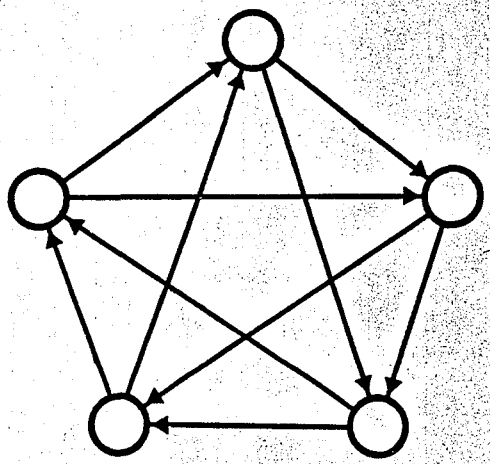
Nr. 4

Werkhefte der
Universität Tübingen

Mario Dal Cin
Elmar Dilger (Eds.)

Self-Diagnosis and Fault-Tolerance



ATTEMPTO
Verlag Tübingen
GmbH

DIAGNOSING ALGORITHMS AND LEARNING

R. Brause, E. Dilger, Th. Risse

Institute for Information Sciences
University of Tübingen, West Germany

ABSTRACT

Algorithms describing the diagnosing mechanism of self-diagnosing fault-tolerant systems are in general very complex. They usually depend on given system parameters. In this lecture we try to improve this approach by introducing the concept of learning. Our conception is to use the system's history, i.e. the information we can get at each system test to achieve a better diagnosing algorithm. This will be done by the method of stochastic approximation. That is, from the diagnosed state at each test point we calculate estimators for the actual system parameters, e.g. the failure rates of single components. Using this additional information we get in nearly all cases an improved diagnosing algorithm. It is shown that describing this model mathematically without loss of generality is too complex. Therefore, we focus on symmetric test structures and symmetric diagnosing strategies, which can be described by a comparatively simple Markov model by reducing the state space of the model to a feasible scale.

1. INTRODUCTION

Various models have been presented to describe and to design fault-tolerant self-diagnosing systems. In investigating these systems two approaches can be distinguished:
a static one in which algorithms for diagnosis are designed, the nature of different systems are compared, repair strategies are developed, etc. (e.g. Preparata et al., 1967 , Kuhl, Reddy 1980)
and a dynamic one where one tries to describe the system's behaviour in time using different methods like renewal theory, markovian processes, operational analysis etc. (e.g. Barlow, Proschan 1967, Buzen 1977).
Both approaches have been combined (compare (Dal Cin/Dilger 1980)).
By considering diagnosis as depending on system parameters, say failure rates of individual units or probabilities of special test results; we get a new link between both aspects if we want to learn

in: M. Dal Cin, E.Dilger (Eds)
Self-Diagnosis and Fault-Tolerance,
ATTEMPTO-Verlag, Tübingen 1981

those system parameters. In this way we constitute the following model:
In each test an updated version of the diagnosing algorithm is applied which is set up by using all the information about the system parameters which is available at that time.
The gauge by which we measure our model is complexity. In our paper we present a diagnosing algorithm of complexity $O(n^3)$. Further complexity analysis for determining the diagnosis matrix in the general case shows that one has to focus on a restricted class of systems.
In expanding the symmetry of the well known D_{1t}-design (Preparata et al.) to repair strategy etc. we come to an appropriate class of systems where we can reduce the state space essentially. So we finally succeed in presenting an explicit formulation of the diagnosis matrix for such systems. With this better insight in the nature of symmetric systems we are able to present a stochastic approximation for the system parameters. In the case of symmetric designs the existence of the limit distribution allows us to quantify the goodness of approximation by establishing an upper bound for the deviation of the mean of the estimation from the actual given parameter.

2. THE MODEL

The multiprocessor or multicomputersystems we consider are assumed to be partitioned in N autonomous units, units which are capable of testing other units, of evaluating their behaviour, of deciding whether another unit is fault-free or faulty, and which are capable of being tested themselves. The decision is '1' if the non-faulty testing unit finds the unit under test to be fault-free, and '0' otherwise. The decisions of a faulty unit are unreliable.
Units and test-connections are represented by a diagnostic graph G=(V,X), (Preparata et al) in which nodes represent units, and an arc

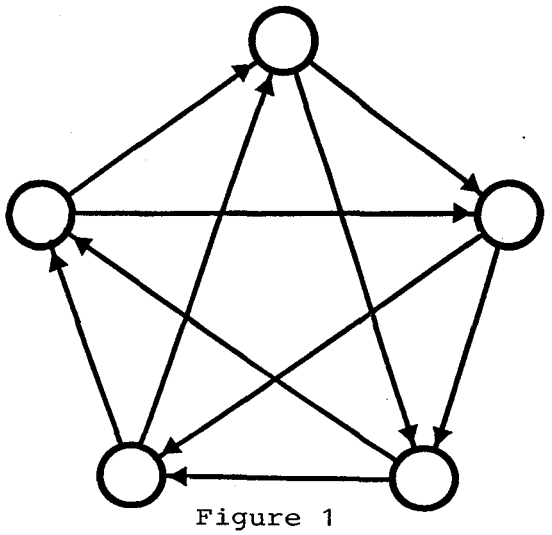


Figure 1

from node i to node j is drawn if, and only if, processing unit P_i is assigned to test processing unit P_j .

The states of the system are described by fault patterns $F=(F(1),\ldots,F(N)) \in \{0,1\}^N$, where $F(j)=1$ if and only if unit P_j is operating. We assume that the reader is familiar with further details of this diagnostic model, especially with the concept of t-diagnosibility with and without repair. Fig. 1 shows as an example a 2-diagnosable diagnostic graph.

It is the D_{12}- graph (Preparata et al). The class of D_{1t}- graphs is optimal in such a sense that, if t or less units are faulty, the faulty units can always be identified, and that for N=2t+1 units there does not exist a graph with a higher t-diagnosibility nor a t-diagnosable graph with fewer test links.

Assumptions:

With regard to the reliability of the units and to our testing and repair strategy, we now make the following assumptions:

(1) The lifetimes $T_n(i)$ of processing unit P_i after the n-th replacement are identically, exponentially distributed and they are independent of the lifetimes of other units. Therefore, the reliability of processing unit P_i at time t is $R_i(t) = \exp(-\lambda_i t)$.

(2) During tests or renewals the system is not available. We shall assume that test- and repair times are small compared with $1/\lambda_i$, the mean of $T_n(i)$.

(3) Tests are scheduled at fixed time intervals of length Δ. 'Working phases' of length Δ alternate with 'test-and-repair phases' of zero length. The begin of a test-and-repair phase is called a checkpoint.

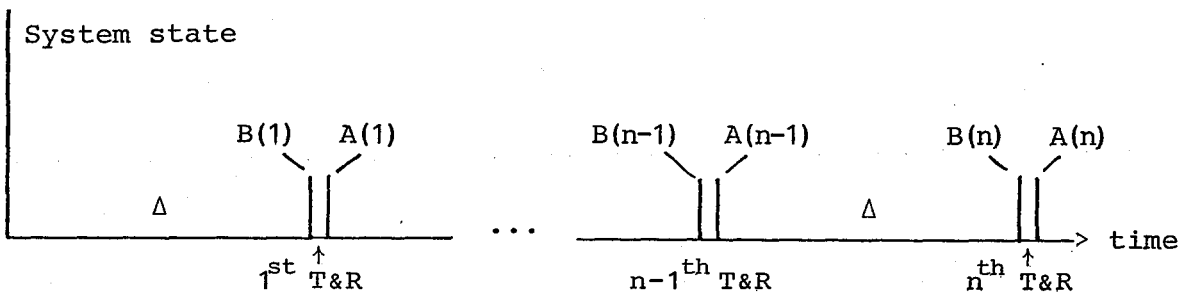


Figure 2

(4) We assume that during a system test, which consists of the execution of all unit tests implied by the diagnostic graph, no failures occur in the system. At each checkpoint the system produces a syndrome

$$S=(S(1),\ldots,S(l)) \in \{0,1\}^l ,$$

where l is the number of test-links due to the diagnostic graph. S(m) is the outcome of the m-th test corresponding to the m-th arc of the graph. Note that the outcome of this test is only reliable if the testing unit itself is reliable. We assume furthermore that the conditional probability

π = P(test-outcome=0 / testing unit is faulty)

is independent of the test time and the testing unit.

(5) Units which are identified as faulty are replaced.

(6) It is further assumed that the diagnosing unit which evaluates the syndrome S is reliable.

In the next section it is shown that with these assumptions the probabilities P(F) and P(S) can be computed as functions of λ and π .

3. DIAGNOSIS

Such a fault pattern as defined above may give rise to several syndromes and a syndrome may have its origin in different fault patterns. Therefore, in order to identify the set of faulty units we have first to find all fault patterns consistent with the given syndrome, i.e. all fault patterns which could have produced this syndrome, and then to decide which one of these fault patterns is likely to be the actual one. Hence, we have to set up a decoding function or algorithm, or strategy which associates with each possible syndrome a consistent fault pattern i.e.

$$\sigma : \{S\} \longrightarrow \{F\}.$$

One possible diagnosing algorithm which has a complexity of $O(N^3)$ is presented in the appendix (for others compare (Kameda et al. 1975), (Kuhl and Reddy 1980)).

We now imagine that the system behaves in the following way.
The system starts with all units up, i.e. with the fault-pattern F1=(1,1,...,1). After a working phase of length Δ, i.e. at the first checkpoint, a test of the system is started and results in a syndrome

S . With the aid of the diagnosing algorithm σ a consistent fault pattern F2 is found. Now we repair the system according to F2, that means, units P_j with $F2(j)=0$ are repaired or replaced. Let the resulting fault pattern be F3, which is the real system's state though we, of course, assume the system to be fault-free. After this test and repair phase the system begins a new working phase of length Δ .

We can describe this system's behaviour by an embedded Marcov chain (Dal Cin and Dilger 1981), as follows:
Denote the state of our system Σ at checkpoint n before and after the test-and-repair phase by b(n) and a(n) respectively, and let F_i and F_j be two fault patterns.
The working phase can be described by the failure matrix $F = (f_{i,j})$, where

$$f_{i,j} = P(b(n)=F_i \;/\; a(n-1)=F_j)$$

which is independent of n .
Given a strategy σ , the test-and-repair phase can be characterized by the repair matrix $D^\sigma = (d_{i,j})$, where

$$d_{i,j} = P(a(n)=F_i \;/\; b(n)=F_j)$$

which is also independent of time. For the calculation of F and D^σ see the appendix of (Dal Cin and Dilger 1981).
For the probability distributions before and after a test-and-repair phase we have

$$A(n) = D^\sigma \cdot B(n) \text{ and}$$
$$B(n) = F \cdot A(n-1) \qquad (3.1)$$
$$\text{and especially } A(n) = D^\sigma \cdot F \cdot A(n-1)$$

which is the state transition equation of a time-homogenous Marcov chain and which allows us to describe the system's behaviour, especially the stationary state distributions A and B

$$A: = \lim_{n\to\infty} (D^\sigma \cdot F)^n \, A(o) \qquad \text{and} \qquad B: = F \cdot A \; . \qquad (3.2)$$

4. A CRITICAL REVIEW

The way we proceeded until now is feasible for small systems. But already for our small example from Fig.1 the state space of our Marcov process has order 2^5 , D and F are 32 x 32 matrices, and the complexity of determining the limiting state distribution is at least

$O(32^3)$, in general, with N units it is $O(2^{3N})$ which is definitely not feasible for larger systems.
Another aspect is , that we should know the exact values of $R_i = \exp(-\lambda_i \Delta)$, i.e. the exact values of λ_i , the failure rate of the i-th unit. Failure rates delivered from the manufacturer may differ from the real rates by some powers of ten.
To avoid this lack of information, and also to get better diagnosing strategies, we can use methods that are well known in pattern recognition problems. To introduce these is the aim of the next paragraph.

5. PATTERN RECOGNITION METHODS

Let us now divide the set of fault patterns (c.f.section 2) into fault classes ω_i, i=1,...,M and let Ω be the set of all possible syndromes. The problem of system diagnosis is to find a partition of Ω into M sets Ω_i where each Ω_i corresponds to a certain fault class $\omega_i = \sigma(\Omega_i)$. The meaning of a fault class can be chosen according to the problem, e.g.such that Ω_i might contain all syndromes which are produced when i units are faulty or is defined to contain all syndromes which can be produced by a single fault state F_i.

SYSTEM DIAGNOSIS

As diagnosing strategy let us choose the minimization of the risk of misclassification. Let L_{ik} be the loss due to the misclassification of the syndrome S produced by ω_k into ω_i. This might be the cost of unnecessary repair or the cost due to not repairing some faulty units. The conditional risk of misclassification (Tou 1974,p.113), when the syndrome S is given, is

$$R(\sigma(S)/S) = \sum_{i=1}^{M} L_{\sigma(S)i} \, P(\omega_i/S)$$

where $P(\omega_i/S)$ is the Probability that the given Syndrom S is produced by the fault class ω_i .
The expected risk for the classification of all syndromes is

$$\text{Risk} = \sum_{S \in \Omega} R(\sigma(S)/S) \; P(S)$$

$$= \sum_{S \in \Omega} \sum_{i=1}^{M} L_{\sigma(S)i} \, P(\omega_i/S) \; P(S)$$

With the 0-1 loss

$$L_{\sigma(S)i} := \begin{cases} 0 & \sigma(S)=i \\ 1 & \sigma(S)\neq i \end{cases}$$

the risk becomes the probability of misclassification

$$\begin{aligned} \text{Risk} &= \sum_{S\in\Omega} \sum_{\substack{i=1 \\ i\neq\sigma(S)}}^{M} P(\omega_i/S)\ P(S) \\ &= \sum_{S\in\Omega} \left(P(S) - P(S/\omega_{\sigma(S)})P(\omega_{\sigma(S)}) \right) \\ &= 1 - \sum_{S\in\Omega} P(S/\omega_{\sigma(S)})\ P(\omega_{\sigma(S)}) \end{aligned}$$

and is minimal when

$$\sum_{S\in\Omega} P(S/\omega_{\sigma(S)})\ P(\omega_{\sigma(S)}) \overset{!}{=} \max \quad .$$

This is achieved by the decision rule

> For every S choose the class $\omega_{\sigma(S)}$ which satisfies (5.1)
> $P(S/\omega_{\sigma(S)})\ P(\omega_{\sigma(S)}) \geq P(S/\omega_i)\ P(\omega_i)$ for all classes ω_i.

It defines the diagnosing function σ which yields the minimal expected error of misclassification.
With this decision rule the set Ω of syndromes is partitioned into M classes. For each syndrome of Ω_k the relation

$$P(S/\omega_k)P(\omega_k) \geq P(S/\omega_i)\ P(\omega_i) \text{ for all other classes } \omega_i$$

holds and, therefore, the functions

$$h_{ik}(S) := P(S/\omega_i)P(\omega_i) - P(S/\omega_k)P(\omega_k)$$

define the boundaries between the class ω_k and the other classes ω_i by

$$h_{ik}(S) \begin{cases} \leq 0 & S\in\Omega_k \\ > 0 & S\notin\Omega_k \end{cases}$$

for i=1..M .
It should be noted that the risk of misclassification with 0-1 loss is the same as

$$\text{Risk} = \sum_{i=1}^{M} \sum_{S \notin \Omega_i} P(\omega_i/S)\ P(S)$$

$$= \sum_{i=1}^{M} (1 - \sum_{S \in \Omega_i} P(S/\omega_i))\ P(\omega_i)$$

$$= 1 - D_{sys}$$

with

$$D_{sys} := \sum_{i=1}^{M} \sum_{S \in \Omega_i} P(S/\omega_i) P(\omega_i)$$

$$= \sum_{i=1}^{M} P(\ \omega_i \text{ was correctly identified }\ )\ P(\omega_i)$$

This diagnosibility D_{sys} of the system was defined by (Blount 1977). Since maximizing D_{sys} is the same as minimizing the Risk he gets the same decision rule (5.1) .

LEARNING

When the exact class-boundaries are not known because sufficient a priori information is not available, an iterative process might be created to update the estimated boundaries.
In the parametric approach (Tsypkin 1973,p.10) the boundaries depend on a parameter, say c (e.g. the failure rates λ_i or the reliabilities R_i) , and the aim of the process is to find the parameter value c* which minimizes a given risk or performance function Risk(c)

$$\text{Risk}(c^*) := \min_{c} \text{Risk}(c)$$

This is only meaningful if Risk(c) and the parameter set {c} are well chosen, e.g. if Risk(c) has a local minimum at c* . The exact assumptions are considered later on.
If the function Risk(c) is explicitly known, it is quite easy to find the extremum with $\nabla_c \text{Risk}(c^*) = 0$ analytically or by the iterations of a hill-climbing algorithm. In the latter case the n-th stage of the learning algorithm

$$c_n := c_{n-1} - \gamma_n \nabla_c \text{Risk}(c_{n-1}) \qquad (5.2)$$

updates the parameter c_{n-1} in the direction of the decrease of Risk(c) and finally leaves c_n unchanged when $\nabla_c \text{Risk}(c) = 0$ at $c = c^*$.
The coefficient γ_n prevents an overestimation of c_n when there are rapid changes in Risk(c) and must be chosen according to the given problem.

LEARNING BY STOCHASTIC APPROXIMATION

Let Risk(c) be the expectation value of a performance measure r(S,c) of the system, e.g. the parameterized risk of the classification of a syndrome and assume that the syndromes are always correctly diagnosed, so they do not depend on the previous diagnosis and repair.
Thus,

$$\text{Risk}(c) = \sum_{S\in\Omega} r(S,c)\ P(S)$$

and assume that the derivative exists

$$j(S,c) := \nabla_c r(S,c)$$

$$J(c) := \nabla_c \text{Risk}(c) = \sum_{S\in\Omega} \nabla_c r(S,c)\ P(S) = E\ (j(.,c)) \qquad (5.3)$$

Let us consider the situation when only the random functions r(S,c) and j(S,c) and neither Risk(c) nor the distribution (S) are known. How is it then possible to compute the root of the regression function J(c) without knowing it ?
This problem is solved by the method of stochastic approximation which was introduced by Robbins and Monro. They showed that the algorithm

$$c_n := c_{n-1} - \gamma_n j(S_n, c_{n-1}) \qquad (5.4)$$

S_n = syndrome of the n-th test

which is apparently a stochastic version of (5.2) , provides the means to update c_{n-1} . The algorithm now converges to c* in the mean square

$$\lim_{n\to\infty} E((c_n - c^*)^2) = 0$$

under the following assumptions

about the random variables :

1) j(S,c) is an unbiased sample of J(c),
i.e. E(j(.,c))=J(c) for all c

2) the variance is finite
$E((J(c)-j(S,c))^2)=\sigma^2 < \infty$

about the function J(c) :

3) J(c) is bounded
$|J(c)| < a|c-c^*| + b < \infty \quad a,b\in\mathbb{R}$

Risk(c) has an unique extremum i.e. a minimum

4a) J(c) has a single root c* with J(c*)=0

4b) J(c') > J(c*) if c' >c*
J(c'') < J(c*) if c''<c*

about the coefficients γ_n:

5)

a) $\lim_{n\to\infty} \gamma_n = 0$ b) $\sum_{n=1}^{\infty} \gamma_n = \infty$ c) $\sum_{n=1}^{\infty} \gamma_n^2 < \infty$

for instance the harmonic sequence $\gamma_n = 1/n$.

6. APPLICATION OF PATTERN RECOGNITION METHODS

The aim is to learn the individual failure rates of the units in order to detect the weak points of the system. To this end we want to apply the methods introduced in section 5 in order to learn these individual rates and to use them in the diagnosis. Therefore, the fault classes ω_i are identified with the fault patterns F_i (c.f.section 2) to reflect the information about every single unit.

SYSTEM DIAGNOSIS

Then the probabilistic diagnosing function σ according to (5.1) is defined as

$$\left|\begin{array}{l}\text{choose } \sigma(S) := F_k \text{ such, that for all other classes} \\ F_i \text{ the relation } \quad h_{ik}(S) \leqq 0 \quad \text{holds.}\end{array}\right. \tag{6.1}$$

Now we want to compare this diagnosing function with the concept of t-diagosibility without repair. The latter implies that for every syndrome which might be produced by a fault pattern F_i with $\leqq t$ faulty units, every other fault pattern F_j which might produce the same syndrome has $>t$ faulty units.
Let Ω^1 be the set of all syndromes which can be produced by a fault pattern with $\leqq t$ faulty units. For this set we can define a deterministic diagnosis

$$\left|\begin{array}{l}\text{For all } S\varepsilon\Omega^1 \text{ choose } \sigma(S) \text{ to be the unique fault} \\ \text{pattern which produces S and has } \leqq t \text{ faulty units.}\end{array}\right. \tag{6.2}$$

It can be shown that the deterministic diagnosing function (6.2) and with this also the diagnosing algorithm in the appendix provides the same decisions for Ω^1 as the probabilistic one (6.1) for $R_i > 0.5$.
It remains to define the diagnosing function $\sigma(S)$ for $\Omega^2 := \Omega - \Omega^1$, i.e. for all syndromes which are produced by more than t faulty units. This can be done by either

$$\left|\begin{array}{l}\text{For all } S\varepsilon\Omega^2 \text{ choose } \sigma(S) := (0,\ldots,0) \text{ , i.e. the whole} \\ \text{system is faulty (the pessimistic strategy of section 9).}\end{array}\right. \tag{6.3}$$

or

$$\left|\begin{array}{l}\text{For all } S\varepsilon\Omega^2 \text{ choose } \sigma(S) \text{ so that for all other} \\ \text{classes } F_i \text{ the relation } \quad h_{i\sigma(S)}(S) \leqq 0 \quad \text{holds.}\end{array}\right. \tag{6.4}$$

With the extension (6.4) for the deterministic diagnosing function (6.2) both probabilistic and deterministic diagnosis are the same for all $S\varepsilon\Omega$.

LEARNING THE FAILURE RATE

For simplicity let us assume that all units have the same failure rate and that after a test interval Δ *we observe* that m units out of N previously working ones have failed.
The estimated probability of survival of the test interval for one unit is taken as

$$\hat{R} := 1 - \frac{m}{N}$$

Because of the assumption of exponentially distributed lifetimes it is equivalent to estimate the fault-rate

$$\hat{\lambda} = \frac{1}{\Delta} \ln \left(\frac{N}{N-m}\right) \qquad (6.5)$$

such that $\hat{R} = 1-\exp(-\hat{\lambda}\,\Delta)$

The stochastic approximation algorithm (5.4) suggests the following iteration algorithm

$$R_n := R_{n-1} - \frac{1}{n} (R_{n-1} - \hat{R}_n) \qquad (6.6)$$

where $\hat{R}_n$ is the value of $\hat{R}$ observed after the n-th test interval.
The regression function is given by

$$J(c) = E(j(.,c)) \qquad \text{with} \qquad j(\hat{R},c) = c-\hat{R} \qquad \text{and} \quad c^*=R\ .$$

Then it is guaranteed that

$$\lim_{n\to\infty} E((R_n-R)^2) = 0$$

and the Risk(c) is the mean square error

$$\text{Risk}(c) = \frac{1}{2} E((c-\hat{R})^2)$$

In this simple case we have a Bernoulli experiment and clearly the assumptions 1)-5) of section 5 are satisfied, R_n converges to $E(\hat{R})=R$. It should be noted that in this way individual failure rates can be learned, too.

7. CRITICAL REVIEW

We have shown that under the assumption that we can observe the system's state by a correct dignosis of S_n the method of stochastic approximation provides a good tool for estimating the reliability R or the corresponding failure rate λ , c.f. figure 3 .

Unfortunately the assumption of unbiased, independent observations of the number m of faulty units does not hold, because we cannot observe the true system state. Instead our knowledge of m after the test interval Δ is based on the diagnosis of the syndrome. All syndromes of Ω^1 might be produced as well by fault patterns with more than t faulty units. So we make a number of misclassifications which increases our estimation of R and decreases that one of λ. If we use on the other hand the pessimistic diagnosing function (6.3) then the resulting offset of $E(\hat{R})$ to R is inverse.
Additional complications arise if we consider that in the case of a wrong diagnosis the system is insufficiently repaired and so the number of fault-free units after the test and repair phase is not N, decreasing our sample estimate of R .

All these complications make the underlying probability distributions differ from those we assumed in section 2.
Since the variations remain finite the convergence to the expectation value of $\hat{R}$ is still garanteed- but what can we say about the deviation from R ? The general answer is too complex because we have to take into account all possible system states and their successors after a wrong diagnosis. For this reason we will focus our attention in the next section on the well-known D_{1t}-design.

8. REDUCTION OF THE STATE SPACE, SYMMETRIES

In connection with the criticism we made in section 4 about our Marcov model we now try to reduce the state space in order to make computations feasible also for larger systems.
In this section we restrict ourselves to the D_{1t}-designs as mentioned in section 2 .
For each $t \in \mathbb{N}$ we get a diagnostic graph with N=2t+1 nodes, i.e. units, numbered from 0 to N-1 and arranged in a ring, where there are test links from unit P_i to units $P_{i+1},\ldots,P_{i+t}$, where + is to be taken modulo N. So the structure is quite symmetric and we know that a system with a D_{1t}-diagnostic graph is t-diagnosable, that means if t or less units fail, the failed units can be localized uniquely.
In other words, we get

$\sigma(\sigma^{-1}(F)) = F$ for all $F \in \{ F$: number of 0-components in F is $\leq t\}$.

In this section we shall use a strategy which makes use of the properties of D_{1t} graphs. It is a symmetrical strategy, in some respects it is a pessimistic strategy (Dal Cin and Dilger 1981):

(a) Whenever a syndrome appears which belongs to a fault-pattern with t or less than t faulty units, we assume that the system state is the aforementioned unique one.

(b) Whenever we are sure that more than t units are faulty (i.e. not as in case a), we assume that all units are defect and we repair or replace the whole system.

We also use a symmetrical reliability structure. We assume that our system is a (N-t)-out-of-N - system, that means that the system is system is operational if at least N-t are units operational.
By using symmetry we must assume that all failure rates $\lambda_i := \lambda$ are equal, i.e. $R_i = R = \exp(-\lambda\Delta)$, for i=0,...,N-1 .

In this case we can reduce the dimension of our state-space from 2^N to N+1 if we identify all the $\binom{N}{i}$ fault-patterns with i faulty units with the state i (i=0,1,...,N).
(There exists a slightly better reduction using a state space with dimension t+2, by lumping together all states with t+1 and more faulty units, but we will not use it here.) As indicated in (Dal Cin, Dilger 1981) the computation of the repair matrix D^σ is quite complicated (c.f. section 3). By group-theoretical arguments we can state the matrix elements directly:

$$d_{i,j} = \binom{i}{j} (1-\pi)^{i(i-1)/2} \pi^{t\, i-i(i-1)/2} \qquad \text{for } i \geq j > t$$

$$d_{i-t,i} = \left[\frac{N}{i}\right] (1-\pi)^{(i-t)(3t-i+1)/2} \pi^{(i-t)(i-t-1)/2} \qquad \text{for } i > t$$

$$d_{i,0} = 1 - \sum_{j=1}^{N} d_{i,j} \qquad \text{for } t+1 \leq i \leq N$$

$$d_{0,i} = 1 \qquad \text{for } 0 \leq i \leq t \text{ and all other } d_{i,j} \text{ are zero.}$$

The elements of the failure matrix F are easily computed:

$$f_{i,j} = \binom{N-j}{N-i} R^{N-i} (1-R)^{i-j} \qquad \text{for } i \geq j$$

$$f_{i,j} = 0 \text{ else.}$$

9. STOCHASTIC APPROXIMATION IN THE CASE OF SYMMETRIC DESIGNS

As an example for the convergence behaviour of the stochastic approximation we consider now the case that the lifetimes of all units are exponentially distributed with the s a m e but unknown mean $1/\lambda$. We do not then need to distinguish between states with the same number of failures and can therefore use the results of section 8 . We want now to learn the parameter λ .

Let m_n be the number of failed units as evaluated by the diagnosis in the n-th test. Using e.g. maximum-likelyhood-estimation we get

$\hat{\lambda} = 1/\Delta \quad \ln(n/(n-m_n))$

as a new estimator for the failure rate λ . Giving to all the so far computed estimations the same weight we compute the estimation in the n-th test phase as

$\lambda_n = (1-1/(n+1))\ \lambda_{n-1} + 1/(n+1)\ \hat{\lambda}$, where λ_0 is for example the failure-rate given by the manufacturer.

We would like to show that the λ_n are converging stochastically or in the mean and to characterize the deviation of this limit from the unknown λ subject to the system parameters and the chosen diagnosis strategy.

However, we have to face two problems:

1) To determine for example $E(\lambda_1)$ utilizing (3.1) we have to know the diagnosis matrix. D^σ .
2) If we assume that the diagnosis function σ of the n-th test is evaluated using the (n-1)-th estimation λ_{n-1} with, for example the constraint to minimize the 0-1-loss then the estimators for λ are not independent.

To get rid of the first problem we restrict ourselves to the investigation of symmetric designs as in section 8 . The second problem vanishes if we are only interested in getting an upper bound for the expectation of estimators, in which case we choose for each test the pessimistic diagnosis function.

Because we assigned the same weight to each new estimator we do not need to compute the mean of each estimator and then the limit of these expectation values. Instead we compute the stationary state distribution B using the very good converging behaviour of the markov chain, as it is described by the failure matrix F and the diagnosis matrix D^σ and evaluate the mean $E(\lambda_\infty)$ of the estimator in the limit case, i.e. when the system's behaviour obeys the stationary state distribution.

From section 3 it is easily computed how many units are said to be faulty assuming that the system is in state j . So we get the (n+1)x(n+1) matrix Nr with Nr(i,j):=number of repaired units, given the system is in state j and after repair the system is in state i . Choosing an appropriate time unit we assume without loss of generality that $\Delta = 1$. Using the diagnosis matrix D^{σ} from section 8 we get

$$E(\lambda_{\infty}) = \sum_{m=0}^{N} \ln\left(\frac{N}{N-m}\right) P(\text{m units are to be repaired in the limit case})$$

$$= \sum_{m=0}^{N} \ln\left(\frac{N}{N-m}\right) \sum_{i,j:Nr(i,j)=m} d_{i,j} \cdot B_j$$

The figures below give an impression of the relationships for the D_{1t} design with five units.

In this case we conclude that the approximation $E(\lambda_{\infty}) \cong 1.12\,\lambda$ holds for small values of λ . For the other strategies the mean of the estimation in the limit case has a smaller distance to λ than that one computed for the pessimistic strategy.

Figure 3

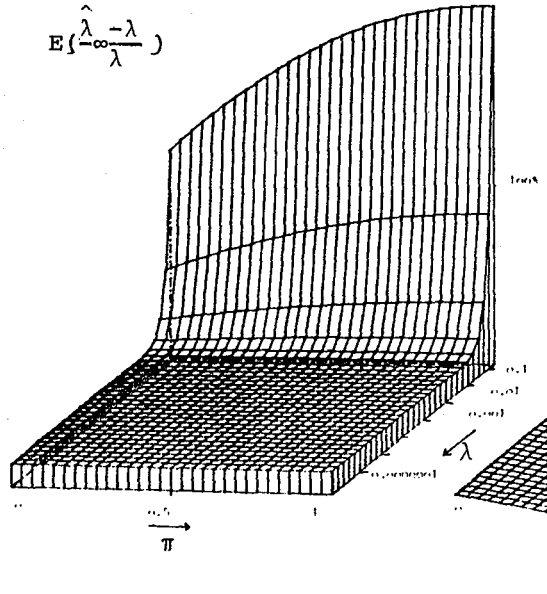
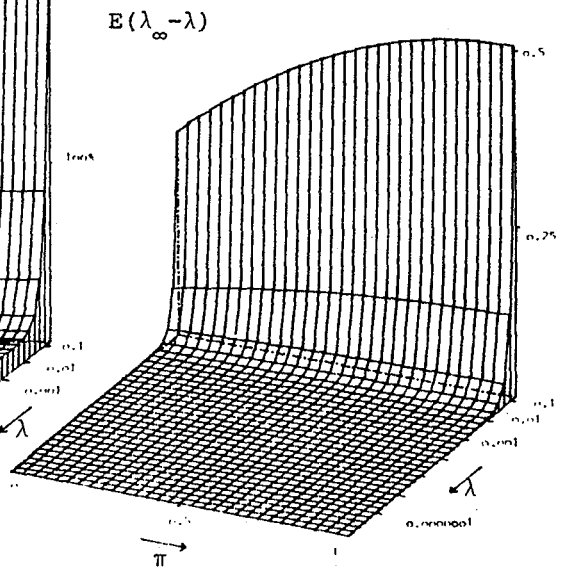


Figure 4



10. CONCLUSION

In this lecture we presented a probabilistic approach to describe the behaviour in time of self-diagnosing multi-processor systems, using Markov chains. We pointed out that in general it will be too complex to describe larger systems in this way (compare for example the presented diagnosing algorithm in the appendix). Therefore we proposed a reduction of the state space making exhaustive use of symmetries.

Complexity is increasing essentially if we have to deal with systems which are designed to show an adaptive behaviour. After an excursion in learning theory we focussed on systems which have to learn their system parameters in order to achieve a better diagnosing strategy. Due to complexity reason and because of the unusual way of questioning in learning theory we showed solutions only for symmetric systems. It should be mentioned that the assumption we made in section 2 about zero - test- and repair-time can be dropped if we use decentral diagnosis as proposed by Kuhl and Reddy 1980 .

ACKNOWLEDGEMENTS

This work was supported by the Deutsche Forschungsgemeinschaft. The authors appreciate fruitful discussions with M. Dal Cin, E. Ammann and H.J. Fuchs.

APPENDIX : A DIAGNOSING ALGORITHM

To develop a diagnosing algorithm we introduce two operations on test results. The test result $d_{i,j}$ of unit i on unit j (not to be confused with the elements of the repair matrix) can have four values:

0(1) iff the tested unit j is judged to be faulty (fault-free),
x iff the testing unit i has no information about the unit j to be tested, and
- iff the testing unit i is identified as faulty.

Let us now imagine that each unit has to form its view of the system state.
If $d_{i,j} = 1$ then unit i can use all test results computed by unit j , i.e. unit i gets the -indirect- result $d_{j,k}$ about unit k . Otherwise the results of unit j must not be used by unit i .
If on the other hand unit i obtaines two different results $d_{i,j}$ and $d'_{i,j}$ about the state of unit j then unit i has to be faulty.

If it turned out that unit i must be faulty and if $d_{k,i} = 1$ then also unit k has to be faulty. In this way we can set up two operations on the test results:

$d_{i,j} \cdot d_{j,k}$

	0	1	x	-
0	x	x	x	x
1	0	1	x	-
x	x	x	x	x
-	-	-	-	-

$d_{i,j} + d'_{i,j}$

	0	1	x	-
0	0	-	0	-
1	-	1	1	-
x	0	1	x	-
-	-	-	-	-

As one easily verifies both operations are associative, addition is commutative and multiplication is not. Only with respect to multiplication from the left are the two operations distributive.
Using the above defined addition and multiplication we can define a multiplication of matrices over $\{0,1,x,-\}$. However, this multiplication is not associative. Therefore, we have to fix the order of brackets to define powers of such matrices.

Let $D^n := D\ (\ D^{n-1}\ )$ and $D^1 := D$. Using the distributivity from the left we can determine the elements $(d^{(n)}_{i,j})$ of D^n as

$$d^{(n)}_{i,j} = \sum_k d_{i,k} \cdot d^{(n-1)}_{k,j} = \sum_{k_1} \sum_{k_2} \cdots \sum_{k_{n-1}} d_{i,k_1} \cdot d_{k_1,k_2} \cdot \ldots \cdot d_{k_{n-1},j} .$$

Given a syndrome we define the so called syndrome matrix $D = (d_{i,j})$ where $d_{i,j}$ is the label of the test link i-->j if it exists and $d_{i,j}$ is set to 'x' otherwise. Set $d_{i,i}=1$ for all i. So $d^{(n-1)}_{i,j}$ becomes the test result on the state of unit j which unit i can compute using test results of no more than n-1 other units. Starting with the syndrome matrix D of a system with n units each test result of unit i in D^n contains all the information about the tested unit which unit i can use.
If the syndrome to be diagnosed is consistent to a system's state with no more than t faults then D^n has at least n-t consistent rows. The component sum of these rows becomes the diagnosed state in setting all 'x' to '0' . The rows corresponding to a faulty unit may contain '-' indicating that this unit can not be fault-free. We observe now that

$$d^{(n)}_{i,j} = d^{(n-1)}_{i,j} + \sum_{k,d_{i,k}=1} d^{(n-1)}_{k,j} =$$

$$= d_{i,j} + \sum_{d_{i,k_1}=1} d_{k_1,j} + \ldots + \sum_{d_{i,k_1}=1} \sum_{d_{k_1,k_2}=1} \cdots \sum_{d_{k_{n-2},k_{n-1}}=1} d_{k_{n-1},j}$$

It should be noted that $D^n = D^n + D^{n-1} + \ldots + D$ holds.
We now want to compute the systems state evaluated by unit i using test results of other appropriate units. To determine $d_{i,k}$ for all k

we proceed as follows: How does a test result $d_{j,k}$ of a unit j where $d_{i,j}$ equals '1' affect the result $d_{i,k}$ evaluated by unit i itself if we look for such units j step by step ?
We then use $d_{i,k} = d_{i,k} + d_{j,k}$ for $d_{i,j}=1$ and distinguish three cases:

- We arrive at a contradiction of test results, i.e. the new value of $d_{i,k}$ is '-'. Then we conclude that unit i and all units k with $d_{k,i} = 1$ cannot be fault-free.
- $d_{i,k}$ changes its value from 'x' to '1' . Then only in the case where we previously had to refuse to use test results of unit k (because $d_{i,k} \neq 1$ was at that time) we now have to update the system's state with the aid of unit k . Otherwise we will have corrected the condition on $d_{i,k}$ when coming to update the pretended systems state established so far by unit i with the help of unit k .
- in all other cases the so far computed results remain unchanged.

So we get the following algorithm, which we present here in a Pascal-like manner:

```
PROCEDURE Falsifiz(I);
BEGIN
 FOR L:=1 TO N DO D(I,L):='-';
 FOR L:=1 TO N DO IF D(L,I)=1 THEN Falsifiz(L)
END;
(* Main Algorithm, a special stack  is defined *)
FOR I:=1 TO N DO
 BEGIN
  FOR J:=1 TO N DO IF (I≠J) AND (D(I,J)=1) THEN
   BEGIN
    JJ:=J;Clear-Stack;
ZADD:  FOR K:=1 TO N DO
        BEGIN
         Dnew:= D(I,K) + D(J,K);
         IF Dnew='-' THEN BEGIN Falsifiz(I);GOTO NextI END;
         IF (Dnew=1) AND (D(I,K)='x') AND (K < JJ) THEN Push(K);
         D(I,K):=Dnew
        END;
    IF NOT Empty-Stack THEN BEGIN Pop(JJ);GOTO ZADD END;
   END
NextI: END
```

We now want to determine the complexity of the above algorithm. Let the number of computations of Dnew be a measure of complexity. We try therefore to detect how often the (innermost) k-loop is performed for each i , i.e. the number of times we have $d_{i,j} = 1$ or we stacked a new k with $d_{i,k}= 1$ with $k < j$. So we see that we could have stacked at the beginning all those $j = i$ with $d_{i,j}= 1$ and then continued as

indicated. Then obviously maximal N different elements have to be stacked. Therefore the k-loop is performed maximal N times for each i and thus the complexity of the algorithm is $O(N^3)$. If we consider the complete directed graph with N nodes in which all edges are labeled with '1' then we note that this upper bound is reached.
The comparison of the n rows to detect whether or not there are at least N-t consistent rows can also be done in less than $O(N^3)$ steps. Namely, for each of the N rows we compute the number of the consistent rows and simultaneously sum them up to get the diagnosed system state. We now detect whether or not there exists in the first t rows a row which has at least N-t-1 consistent rows. In the first case we substitute 'x' by '0' if necessary to finish the diagnosis, whereas in the latter case we conclude that more than t failures occured and, if we are pessimistic, we assume that all units are faulty.

REFERENCES

Barlow, R., Prochan, F.: Mathematical Theory of Reliability, New York 1967

Blount, M.L.: Probabilistic Treatment of Diagnosis in Digital Systems, FTCS-7,p. 72-77 , 1977

Buzen, J.P.: Operational Analysis - an Alternative to Stochastic Modeling, Lincoln, 1977

Dal Cin, M., Dilger, E.: On the diagnosibility of self-testing multi-microprocessor systems, Microprocessing and Microprogramming 7 (1981), 177-184

Dal Cin. M, Dilger, E.:Self-testing and self-diagnosing multicomponent systems, Digest of papers FTCS-10, Kyoto (1980)

Kameda et al.: A diagnosing Algorithm for Networks, Information and Control 29, 141-148 (1975)

Kuhl, J.G., Reddy, S.M.: Distributed Fault Tolerance for Large Multiprocessor Systems, Proc. 7-th Ann. Symp. on Computer Architecture, La Baule, 1980, 23-30

Preparata, F.P. et al.:On the Connection Assignment Problem for Diagnosable Systems, IEEE Trans. Electr. Comp. Vol EC-16, 848-854 (1967)

Robbins, H., Monro, S.: A Stochastic Approximation Method, Ann. Math. Stat. 22 ,p.400-407,(1951)

Tou, J.T., Gonzales, R.C.: Pattern Recognition Principles, London 1974

Tsypkin, J.S.: Foundations of the Theory of Learning Systems, New York 1973