SRC TR 89-17

An Optimal Diagnosis Structure for Distributed Systems

By

F.H.W. Shih
and
K. Nakajima

An Optimal Diagnosis Structure for Distributed Systems

by

Feng-Hsien Warren SHIH

IBM Research Division

T. J. Watson Research Center

Yorktown Heights, NY 10598

and

Kazuo NAKAJIMA†

Electrical Engineering Department,

Institute for Advanced Computer Studies,

and Systems Research Center

University of Maryland

College Park, MD 20742

† The work of this author was supported in part by the National Science Foundation Grants MIP-84-51510 and CDR-88-03012 (Engineering Research Centers Program), and in part by a grant from AT&T.

An Optimal Diagnosis Structure for Distributed Systems

by

Feng-Hsien Warren Shih and Kazuo Nakajima

Abstract

We propose a simple structure which provides optimal system-level fault diagnosis. Each unit of a system can diagnose itself using test outcomes generated within its associated structure. This property makes the structure suitable for systems in a distributed environment where no central processor exists for fault diagnosis. We present a simple diagnosis algorithm based on this structure. Furthermore, we show that the proposed structure always exists for every unit in a system whose graph connectivity is not less than the number of faulty units. We provide some examples of applying this unit-diagnosis approach to such systems as Boolean N-cube, generalized hypercube and DeBruijn networks. Distributed diagnosis can, therefore, be efficiently done for these fault-tolerant systems with results superior to those previously obtained.

I. Introduction

In the area of system-level fault diagnosis, Preparata, Metze, and Chien [18] introduced a graph theoretical model called the *PMC model* to deal with system-level faults in a multiprocessor system. For the purpose of fault diagnosis, the system is decomposed into independent subsystems or units. A connection assignment of test links among all units is established in such a way that each unit is tested by a subset of the other units. The outcome of each test is either "pass" or "fail". The outcomes from all test links are then used to diagnose the condition of each unit in the system. Each unit is identified as either fault-free or faulty. In the PMC model, it is assumed that the test outcome is reliable only if the testing unit is fault-free; otherwise it is unreliable regardless of the condition of the other unit involved in the test. Preparata *et al.* [18] defined a system to be *t-diagnosable* if all faulty units can be identified on the basis of any possible collection of test outcomes, provided that the number of faulty units does not exceed the given bound, t. Hakimi and Amin [8] and later Allan, Kameda, and Toida [1] gave necessary and sufficient conditions on the connection assignment of test links for a t-diagnosable system under the PMC model. Since then considerable research has been done on the characterization, diagnosis and diagnosability of t-diagnosable systems based on the PMC model [4-6,11,15,16,20,21].

In most diagnosis algorithms developed for t-diagnosable systems [5,6,11,16,21], the existence of a central processor that monitors and analyzes test outcomes from all test links is assumed. This assumption is unrealistic in the distributed environment as pointed out in [13,14,17]. Each individual unit has to carry out some portion of the diagnosis task in a distributed system. Kuhl and Reddy [13,14] gave a necessary and sufficient condition for a system to be t-diagnosable in a distributed manner and diagnosis algorithms were developed for such systems [10,13,14]. Very recently, Kreutzer and Hakimi [12] considered this distributed diagnosis problem from the point of view of a system user who does

not have a testing capability. By modifying Kuhl and Reddy's algorithms [13,14], they developed a distributed diagnosis algorithm. However, in the system model of Kuhl and Reddy [13,14], the communication of all test links occurs in both directions, which may not always be true in some system designs [19]. Pradhan and Reddy [17] proposed a distributed diagnosis algorithm for DeBruijn communication networks. Their algorithm distributes the diagnosis task to a set of so-called root-nodes defined in their test digraph. Each root-node examines all test outcomes from its neighbors, and proceeds to perform the diagnosis. However, this algorithm is only applicable to a special type of system topology and can identify only a half of the number of faulty units that can be identified in some other way. Furthermore, in their system model, the direction of test outcome transmission is opposite to that of the test performed through each test link and test outcomes are transmitted through many units and test links before diagnosis.

In this paper, we propose a simple diagnosis structure for distributed systems. Mallela [15] considered a special version of this structure, and assuming its existence, developed simple diagnosis algorithms. Chwa and Hakimi [4] also considered the same structure as Mallela's [15] for the diagnosis of a special class of systems in which such a structure always exists for each unit. We show that the proposed structure always exists for every unit in a large class of systems, where their graph connectivity is not less than the number of faulty units. Many distributed systems fall into this class. Furthermore, during the diagnosis process of such systems, test outcomes are transmitted in the same direction as the test link and through at most one test link. Each unit can diagnose itself based on test outcomes generated within the structure associated with it.

In the next section we give basic model assumptions and definitions. The unit-diagnosis structure and a simple diagnosis algorithm along with its complexity analysis are presented in Section III. We also describe how to construct such a structure for each

unit of a system. Relationship between the existence of the structure and the connectivity of a system is established in Section IV. Finally, applications of the unit-diagnosis approach to such fault-tolerant networks as Boolean N-cube [2], generalized hypercube [3], and DeBruijn networks [7,17] and their variation [19] are provided in Section V.

II. Preliminaries

Consider a system S consisting of a set of n units $U = \{u_1, u_2, \ldots, u_n\}$ and a connection assignment of test links, denoted by T. Each test is performed on a unit by a subset of the other units. The system S can thus be represented by a digraph $G = (U, T)$, which we call a *system digraph*, where U is the set of vertices of G and $T = \{(u_i, u_j) | u_i, u_j \in U$ and u_i is assigned to test $u_j\}$ is the set of directed edges in G. Each test outcome is either "pass" (fault-free) or "fail" (faulty) and can be expressed as a binary weight $w(u_i, u_j)$ of edge $(u_i, u_j) \in T$. In particular, $w(u_i, u_j) = 0$ if u_i judges u_j to be fault-free and $w(u_i, u_j) = 1$, otherwise. In the PMC model, the test outcome $w(u_i, u_j)$ is reliable only if u_i is a fault-free unit. The set of all test outcomes of S is called the *syndrome* of S.

Definition 1 [18]: A system S is said to be *t-diagnosable*, if given a syndrome of S, all faulty units in S can correctly be identified, provided that the number of faulty units present does not exceed t.

The above definition implies that a complete syndrome of S should be available to a central monitor which performs the diagnosis task to identify the faulty units in the system. If only a partial syndrome of S is available, the system may not be t-diagnosable.

In the distributed environment, no central monitor processor exists in the system. The diagnosis task needs to be performed based on a partial syndrome available to each unit. The availability of such a syndrome is usually limited by the number of communication paths or test links assigned to that unit.

In the distributed system model proposed in [13,14,17], it is assumed that a bidirec-

tional communication path for the test outcome transmission between two units is always available even when the test link between them is unidirectional. The diagnosis algorithms developed are also based on the assumption that unit u receives diagnostic information from unit v only after the receiver u judges the sender v to be fault-free. In this paper, we make the following assumptions:

(1) For each test link $(u, v) \in T$, the test outcome $w(u, v)$ is available to unit v.

(2) The test outcome $w(u, v)$ is available to unit x only if $(v, x) \in T$.

The above assumptions are natural as the directions of communication paths coincide with those of test links defined in the system digraph. The partial syndrome available to each unit is defined in such a way that a unit is aware of its own diagnosis outcomes when it is tested by others and can broadcast them to the units it tests. We assume that no fault occurs on the channels for data communication. Note that it is possible for a test outcome to be transmitted to more than one destination along test links. However, the unit-diagnosis structure proposed in the next section does not depend on this property.

Before we proceed, we give some graph-theoretical definitions and notation used in this paper. Let $G = (V, E)$ be a digraph. For a subset $W \subset V$, let $E(W) = \{(v, w) \in E | v,\ w \in W\}$. A *subgraph* of G is a graph (W, F) such that $W \subset V$ and $F \subset E(W)$. For subsets $X,\ Y \subset V$ with $X \cap Y = \phi$, let $E(X, Y) = \{(x, y) \in E | x \in X,\ y \in Y\}$. G is called a *bipartite graph* if $V = X \cup Y,\ X \cap Y = \phi$, and $E \subset \{(x, y),\ (y, x) | x \in X,\ y \in Y\}$. A subset $M \subset E$ is called a *matching* of G if no two edges of M share a common vertex, and a vertex is said to be *matched* with respect to M if it is an endpoint of some edge in M. A matching of maximum cardinality is called a *maximum matching.* A *dipath* of length $k-1$ from v to w is a sequence of k distinct vertices $[v = v_1, v_2, \ldots,\ v_k = w]$ such that $(v_i, v_{i+1}) \in E$ for $i = 1, 2, \ldots,\ k-1$, and it is said to *end at vertex* w. A dipath is said to be *maximal* if it cannot be made longer by adding a new vertex to it. Dipaths are said to be *vertex disjoint*

if no two of them share a common vertex. G is called a *bundle of dipaths* of length k *tied at vertex* v if all dipaths in G end at the vertex v and the removal of v and its incident edges from G results in a graph consisting of vertex disjoint dipaths of length at most $k-1$. G is said to be *strongly connected* if for every pair of vertices $v, w \in V$, there exist dipaths from v to w and from w to v. The *connectivity* of G, denoted by $k(G)$, is the minimum number of vertices whose removal yields a digraph that is not strongly connected. For a vertex $v \in V$, let $\Gamma^{-1}(v) = \{u \in V | (u,v) \in E\}$, $d_{in}(v) = |\Gamma^{-1}(v)|$, $\Gamma(v) = \{w \in V | (v,w) \in E\}$, and $d_{out}(v) = |\Gamma(v)|$. For a subset $W \subset V$, let $\Gamma^{-1}(W) = \{u \in V - W | (u,v) \in E, v \in W\}$.

III. Unit-diagnosis Structure and Algorithm

In this section, we propose a simple structure to be used for the diagnosis of each unit of a system. We then describe how to find such a structure. Finally we present an algorithm to determine the condition of a unit, using test outcomes available within this structure.

Definition 2: A *unit-diagnosis digraph* for unit u in a system S, denoted by $S(u;s)$, is a subgraph of its system digraph $G = (U,T)$ that is a bundle of dipaths of length one or two tied at u and consists of $2s+1$ vertices. The unit u is called the *base unit* of $S(u;s)$.

As shown in Figure 1, $S(u;s)$ consists of the base unit u and two sets of units $U_a = \{a_1, a_2, \ldots, a_r\}$ and $U_b = \{b_1, b_2, \ldots, b_{2s-r}\}$ with $0 \leq r \leq s$, where $U_b \subset \Gamma^{-1}(u)$ and $(a_i, b_i) \in T$ for $i = 1, 2, \ldots, r$. Note that the number of edges in $S(u;s)$ is even. The value of s determines the maximum number of faulty units allowed in $S(u;s)$ such that the base unit u can still diagnose its own condition, as shown in the proof of the following theorem.

Theorem 1: Let $S(u;s)$ be a unit-diagnosis digraph for unit u of a system S. If test outcomes from all $2s$ test links in $S(u;s)$ are known, the condition of the base unit u can always be determined, provided that the number of faulty units in $S(u;s)$ does not exceed s.

Proof: Let h be the number of tests between units in U_a and those in U_b whose outcomes are "fail". Note that at least one of the units a_i and b_i is faulty when $w(a_i, b_i) = 1$. Thus, if $h = s$, unit u must be fault-free. If $h < s$, by removing from $S(u; s)$ the $2h$ vertices and h edges involved in those tests, we obtain an $S(u; s - h)$ digraph which contains at most $s - h$ faulty units. By reindexing the remaining units, as shown in Figure 2, we can assume that $w(a_i, b_i) = 0$ for $i = 1, \ldots, r$; $w(b_i, u) = 0$ for $i = 1, \ldots, p$ and $w(b_i, u) = 1$ for $i = p + 1, \ldots, r$; $w(b_i, u) = 0$ for $i = r + 1, \ldots, q$ and $w(b_i, u) = 1$ for $i = q + 1, \ldots, 2(s - h) - r$. If u is fault-free, the number of faulty units would be $f_0 = 2(r - p) + 2(s - h) - r - q = 2(s - h) - 2p - q + r$, and if u is faulty, it would be $f_1 = 1 + 2p + q - r$. Note that $f_0 + f_1 = 2(s - h) + 1$. Since $S(u; s - h)$ contains at most $s - h$ faulty units, $f_0 \leq s - h$ and $f_1 \leq s - h$, and hence $f_0 + f_1 \leq 2(s - h)$. Therefore, exactly one of the situations, either $f_0 > s - h$ or $f_1 > s - h$, must occur and thus the condition of the base unit u is always determined. □

Definition 3: The maximum value of s for which $S(u; s)$ exists is called the *unit-diagnosability* of u and is denoted by s_u. A unit-diagnosis digraph $S(u; s_u)$ is said to be *optimal.*

It should be noted that $s_u \leq d_{in}(u)$ for any $u \in U$. According to the assumptions on test outcome transmissions given in the previous section, $2s$ test outcomes in $S(u; s)$ form a partial syndrome of S that is available to the base unit u. The above theorem implies that using this partial syndrome, each unit u can determine its own condition if the number of faulty units in S does not exceed the unit-diagnosability s_u. We now introduce a class of diagnosable systems.

Definition 4: A system S is said to be *unit-based t-diagnosable*, or in short *u-t-diagnosable*, if given a partial syndrome available to each unit within its optimal unit-diagnosis digraph, all faulty units in S can correctly be identified, provided that the number

of faulty units present does not exceed t.

Note that in a u-t-diagnosable system $s_u \geq t$ for every unit u of the system. By definition, a u-t-diagnosable system is t-diagnosable. However, the converse is not always true. For example, consider the system shown in Figure 3 (taken from Figure 1 of [5]). This is a 4-diagnosable system. Figure 4 shows optimal unit-diagnosis digraphs for all units in the system. Since the unit-diagnosabilities are three or four, the system is u-3-diagnosable but not u-4-diagnosable.

Definition 5: The *unit-based diagnosability* of a system S is the maximum value of t for which S is u-t-diagnosable.

Given the system digraph $G = (U, T)$ of S, its unit-based diagnosability is determined by finding the minimum value of the unit-diagnosability among all units. We briefly describe how to compute the unit-diagnosability s_u and construct an optimal unit-diagnosis digraph $S(u; s_u)$ for a given unit $u \in U$.

We first extract from G a bipartite digraph $H = (X \cup Y, T(X, Y))$ such that $Y = \Gamma^{-1}(u)$ and $X = \Gamma^{-1}(Y) \cap (U - Y - \{u\})$. We then find a maximum matching M of H. Let X' be the set of the matched vertices in X. If $|Y|+|X'|$ is even, set $s_u = (|Y|+|X'|)/2$, $U_a = X'$ and $U_b = Y$. Otherwise, set $s_u = (|Y| + |X'| - 1)/2$, $U_a = X' - \{x'\}$ with x' being any vertex in X' if $X' \neq \phi$ and $U_b = Y$, or if $X' = \phi$, $U_a = \phi$ and $U_b = Y - \{y'\}$ with y' being any vertex in Y.

Using Hopcroft and Karp's algorithm [9], we can find in $O(n^{2.5})$ time a maximum matching of the bipartite digraph H. Therefore, it takes $O(n^{3.5})$ time to construct all optimal unit-diagnosis digraphs and determine the unit-based diagnosability of S.

Let $G = (U, T)$ be the system digraph of a u-t-diagnosable system S. Since $s_u \geq t$ for each unit u in U, we can always construct in $O(n^{2.5})$ time a unit diagnosis digraph $S(u; t)$ for u. As defined before, let $U_a = \{a_1, a_2, \ldots, a_r\}$ and $U_b = \{b_1, b_2, \ldots, b_{2t-r}\}$ be the sets

of units in $S(u;t)$, where $0 \leq r \leq t$. The following algorithm determines the condition of unit u.

Algorithm 1:

Step 0:

Set $i \leftarrow 1$, $q \leftarrow t$, $f_0 \leftarrow 0$, and $f_1 \leftarrow 1$.

Step 1:

If $f_0 > q$, stop. u is a faulty unit.

If $f_1 > q$, stop. u is a fault-free unit.

If $i > r$, set $\ell \leftarrow 1$, and go to Step 2.

Otherwise, seek the outcome of test (a_i, b_i).

If $w(a_i, b_i) = 1$, set $q \leftarrow q - 1$ and $i \leftarrow i + 1$, and return to Step 1.

If $w(a_i, b_i) = 0$, set $\ell \leftarrow 2$.

Step 2:

Seek the outcome of test (b_i, u).

If $w(b_i, u) = 1$, set $f_0 \leftarrow f_0 + \ell$.

If $w(b_i, u) = 0$, set $f_1 \leftarrow f_1 + \ell$.

Set $i \leftarrow i + 1$, and go to Step 1. □

It is not difficult to see that in some cases Algorithm 1 examines only t or $t + 1$ test outcomes to diagnose a base unit. In general we have the following theorem.

Theorem 2: Algorithm 1 correctly identifies the condition of a base unit after examining $t + k$ test outcomes with $0 \leq k \leq t$. Furthermore, if $k \geq 1$, k or $k - 1$ other units can be identified as faulty.

Proof: The termination and the correctness of the algorithm follow from the proof of Theorem 1. Let c be the total number of test outcomes to be examined during the execution

of Algorithm 1. There are three ways to lead the algorithm towards its termination. Each time $w(a_i, b_i) = 1$, the value of q is decremented by one. Let h be the number of such tests (a_i, b_i) that $w(a_i, b_i) = 1$. Clearly $h = t - q$. Each time $w(a_i, b_i) = 0$ and $w(b_i, u) = 1$ (respectively, $w(b_i, u) = 0$), the value of f_0 (respectively, f_1) is incremented by two. When $i > r$, the value of f_0 or f_1 is incremented by one each time $w(b_i, u)$ is examined. Therefore, $c = h + f_0 + f_1 - 1$ (note that the initial value of $f_1 = 1$).

Since the algorithm terminates when $f_0 > q$ or $f_1 > q$, clearly $c \geq t$. Furthermore, $c \leq 2t$ from the proof of Theorem 1. When the termination occurs after $T = t+k = q+h+k$ test outcomes are examined, f_0 or $f_1 = q + 1$ or $q + 2$. If $f_0 > q$, the base unit is faulty and hence all the units a_i and b_i that contribute to f_1 are also faulty. Likewise, if $f_1 > q$, the base unit is fault-free and all the units a_i and b_i that contribute to f_0 are faulty. By substituting $q+1$ or $q+2$ for f_0 (respectively, f_1) in the equation $h+f_0+f_1-1 = h+q+k$, we obtain that f_1 (respectively, f_0) $= k$ or $k-1$. Therefore, at the termination with $k \geq 1$, k or $k-1$ other units are diagnosed as faulty. □

Theorem 2 implies that if $k \geq 1$, $t+k$ test outcomes are examined to diagnose a total of k or $k+1$ units. Using Algorithm 1, all faulty units are identified after examining $O(nt)$ test outcomes. Note that the number of test links required for a system to be t-diagnosable is at least nt [8,18]. Thus, the above algorithm is optimal in this sense.

IV. Existence of Unit-diagnosis Digraphs

Let $G = (U, T)$ be the system digraph of S with n units. It has been shown [8] that if $n \geq 2t + 1$ and $k(G) \geq t$, S is t-diagnosable. In this section we present a stronger result by showing that if the same conditions hold, S is u-t-diagnosable. We first consider a class of t-diagnosable systems in which no two units test each other. Let $S(u; a, b)$ be another representation of $S(u; t)$ with $a = |U_a|$, $b = |U_b|$ and $a + b = 2t$. Note that a digraph $S(u; t, t)$ is the structure considered by Mallela [15] and Chwa and Hakimi [4].

Theorem 3: Consider a system S of n units in which no two units test each other. Let $G = (U, T)$ be its system digraph. If $n \geq 2t + 1$ and $k(G) \geq t$, then for every unit $u \in U$, there exists a digraph $S(u; t, t)$.

Proof: Let $Y = \Gamma^{-1}(u)$ and $X = \Gamma^{-1}(Y) \cap (U - Y - \{u\})$. Since $k(G) \geq t$, $|Y| \geq t$. Since unit u does not test any unit in Y, all dipaths from u to any unit in Y must pass through vertices in X and hence $|X| \geq t$ due to the connectivity assumption. Let M be a maximum matching of the bipartite digraph $H = (X \cup Y,\ T(X, Y))$, and X' and Y' be the sets of the matched vertices in X and Y, respectively. Suppose that $|M| < t$. Since $|X'| = |M| < t$ and $|X| \geq t$, there is a vertex $w \in X - X'$. However, there is no vertex z in $Y - Y'$ such that $(w, z) \in T$, since M is a maximum matching. Thus, the removal of the vertices in Y' eliminates all dipaths from w to u, which contradicts the assumption that $k(G) \geq t$. Therefore, $|M| \geq t$ and there exists an $S(u; t, t)$ digraph with $U_a = X'$ and $U_b = Y'$. □

We now consider systems in which units are allowed to test each other.

Theorem 4: Let S be a system with n units and $G = (U, T)$ be its system digraph. If $n \geq 2t + 1$ and $k(G) \geq t$, then for every unit $u \in U$, there exists a unit-diagnosis digraph $S(u; t)$.

Proof: Let $Y = \Gamma^{-1}(u)$, $X = \Gamma^{-1}(Y) \cap (U - Y - \{u\})$, and $V = U - X - Y - \{u\}$. Since $k(G) \geq t$, $|Y| \geq t$. There are two cases to be considered separately.

Case 1: $V \neq \phi$.

Let $v \in V$. Since $k(G) \geq t$, there are at least t vertex disjoint dipaths from v to u that must pass through vertices in X and Y. Therefore, there exists an $S(u; t, t)$ digraph.

Case 2: $V = \phi$.

Let $|Y| = t + k$ for some integer $k \geq 0$. Since $n \geq 2t + 1$, $|X| \geq t - k$. Let M be a maximum matching of the bipartite digraph $H = (X \cup Y,\ T(X, Y))$. Let X' and Y' be

the sets of all matched vertices in X and Y, respectively. If $|M| < t-k$, $X - X' \neq \phi$ and all dipaths from a vertex $w \in X - X'$ to u are destroyed by removing less than t vertices of Y', a contradiction. Therefore, $|M| \geq t-k$ and there exists an $S(u;t)$ digraph with $U_b = Y$ and $U_a \subset X'$ such that $|U_a| + |U_b| = 2t$. Note that this argument holds up to $|M| = t$ if $|X| \geq t$ and otherwise $|Y| > t$ and $X' = X$. □

Whether or not units are allowed to test each other in a system, we derive the following corollary from Theorems 1, 3, and 4.

Corollary 1: Let S be a system with n units and G be its system digraph. S is u-t-diagnosable if $n \geq 2t+1$ and $k(G) \geq t$.

The above corollary implies that the unit-based diagnosability of a system is at least as large as the connectivity $k(G)$ of its system digraph G. Since the degree of fault-tolerance of a system is defined to be $k(G)-1$ [17,19], the diagnosis approach proposed in this paper is optimal in the sense that the total number of faulty units that can be identified is lower bounded by the maximum fault-tolerance of the system plus one.

We now show that the converse of Corollary 1 is not always true. Consider the system S_1 shown in Figure 5. Clearly, the connectivity of its system digraph is 0. However, it is clear that an $S(u;2)$ digraph exists for every unit u in S_1. Therefore, S_1 is u-2-diagnosable. In fact, S_1 is a $D(10,2,\{2,4\})$ system, where in general a $D(n,t,X)$ system is defined as follows. $X = \{x_1, x_2, \ldots, x_t\}$ is a set of t integers with $1 \leq x_1 < x_2 \ldots < x_t < n/2$. A $D(n,t,X)$ system consists of a set of n units $U = \{u_1, u_2, \ldots, u_n\}$ and a set of test links T such that $(u_i, u_j) \in T$ if and only if $(i-j) \bmod n \in X$. As shown by Chwa and Hakimi [4], an $S(u;t,t)$ digraph always exists for every unit u in such a system. Thus, a $D(n,t,X)$ system is u-t-diagnosable.

V. Applications of Unit-diagnosis Digraph

Many computer networks with different topologies have been proposed for distributed

systems, such as Boolean N-cube [2], generalized hypercube [3], DeBruijn network [7,17] and its variation [19]. The following examples show that a unit-diagnosis digraph $S(u;t,t)$ exists for each unit u in all such systems. Improvements on diagnosis by our approach over those previously reported are also pointed out.

Example 1: A Boolean N-cube system [2] consists of $n = 2^N$ processors identified by N-bit binary numbers from 0 to $2^N - 1$. Two processors can communicate with each other if their N-bit representations differ in only one position. Figure 6(a) shows the connection assignment of processor links for a Boolean 3-cube network. If two directly connected processors in a Boolean N-cube can completely test each other, it is N-diagnosable since its connectivity is N. Armstrong and Gray [2] gave a rather complex diagnosis algorithm to be carried out by a central diagnosis processor. Since the connectivity is N, a Boolean N-cube system is also u-N-diagnosable. In Figure 6(b), we show an optimal unit-diagnosis digraph $S(u;N,N)$ for every unit u, where $u = [x_1, \ldots, x_n]$ is the N-bit binary representation with $x_i = 0$ or 1 and x_i' denotes the binary complement of x_i. Since such a unit-diagnosis digraph always exists, each unit can efficiently diagnose itself.

Example 2: The generalized hypercube (GHC) structure proposed by Bhuyan and Agrawal [3] consists of $n = m_r * \ldots * m_2 * m_1$ processors with each processor represented by $u = [x_1^0, \ldots, x_r^0]$, where $x_i^0 \in \{0, 1, \ldots, m_r\}$. Let x_i^j, $j = 1, \ldots, m_r - 1$, be those $m_r - 1$ values that are different from x_i^0 and $M = m_1 + m_2 + \ldots + m_r - r$. Two processors can communicate with each other if their r-bit representations differ in only one position. This means that a unit always has direct links to M other units in the system. Figure 7(a) shows the connection assignment of processor links for a $4 * 3 * 2$ GHC system. The Boolean N-cube system given in Example 1 is a special case of this GHC system. Since the connectivity of a GHC structure is M, it is u-M-diagnosable. We show the general $S(u;M,M)$ digraph in Figure 7(b). Thus, efficient self-diagnosis of a GHC system can be

performed.

Example 3: Various topologies have been proposed for DeBruijn type communication networks [7,17,19]. The number of processors (or nodes) in such a system is assumed to be r^m. Accordingly, each node has radix-r representation. Pradhan and Reddy [17] gave an undirected graph topology for the network and showed that there are exactly $n - r^2$ nodes of degree $2r$, r nodes of degree $2r-2$, and $r^2 - r$ nodes of degree $2r-1$ depending on the representation of each node. This can be seen from the 16-node network shown in Figure 8(a) with $m = 4$ and $r = 2$. They also presented a distributed diagnosis algorithm which can identify up to $r-1$ faulty nodes in $O(nr)$ time. Bidirectional communication links are assumed between pairs of processors in their diagnosis. Since the connectivity of this topology is shown to be $2r-2$ by Esfahanian and Hakimi [7], the system is $u-(2r-2)$-diagnosable. In Figure 8(b), we show the three general $S(u; 2r-2, 2r-2)$ digraphs for the units in the network with different types of radix-r representations. Thus our approach can identify up to $2r-2$ faulty nodes in $O(nr)$ time, which doubles the maximum number of faulty nodes that can be diagnosed by Pradhan and Reddy's approach [17]. For a digraph topology of DeBruijn networks, Sengupta et al. [19] proposed an architecture using unidirectional links. The connection assignment of processor links is defined in such a way that for each node u, $d_{in}(u) = d_{out}(u) = r$. An 8-node network of this type is shown in Figure 9(a) with $m = 3$ and $r = 2$. The connectivity of the topology is shown to be r [19], the maximum possible connectivity, and thus the network is called optimally fault-tolerant. Clearly it is a u-r-diagnosable system. We show two types of $S(u; r, r)$ digraphs for this topology in Figure 9(b).

VI. Conclusions

We have presented a simple unit-diagnosis structure and a diagnosis algorithm. They were shown to be an optimal approach to system-level fault diagnosis for distributed sys-

tems under the PMC model. This unit-based diagnosis approach resolves the diagnosis problem for a system user, which was recently posed by Kreutzer and Hakimi [12], by providing correct and complete diagnosis results. We have introduced a class of u-t-diagnosable systems. We have shown that an optimal unit-diagnosis digraph is constructed in $O(n^{2.5})$ time for each unit in a system of n units and thus its unit-based diagnosability is determined in $O(n^{3.5})$ time. We have then shown that the diagnosability is lower bounded by the connectivity of the system graph. This implies that as the degree of fault-tolerance of a system increases, so does its diagnosability. Since the complexity of fault diagnosis depends only on the number of test links in the system, a topology with the diagnosability being equal to the maximum fault-tolerance of the network plus one can be viewed as optimally diagnosable. The network examples discussed in this paper are all optimal in this sense. The data communication assumptions used in the diagnosis are natural and were shown to be applicable to various fault-tolerant network architectures. The results obtained here are superior to those previously known. Finally, with possible design of a fault-free diagnosis hardware [22], the implementation of our unit-based diagnosis approach may become practical for some fault-tolerant networks.

References

[1] F. J. Allan, T. Kameda, and S. Toida, "An approach to the diagnosability analysis of a system," *IEEE Trans. Comput.*, vol. C-24, pp. 1040-1042, Oct. 1975.

[2] J. R. Armstrong and F. G. Gray, "Fault diagnosis in boolean n-cube array of microprocessors," *IEEE Trans. Comput.*, vol. C-30, pp. 590-596, Aug. 1981.

[3] L. N. Bhuyan and D. P. Agrawal, "Generalized hypercube and hyperbus structures for a computer network," *IEEE Trans. Comput.*, vol. C-33, pp. 323-333, Apr. 1984.

[4] K. Y. Chwa and S. L. Hakimi, "On fault identification in diagnosable systems," *IEEE Trans. Comput.*, vol. C-30, pp. 414-422, June 1981.

[5] A. T. Dahbura and G. M. Masson, "An $O(n^{2.5})$ fault identification algorithm for diagnosable systems," *IEEE Trans. Comput.*, vol. C-33, pp. 486-492, June 1984.

[6] A. T. Dahbura, G. M. Masson and C. L. Yang, "Self-implicating structures for diagnosable systems," *IEEE Trans. Comput.*, vol. C-34, pp. 718-723, Aug. 1985.

[7] A. H. Esfahanian and S. L. Hakimi, "Fault-tolerant routing in DeBruijn communication networks," *IEEE Trans. Comput.*, vol. C-34, pp. 777-789, Sept. 1985.

[8] S. L. Hakimi and A. T. Amin, "Characterization of connection assignment of diagnosable systems," *IEEE Trans. Comput.*, vol. C-23, pp. 86-88, Jan. 1974.

[9] J. E. Hopcroft and R. M. Karp, "An $O(n^{2.5})$ algorithm for maximum matchings in bipartite graphs," *SIAM J. Comput.*, vol. 2, Dec. 1973.

[10] S. H. Hosseini, J. G. Kuhl, and S. M. Reddy, "On self-fault diagnosis of the distributed systems," *IEEE Trans. Comput.*, vol. C-37, pp. 248-251, Feb. 1988.

[11] T. Kameda, S. Toida, and F. J. Allan, "A diagnosing algorithm for networks," *Informat. Contr.*, vol. 29, pp. 141-148, 1975.

[12] S. E. Kreutzer and S. L. Hakimi, "Distributed diagnosis and the system user," *IEEE Trans. Comput.*, vol. C-37, pp. 71-78, Jan. 1988.

[13] J. G. Kuhl and S. M. Reddy, "Distributed fault-tolerance for large multiprocessor systems," in *Proc. 7th Annu. Symp. Comput. Arch.*, May 1980, pp. 23-30.

[14] J. G. Kuhl and S. M. Reddy, "Fault diagnosis in fully distributed systems," in *Proc. 11th Int. Symp. Fault-Tolerant Comput.*, June 1981, pp. 100-105.

[15] S. Mallela, "On diagnosable systems with simple algorithms," in *Proc. 17th Annu. Conf. on Informat. Sci. Syst.*, Princeton, NJ, Mar. 1983, pp. 545-549.

[16] G. G. L. Meyer and G. M. Masson, "An efficient fault diagnosis algorithm for symmetric multiple processor architecture," *IEEE Trans. Comput.*, vol. C-27, pp. 1059-1063, Nov. 1978.

[17] D. K. Pradhan and S. M. Reddy, "A fault-tolerant communication architecture for distributed systems," *IEEE Trans. Comput.*, vol. C-31, pp. 863-870, Sept. 1982.

[18] F. P. Preparata, G. Metze, and R. T. Chien, "On the connection assignment problem of diagnosable systems," *IEEE Trans. Electron. Comput.*, vol. EC-16, pp. 848-854, Dec. 1967.

[19] A. Sengupta, A. Sen and S. Bandyopadhyay, "On an optimal fault-tolerant multiprocessor network architecture," *IEEE Trans. Comput.*, vol. C-36, pp. 619-623, May 1987.

[20] G. Sullivan, "A polynomial time algorithm for fault diagnosability," in *Proc. 25th Annu. Symp. Found. of Comput. Sci.*, Oct. 1984, pp. 148-155.

[21] G. Sullivan, "An $O(t^3 + |E|)$ fault identification algorithm for diagnosable systems," in *Proc. 21st Annu. Conf. on Informat. Sci. Syst.*, Baltimore, MD, Mar. 1987, pp. 203-207.

[22] F.-H. W. Shih and K. Nakajima, "A hardware accelerator for real-time system diagnosis," Tech. Report RC 13944, IBM Research Division, Yorktown Heights, NY, April 1988.

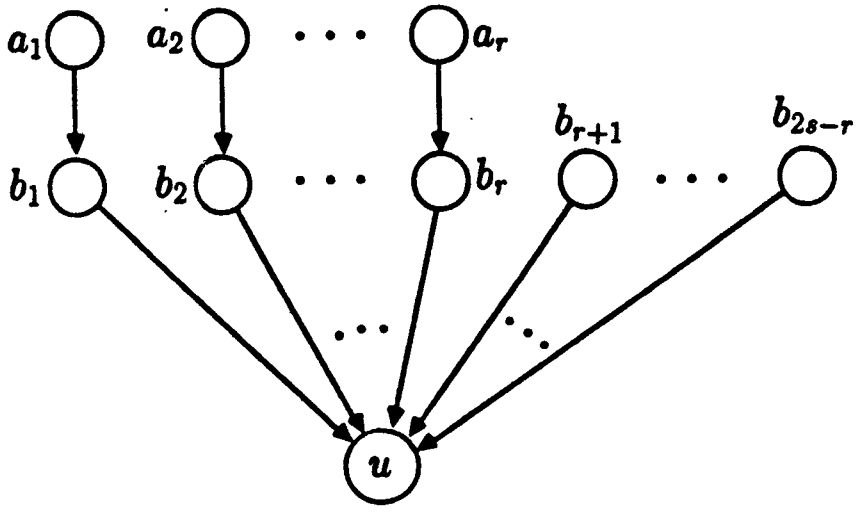


Fig. 1. A unit-diagnosis digraph $S(u; s)$.

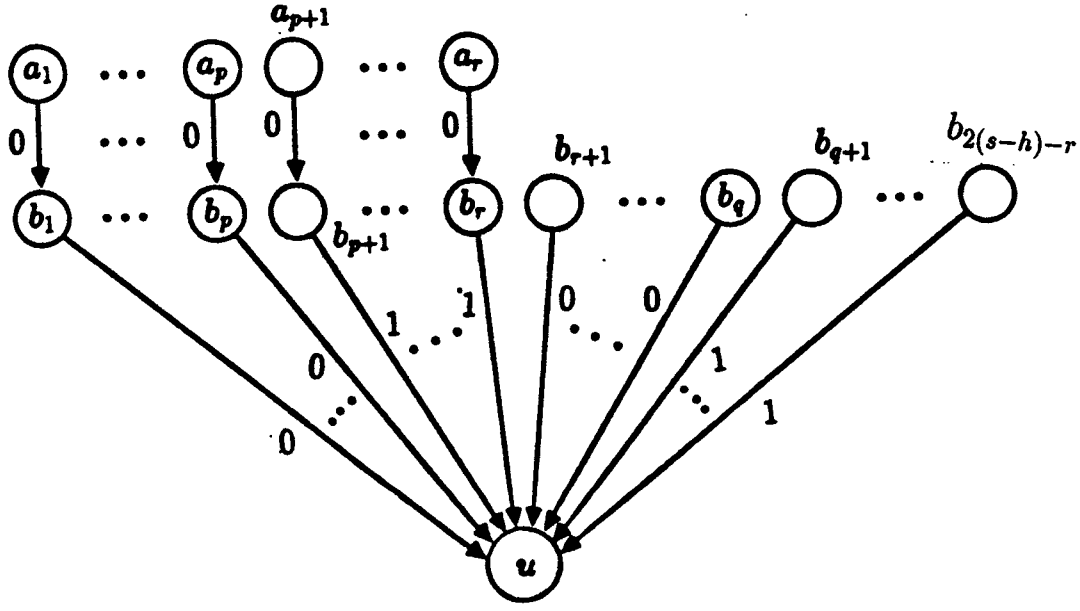


Fig. 2. A general syndrome for $S(u; s - h)$.

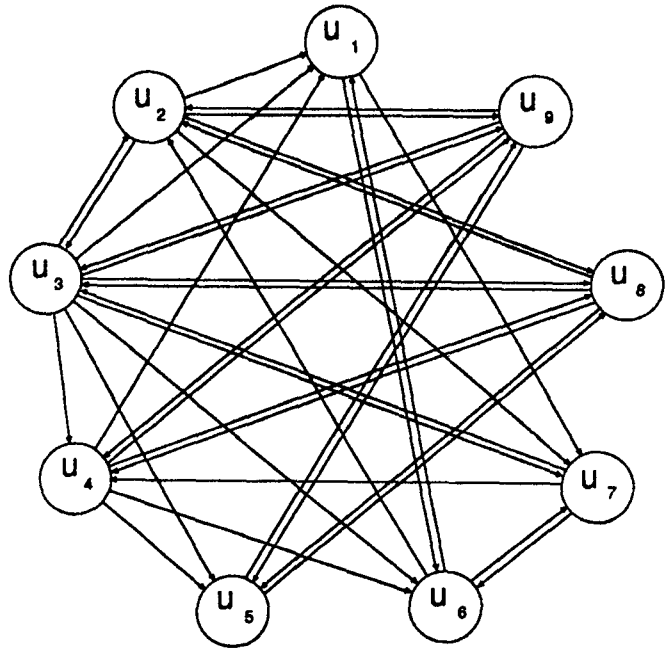


Fig. 3. A 4-diagnosable and u-3-diagnosable system.

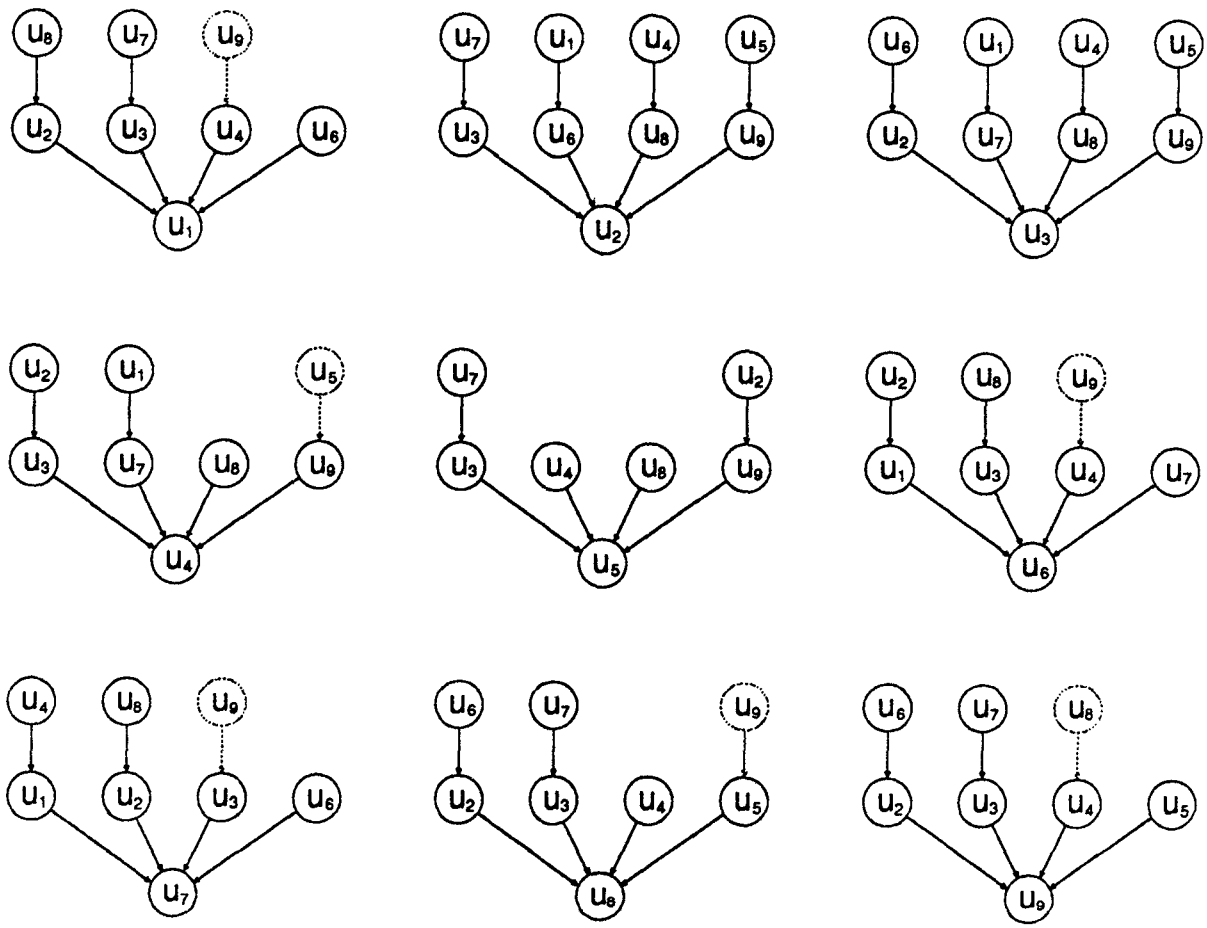


Fig. 4. Optimal unit-diagnosis digraphs for the system of Fig. 3 (vertices and edges in dotted lines are alternative choices).

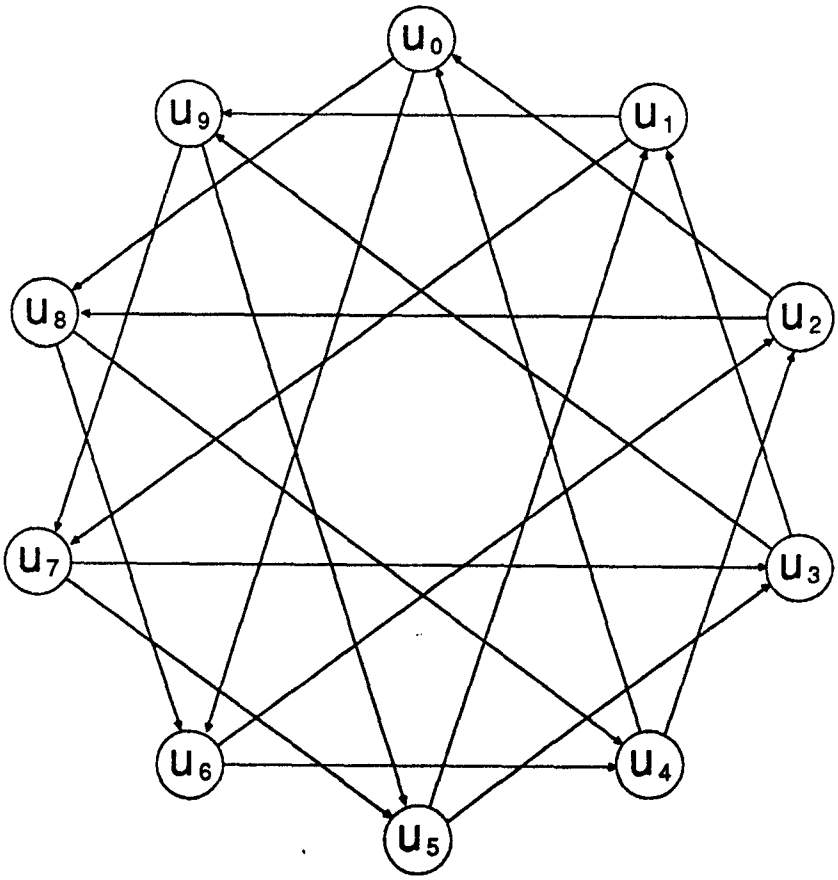
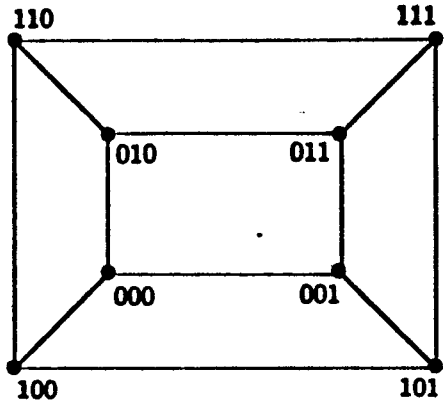
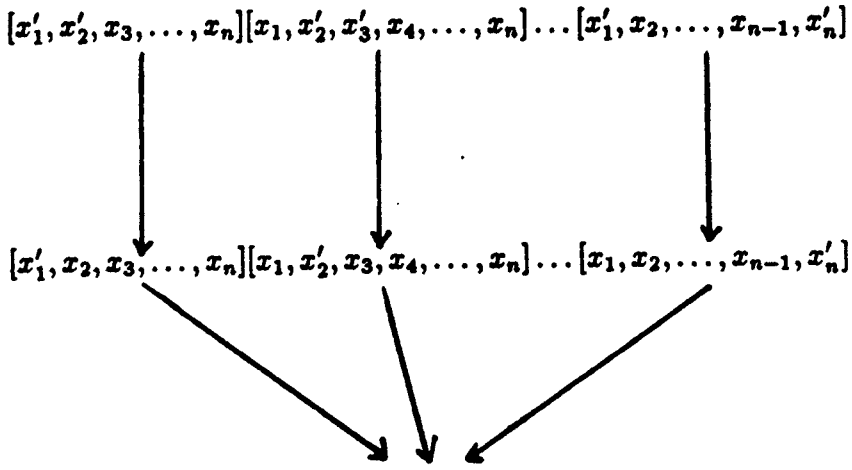


Fig. 5. A u-2-diagnosable system S_1 with $k(G) = 0$.

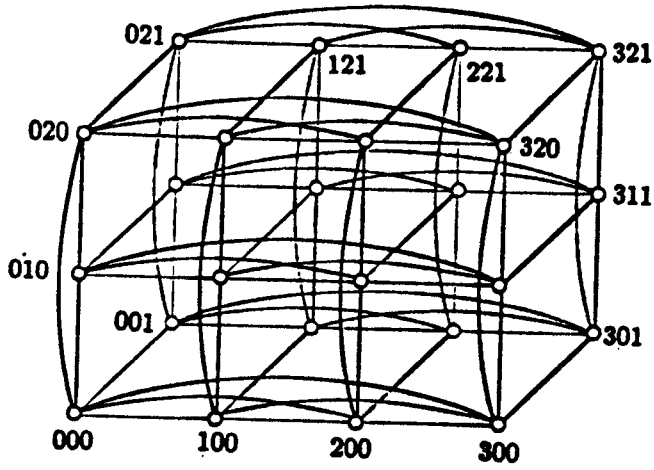


(a) A Boolean 3-cube network.

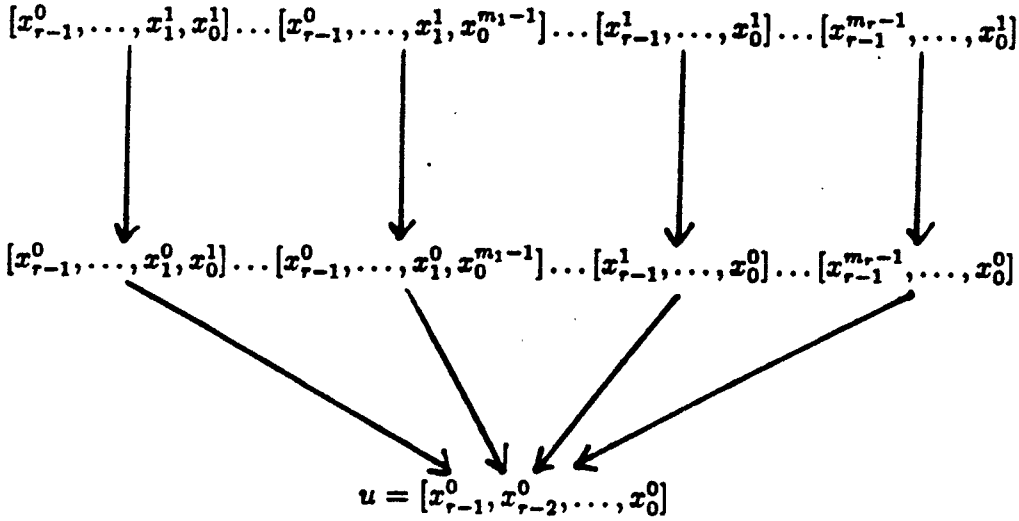


(b) $S(u; N, N)$.

Fig. 6. A Boolean N-cube network and $S(u; N, N)$ digraph.

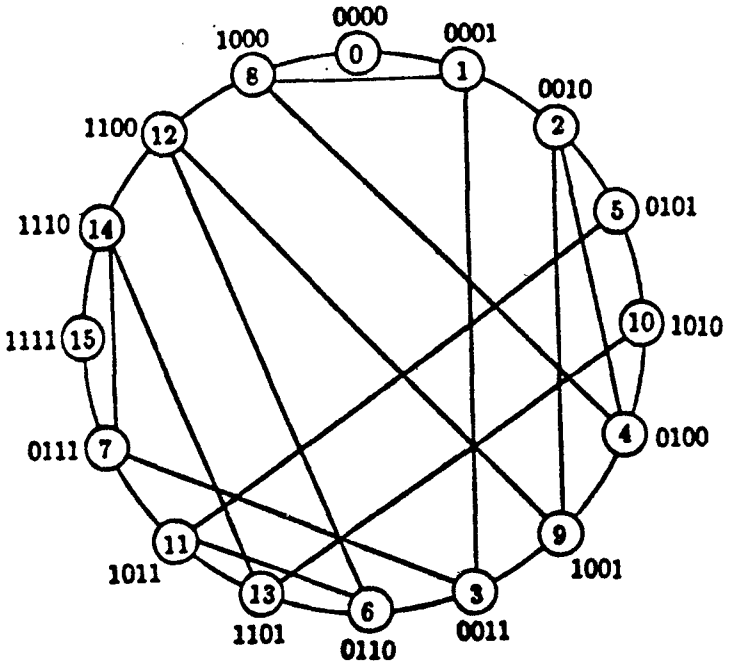


(a) A $4 \cdot 3 \cdot 2$ generalized hypercube structure.



(b) $S(u; M, M)$.

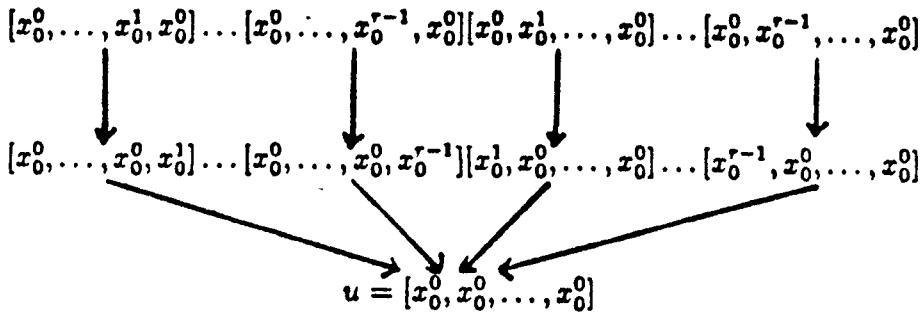
Fig. 7. A generalized hypercube structure and $S(u; M, M)$ digraph.



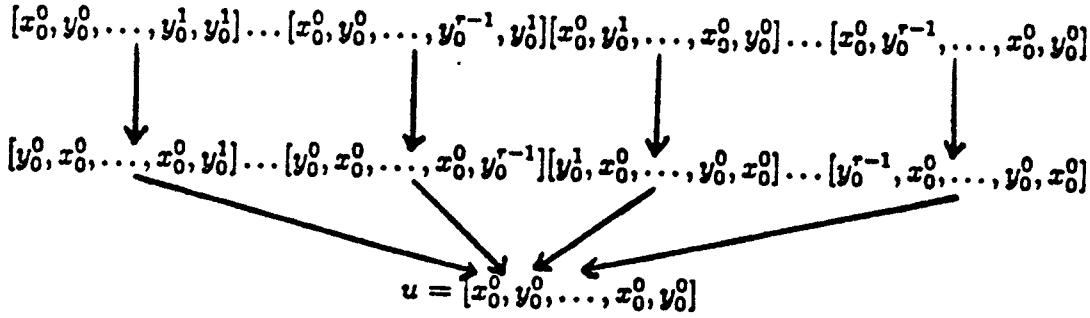
(a) A DeBruijn 16-node network with $m = 4$ and $r = 2$.

Fig. 8. A DeBruijn network and $S(u; 2r-2, 2r-2)$ digraphs.

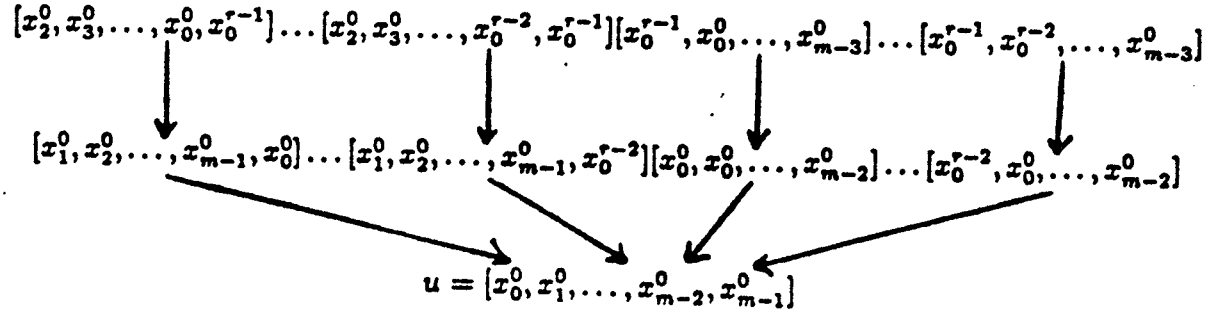
Case 1: $d = 2r - 2$.



Case 2: $d = 2r - 1$, $x_0^0 \neq y_0^0$

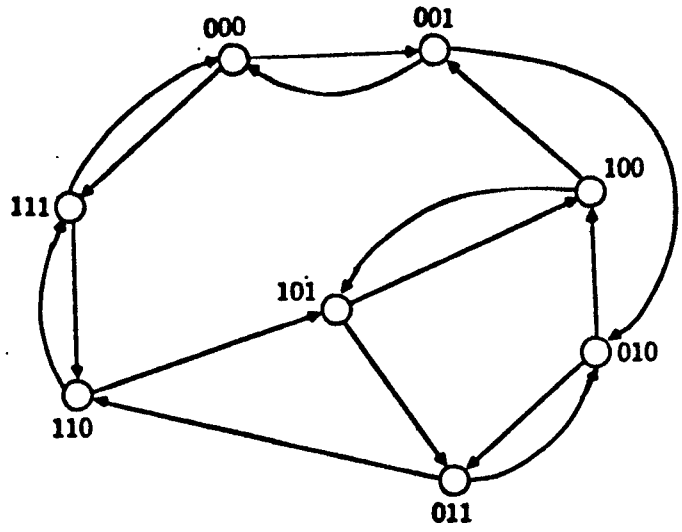


Case 3: $d = 2r$.



(b) $S(u; 2r-2, 2r-2)$.

Fig. 8. A DeBruijn network and $S(u; 2r-2, 2r-2)$ digraphs (Continued).



(a) A directed DeBruijn 8-node network with $m = 3$ and $r = 2$.

Case 1:

$$[x^1_{m-1}, x^0_0, \ldots, x^0_{m-2}][x^2_{m-1}, x^0_0, \ldots, x^0_{m-2}] \ldots [x^{r-1}_{m-1}, x^0_0, \ldots, x^0_{m-2}][x^0_{m-1}, x^0_0, \ldots, x^1_{m-2}]$$

$$[x^0_0, \ldots, x^0_{m-2}, x^1_{m-1}][x^0_0, \ldots, x^0_{m-2}, x^2_{m-1}] \ldots [x^0_0, \ldots, x^0_{m-2}, x^{r-1}_{m-1}][x^0_{m-1}, x^0_0, \ldots, x^0_{m-2}]$$

$$u = [x^0_0, \ldots, x^0_{m-2}, x^0_{m-1}]$$

Case 2:

$$[x^1_0, x^0_0, \ldots, x^0_0, x^0_0][x^2_0, x^0_0, \ldots, x^0_0, x^0_0] \ldots [x^{r-1}_0, x^0_0, \ldots, x^0_0, x^0_0][x^{r-1}_0, x^{r-1}_0, \ldots, x^{r-1}_0, x^{r-2}_0]$$

$$[x^0_0, x^0_0, \ldots, x^0_0, x^1_0][x^0_0, x^0_0, \ldots, x^0_0, x^2_0] \ldots [x^0_0, x^0_0, \ldots, x^0_0, x^{r-1}_0][x^{r-1}_0, x^{r-1}_0, \ldots, x^{r-1}_0, x^{r-1}_0]$$

$$u = [x^0_0, x^0_0, \ldots, x^0_0, x^0_0]$$

(b) $S(u; r, r)$.

Fig. 9. A directed DeBruijn network and $S(u; r, r)$ digraphs.