中,若节点 u 为正常节点,则节点 u 测试其他节点的结果可信。当被测节点 v 为故障节点时,测试结果为 1;当被测节点 v 为正常节点时,测试结果为 0。若节点 u 本身为故障节点,则节点 u 测试其他节点的结果不可信。节点 u 测试任何其他节点(无论故障与否)的结果是随机的,为 0 或 1。PMC 模型得到了广泛的研究[17-20],特别地,对于 N 个顶点 M 条边的 t-可诊断系统,Sullivan 等[21]提出了精确诊断算法,其时间复杂度为 $O(t^3+|E|)$;Dahbura 等[22]在 Meyer[23]的理论基础上提出了 N 个节点的 t-可诊断系统精确诊断算法,其时间复杂度为 $O(N^{2.5})$。为了提高系统的诊断效率,节省诊断时间,本文提出了一种基于 PMC 模型的分布式精确诊断并行算法,在节点个数为 N 的超立方网络中,该算法的并行时间复杂度不超过 $O(N)$。

2 定义定理

n 维超立方网络用 $Q_n=(V,E)$ 表示,其递归定义为:

$$Q_n=\begin{cases}K_2, & n=1\\ Q_{n-1}\cdot K_2, & n>1\end{cases}$$

其中,Q_n 节点的个数 $|V|$ 等于 2^n,Q_n 是 n-正则的,那么 Q_n 边的数量 $|E|$ 为 $2^{n-1}\cdot n$。$Q_1(=K_2)$,Q_2,Q_3 和 Q_4 如图 1 所示。

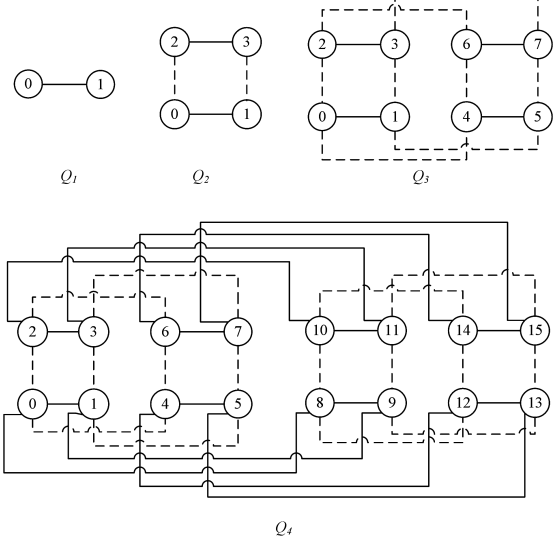

图 1 超立方体网络 $Q_1(=K_2)$,Q_2,Q_3 及 Q_4

Fig. 1 Hypercube systems of $Q_1(=K_2)$, Q_2, Q_3 and Q_4

定理 1[24] 用图 $G=(V,E)$ 表示包含 N 个处理器的系统 S,S 是 t-可诊断系统当且仅当以下 3 个条件成立:

1) $N\geqslant 2t+1$;

2) 对于 $\forall x\in V$,每个节点 x 至少被其他 t 个节点检测;

3) 对于任意满足 $0\leqslant p<t$ 的整数 p 以及满足 $|X|=n-2t+p$ 的 V 的任意子集 X,有 $|\Gamma X|>p$。

定义 1 图 $G=(V,E)$,$V(G)$ 的子集 S 使得 $G-S$ 的分支超过一个,那么集合 S 为 G 的一个分离集。使图 G 剔除 S 后不连通或仅剩一个节点的最小的 $|S|$ 值,被称为 G 的连通度,记为 $\kappa(G)$。若图 G 的 $\kappa(G)$ 至少为 k,则称图 G 为 k-连通的。

引理 1[25] 当 $n\geqslant 2$ 时,n 维超立方体网络 Q_n 的连通度 $\kappa(Q_n)=n$。

证明:对于 $\forall x\in V(Q_n)$,$N(x)$ 恰好是 Q_n 的一个分离集,所以 $\kappa(Q_n)\leqslant n$。下面对 n 进行数学归纳,证明每个分离集大于或等于 n。当 $n\leqslant 1$ 时,Q_n 是节点数量为 $n+1$ 的完全图,所以 $\kappa(Q_n)=n$。当 $n\leqslant 2$ 时,归纳假设,$\kappa(Q_{n-1})=n-1$,设 S 是 Q_n 的一个分离集,若 $Q_{n-1}-S$ 和 $Q'_{n-1}-S$ 皆连通,则 Q_n-S 也连通。若要使 Q_n-S 不连通,则 S 中应至少包含每条 Q_{n-1} 和 Q'_{n-1} 连线所浸润的一个节点,即 S 中节点个数至少为 2^{n-1},当 $n\geqslant 2$ 时,$|S|\geqslant n$。继续假设 Q_n-S 不连通,即 S 中至少含有 $n-1$ 个 Q_n 中的节点,若 $S\cap Q'_{n-1}=\varnothing$,显然,$Q'_{n-1}$ 连通,$Q_{n-1}-S$ 的节点皆与 $Q'_{n-1}-S$ 相连,即 Q_n-S 连通。那么 S 中必定具有一个来自 Q'_{n-1} 的节点,则 $\kappa(Q_n)\geqslant n$,故 $\kappa(Q_n)=n$。证明示意如图 2 所示。

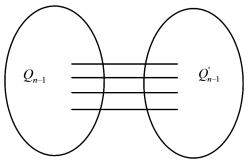

图 2 引理 1 证明图示

Fig. 2 Illustration of proving Lemma 1

定义 2 内部不相交定义为起点 x 至终点 y 的两条路径没有重叠的内部节点。

定义 3 给定一个节点 x 和一个集合 X,定义 x,X-扇为从节点 x 出发到 $\forall u\in F$ 的检测路径的集合,其中任意两条检测路径有且只有一个公共点 x。

引理 2(扩张引理) 若图 G 是 k-连通的,在图 G 中添加一个额外的顶点 y,把节点 y 与 G 中大于或等于 k 个节点直接邻接,由此得到的图记作 G',那么图 G' 依然是 k-连通的。

证明:设图 G' 的分离集为 S,若 $y\in S$,那么 $S-y$ 是图 G 的分离集,所以 $|S|\geqslant k+1$。若 $y\notin S$ 且 $S\supseteq N(y)$,那么 $|S|\geqslant k$。否则,$G'-S$ 其中一个分支包含 $N(y)-S$,这样,S 同样分离了 G,所以 $|S|\geqslant k$,图 G' 依然是 k-连通的,如图 3 所示。

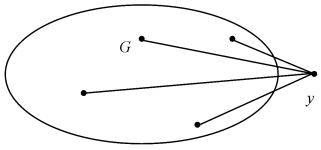

图 3 引理 2 证明图示

Fig. 3 Illustration of proving Lemma 2

定理 2[26](Menger 定理) 对于图 G 中的两个顶点 x 和 y,有 $(x,y)\notin E(G)$,那么 x,y-割的最小值 $\kappa(x,y)$ 等于两两内部不相交的 x,y-路径的最大数量 $\lambda(x,y)$。

定理 3 n 维超立方体网络 $Q_n=(V,E)$,对于 $\forall X\subset V$ 且 $|X|\leqslant n$,$\exists x\in V-X$,一定使得 Q_n 中存在 x,X-扇。

证明:Q_n 是 n-通图,在 Q_n 中添加一个节点 y,使该节点与 X 中的任意一个节点均相邻,如此操作 Q_n 之后的图用 Q_n' 表示,根据引理 2(扩张引理)得:Q_n' 还是一个 n-连通图。由 Menger 定理可知,能够在 Q_n' 中找到 $|X|$ 条 x,y-路径且两两内部不相交。从图 Q_n' 中移除节点 y,即可得到 Q_n 的一个 x,X-扇,证明示意如图 4 所示。

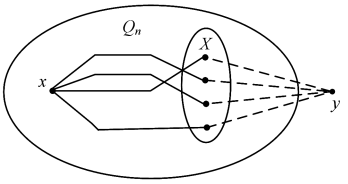

图 4　定理 3 证明示意图

Fig. 4　Illustration of proving Theorem 3

3　算法分析与实现

推论 1　n 维超立方体网络 $Q_n=(V,E)$ 中，故障节点个数不超过 n 时，从非故障节点 x 出发必能找到 $x_{\text{fault-free}}$，F-故障扇，此时集合 F 代表所有故障节点的集合。

证明：在定理 3 中，在取 $X=F$ 即可得结论，如图 5 所示。

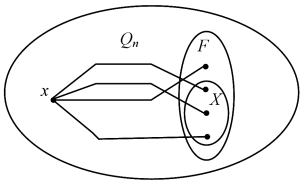

图 5　推论 1 证明图示

Fig. 5　Illustration of proving Corollary 1

那么，当 $X=F$ 时，即当 X 为故障集 F 时，我们就得到了 $x_{\text{fault-free}}$，F-故障扇，F 中的节点皆为故障点，F 外的节点皆正常。为寻找 $x_{\text{fault-free}}$，F-故障扇来确定故障节点集合，下面进行算法分析和实现。

超立方体网络 Q_n 在 PMC 模型下是 n-可诊断的（$n>2$）。事实上，Q_n（$n>2$）中节点总数为 $N=2^n$，$N=2^n>2n+1$（$n>2$）即定理 1 的第一个条件得到满足。注意到 Q_n 中任意节点 x 的度均为 n，即存在 n 个其他节点去检测 x，定理 1 第二个条件也得到满足。下面检验 Q_n 满足定理 1 中的第三个条件：对任意满足 $0\leqslant p<t$ 的整数 p 以及满足 $|X|=n-2t+p$ 的 V 的任意子集 X，有 $|V-X|=|N|-(|N|-2n+p)=2n-p>n$。如果 $|\Gamma X|\leqslant p$，那么 X 邻居节点的集合 $|N(X)|\leqslant p$。$N(X)$ 是 Q_n 的一个割集，其基数小于或等于 p，这与引理 1 的结论 $\kappa(Q_n)=n$ 相矛盾，故 Q_n 满足定理 1 第三个条件。由此可得，超立方体网络 Q_n 在 PMC 模型下是 n-可诊断的（$n>2$）。在 Q_n 中出现故障节点的个数不超过 n，远远小于 Q_n 总节点个数 $N=2^n$。也就是说，正常节点远多于故障节点，我们可以通过 Q_n 中的每个节点分布式地寻找 $x_{\text{fault-free}}$，F-故障扇。如果开始的节点 x 是正常节点，那么由 x 开始寻找到的 x，X-扇一定是 $x_{\text{fault-free}}$，F-故障扇；如果开始节点 x 是故障节点，那么由 x 开始寻找到的 x，X-扇一定不是 $x_{\text{fault-free}}$，F-故障扇。我们可以分布式地让每个节点去试图深度优先搜索 $x_{\text{fault-free}}$，F-故障扇，其优势在于可以均衡各个节点负载的效果。由于正常节点远多于故障节点，可以采用相信大多数的原则来确定最终故障集。

在系统级故障诊断开始前，首先要确定节点间的测试规则，然后通过通信链路收集汇总测试结果，称这个阶段为测试阶段。本文的测试策略是每一个节点 x 测试其所有邻居节点，并且节点 x 保存测试邻居节点结果的列表 l。下面是故障诊断具体的算法过程。

算法 1　$DFS\text{-}VISIT(Q_n,x)$诊断算法

输入：测试结束后的图 Q_n 和搜索起始节点 x

输出：一个潜在的 $x_{\text{fault-free}}$，F-故障扇

BEGIN

步骤 1　在节点 x 的测试结果列表中：

1）若 x.π=∅（x.π 为 x 的前驱），存在测试结果为 1 的记录，将被 x 测试且结果为 1 的节点 y 存入集合 F_x，并且将节点 y 和边（x，y）从 Q_n 中删除。

2）若 x.π≠∅，存在测试结果为 1 的记录，将被测试且结果为 1 的节点 y 存入集合 F_x，并且将颜色标记为黑色的节点从 Q_n 中删除。若起始节点已经没有邻居节点，或 Q_n 中除去输入的起始节点，其余节点的颜色均标记为黑色，则删除所有黑色节点，结束算法。

步骤 2　从节点 x 的测试结果列表 l 中选择一个测试结果为 0 且未被标记为黑色的节点 y（按照节点编码的升序选择），标记节点 y 为黑色，将 y 的前驱节点 y.π 赋值为 x。

步骤 3　执行 DFS-VISIT(Q_n，y)。

END

算法 2　$DFS(Q_n)$算法

输入：测试结束后的图 Q_n

输出：故障节点集合

BEGIN

FOR $x_i=0$ TO N−1，PAR-DO

　DFS-VISIT(Q_n，x_i)

END FOR

　输出出现次数最多的集合 F_x

END

本文算法的复杂性说明：在提出的算法中，对于任意起始节点，当找到了某个节点为潜在故障节点时，删除该路径，也就是说没有回溯过程；当起始节点出发的所有路径被找到，网络中的每个节点至多被搜索一次，在最坏情况下，单个计算分支的时间复杂度不会超过 $O(N)$。由于本文提出的算法是并行的，因此算法的时间复杂度不超过 $O(N)$，明显低于已有的经典算法的复杂度 $O(N^{2.5})$[23]。

例如在图 1 所示的 Q_4 网络系统中，若起始节点为 0 号节点且正常，2，5，9，14 号节点为故障节点，则算法过程如图 6 所示。其中，0 号节点为正常的起始节点，2，5，9，14 号节点用阴影代表加入到故障集中。若起始节点为 0 号节点且正常，只有 2 号节点为故障节点，则算法过程如图 7 所示。

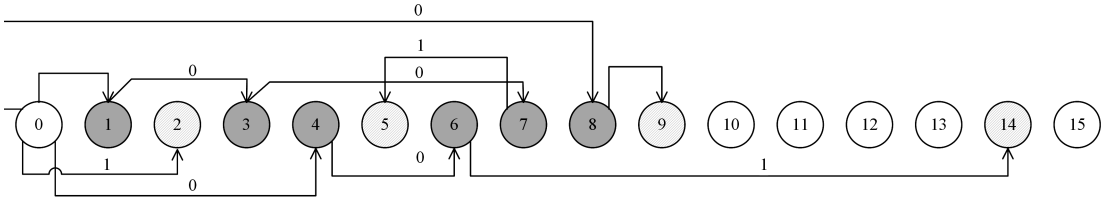

图 6　2，5，9 及 14 号节点为故障节点时的算法执行过程

Fig. 6　Algorithm execution process when nodes 2，5，9 and 14 are faulty nodes

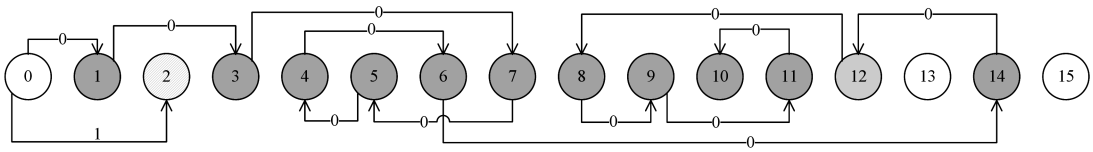

图 7 2 号节点为故障节点时的算法执行过程

Fig. 7 Algorithm execution process when node 2 is faulty node

结束语 本文以超立方体的基础定义为出发点，深入分析了其结构的独特性质，并引入故障扇的定义，证明了 n 维超立方体网络 $Q_n=(V,E)$ 中，对于 $\forall X\subset V$ 且 $|X|\leqslant n$，$\exists x\in V-X$，使 Q_n 中必然存在 x，X-扇结构。根据 x，X-扇的结构特点，设计了并行算法对每个节点进行深度优先搜索，以定位系统中故障节点的位置。由于本文提出的是精确诊断算法，在故障节点不超过 n 时，能够完全精确地诊断出故障节点。本文所提算法的复杂度仅为 $O(N)$，远低于经典算法的复杂度 $O(N^{2.5})$[23]。

参 考 文 献

[1] ZHOU S, LIN L, XU L, et al. The t/k-Diagnosability of Star Graph Networks[J]. IEEE Transactions on Computers, 2015, 64(2): 547-555.

[2] LIANG J R, ZHANG Q, LI H. Structural Properties and t/s-Diagnosis for Star Networks based on the PMC Model[J]. IEEE Access, 2017, 5(11): 26175-26183.

[3] CHENG E, LIPTÁK L. Linearly many faults in Cayley graphs generated by transposition trees[J]. Information Sciences, 2007, 177(22): 4877-4882.

[4] GUO L, HOU W, GUO P. Designs of 3D mesh and torus optical Network-on-Chips: Topology, optical router and routing module [J]. China Communications, 2017, 14(5): 17-29.

[5] LIN L, ZHOU S, XU L, et al. The Extra Connectivity and Conditional Diagnosability of Alternating Group Networks[J]. IEEE Transactions on Parallel & Distributed Systems, 2015, 26(8): 2352-2362.

[6] YE L C, LIANG J R, LIN H X. A fast pessimisticDiagnosis Algorithm for Hypercube-Like Networks under the Comparison Model[J]. IEEE Transactions on Computers, 2016, 65(9): 2884-2888.

[7] YE T L, HSIEH S Y. A Scalable Comparison-Based Diagnosis Algorithm for Hypercube-Like Networks [J]. IEEE Transactions on Reliability, 2013, 62(4): 789-799.

[8] LEE S C, HOOK L R. Logic and Computer Design in Nanospace [J]. IEEE Transactions on Computers, 2008, 57(7): 965-977.

[9] ROOSE D, BOMANS L, HEMPEL R. The Argonne/GMD macros in FORTRAN for portable parallel programming and their implementation on the Intel iPSC/2 [J]. Parallel Computing, 1990, 15(1): 119-132.

[10] ZHU Q. On conditional diagnosability and reliability of the BC networks[J]. Journal of Supercomputing, 2008, 45(2): 173-184.

[11] YE L C, LIANG J R. Five-Round Adaptive Diagnosis in Hamiltonian Networks[J]. IEEE Transactions on Parallel & Distributed Systems, 2015, 26(9): 2459-2464.

[12] TSAI C H, CHEN J C. Fault isolation and identification in general biswapped networks under the PMC diagnostic model [J]. Theoretical Computer Science, 2013, 501(3): 62-71.

[13] CHANG G Y, CHANG G J, CHEN G H. Diagnosabilities of Regular Networks[J]. IEEE Transactions on Parallel & Distributed Systems, 2005, 16(4): 314-323.

[14] LAI P L, TAN J J M, CHANG C P, et al. Conditional Diagnosability Measures for Large Multiprocessor Systems[J]. IEEE Transactions on Computers, 2005, 54(2): 165-175.

[15] CHESSA S, ERRICO W, MAESTRINI P, et al. Self-diagnostic tools of the APEmille parallel machine[J]. IEE Proceedings-Computers and Digital Techniques, 2002, 149(6): 273-279.

[16] PREPARATA F P, METZE G, CHIEN R T. On the Connection Assignment Problem of Diagnosable Systems[J]. IEEE Transactions on Electronic Computers, 1967, 16(6): 848-854.

[17] LIANG J, ZHANG Q. The t/s-diagnosability of Hypercube Networks under the PMC and Comparison Models[J]. IEEE Access, 2017, 5(1): 5340-5346.

[18] CHANG N W, HSIEH S Y. Conditional Diagnosability of (n,k)-Star Graphs under the PMC Model[J]. IEEE Transactions on Dependable & Secure Computing, 2016, 15(2): 207-216.

[19] CHANG N W, HSIEH S Y. Structural Properties and Conditional Diagnosability of Star Graphs by Using the PMC Model [J]. IEEE Transactions on Parallel & Distributed Systems, 2014, 25(11): 3002-3011.

[20] YUAN J, LIU A, MA X, et al. The g-Good-Neighbor Conditional Diagnosability of,-Ary,-Cubes under the PMC Modeland MM* Model[J]. IEEE Transactions on Parallel & Distributed Systems, 2014, 26(4): 1165-1177.

[21] SULLIVAN G F. An $O(t^3+|E|)$ fault identification algorithm for diagnosable systems[J]. IEEE Transactions on Computers, 1988, 37(4): 388-397.

[22] DAHBURA A T, MASSON G M. An $O(n^{2.5})$ Fault Identification Algorithm for Diagnosable Systems[J]. IEEE Transactions on Computers, 1984, 33(6): 486-492.

[23] MEYER G G L. A fault diagnosis algorithm for asymmetric modular architectures [J]. IEEE Transactions on Computers, 1981, 30(1): 81-83.

[24] HAKIMI S L, AMIN A T. Characterization of Connection Assignment of Diagnosable Systems [J]. IEEE Transactions on Computers, 1974, 23(1): 86-88.

[25] 徐俊明. 组合网络理论[M]. 北京：科学出版社，2007：96.

[26] AHARONI R. Menger's Theorem for Graphs Containing no Infinite Paths[J]. European Journal of Combinatorics, 1983, 4(3): 201-204.