DISTRIBUTED DIAGNOSIS OF SYSTEMS WITH MULTIPLE FAULTS

Hector Geffner
Judea Pearl

October 1986
CSD-860023

DISTRIBUTED DIAGNOSIS OF SYSTEMS WITH MULTIPLE FAULTS*

Hector Geffner & Judea Pearl

Cognitive Systems Laboratory, UCLA Computer Science Department, L.A., CA. 90024

ABSTRACT: The paper describes a distributed scheme for finding the most likely diagnosis of systems with multiple faults. The scheme uses the independencies embedded in a system to decompose the task of finding a best overall interpretation into smaller sub-tasks of finding the best interpretations for subparts of the net, then combining them together. This decomposition yields a globally-optimum diagnosis by local and concurrent computations using a message-passing algorithm. The proposed scheme offers a drastic reduction in complexity compared with other methods: attaining linear time in singly-connected networks and, at worst, exp(| *cycle–cutset* |) time in multiply-connected networks.

1. Introduction

The diagnosis problem is to determine those components of a system which, when assumed to be functioning abnormally, will "best" explain the discrepancy between the observed and correct system behavior. Diagnostic reasoning systems appear to adopt one of the following two approaches: empirical reasoning and model-based reasoning (sometimes called "reasoning from first principles"). In the first approach, typified by the MYCIN system [Buchanan & Shortliffe 1984], we codify the *behavior of a human diagnostician*, while the structure of the real world system being diagnosed is only weakly represented. In the second approach the only available information is the *system description*, i.e., its design or structure, together with some observations of the system's behavior and an incomplete, often statistical, characterization of the type of failures one should anticipate or their relative likelihood. Notable examples of model-based approaches are [Davis 1984, Genesereth 1984, Cooper 1984, Reiter 1985 and de Kleer & Williams 1986a].

This paper adopts the latter approach and is based on [Pearl 1986b]; causal knowledge of system behavior is represented in belief networks, and diagnosis is performed using Bayesian inference. However, while previous work on belief networks focused on computing a numerical degree of belief in individual propositions [Pearl 1986a], this paper uses the same formalism to generate a categorical, multi-hypotheses description constituting the best global explanation of the observed behavior.

In principle, the task of diagnosing systems with multiple faults seems intractable because enumerating and rating all possible fault combinations is computationally prohibitive, even in systems of moderate size. It is not surprising, therefore, that the problem has been treated with a variety of heuristics and search-pruning techniques. Ben-Bassat [1980] and Pople [1982] have used heuristics to assemble a suboptimal set of hypotheses based on how they rank individually. Reggia et al. [1983], de Kleer et al. [1986a] and Reiter [1985] have developed techniques for identifying those diagnoses which are "minimal," i.e., they contain no proper subset of faults which equally explains the symptoms observed. de Kleer & Williams [1986b] also extended these techniques to include probabilistic information about device failure rates.

The computational problem is further aggravated in cases where the system description itself is non-deterministic, e.g., in medical diagnosis. The notion of "minimality" is no longer helpful because the extent to which a diagnosis explains or "covers" a set of symptoms is a matter of degree; so, it is no longer obvious to decide that a given diagnosis cannot be improved by postulating a larger set of faults. To our knowledge, the systems of Cooper [1984] and Peng & Reggia [1986] are the only ones which take into account both probabilistic information about fault likelihood and uncertainty about system behavior and still return an optimal, i.e., most likely, diagnosis. However, these systems employ a branch-and-bound search algorithm which often runs in exponential time and often misses those structural properties of the diagnosed system which could make the search significantly faster, if not superfluous. In addition, although the outcome of the search is globally optimal, it is hard to justify in meaningful terms because the global process of searching for that outcome is very different from the local mental process exercised by human diagnosticians.

This paper departs from previous work by basing the diagnostic process on a *local* and *distributed* mechanism of belief revision, while still guaranteeing a globally-optimal solution [Pearl 1986b]. The impact of each new piece of evidence is viewed as a perturbation that propagates through a network via local communication among neighboring variables, with minimum external supervision. At equilibrium, each variable has a definite value which, together with all other value assignments, is the best global interpretation of the evidence. The main reason for adopting this distributed message-passing paradigm is that it provides a natural mechanism for exploiting the independencies embodied in the system's description and translating them (by sub-task decomposition) into an order of magnitude reduction of complexity. Another reason is that it leads to a "transparent" belief revision process in which the intermediate steps are conducted at the device level of description and, thus, are conceptually meaningful, instilling confidence in the final conclusion. Additionally, the method lends itself naturally to object-oriented specification. Finally, the method facilitates the generation of qualitative justifications by tracing the sequence of operations along the activated pathways and then, using their causal or

* This work was supported in part by the National Science Foundation, Grant DCR 85-01234.

diagnostic semantics, translating them into appropriate verbal expressions.

In section 2, we shall present a brief review of belief-network formalisms, followed by an operational account of the distributed revision method established in Pearl [1986b]. In section 3, we shall demonstrate the applicability of the method to circuit diagnosis problems and illustrate the propagation process by working in detail an example discussed in Davis [1984], Genesereth [1984] and de Kleer [1986]. Section 4 discusses the problem of degeneracy (multiple best explanations) and offers a working solution.

2. Review of Belief Revision in Bayesian Networks

Bayesian belief networks are directed acyclic graphs in which the nodes represent propositional variables, the arcs signify the existence of direct causal influences between the linked propositions, and the strength of these influences are quantified by the conditional probability of each variable, given the state of its parents [Pearl 1986a].

A Bayesian network (e.g., Figure 1) provides a clear graphical representation for the essential independence relationships embedded in the underlying causal model. These independencies can be detected by the following *di-graph separation* criterion: if all paths between X_i and X_j are "blocked" by a subset S of variables, then X_i is independent of X_j, given the values of the variables in S. A path is "blocked" by S if it contains a member of S between two diverging or two cascaded arrows or, alternatively, if it contains two arrows converging at node X_k and neither X_k nor any of its descendants is in S. In particular, each variable X_i is independent of both its grandparents and its non-descendant siblings, given the values of its parents. In Figure 1, for example, X_2 and X_3 are independent, given either $\{X_1\}$ or $\{X_1, X_4\}$, because the two paths between X_2 and X_3 are blocked by either one of these sets. However, X_2 and X_3 may not be independent given $\{X_1, X_6\}$ because X_6, as a descendant of X_5, "unblocks" the head-to-head connection at X_5, thus opening a pathway between X_2 and X_3.

Once a Bayesian network is constructed, it can be used as an interpretation engine, namely, newly arriving information will set up a parallel constraint-propagation process which ripples multidirectionally through the network until, at equilibrium, every variable is assigned a value corresponding to the most likely interpretation of the data. Incoming information corresponds to direct observations which unequivocally determine the values of some variables in the network.

Let variable names be denoted by capital letters, e.g., U, V, X, Y, Z and their associated values by lower case letters, e.g., u, v, x, y, z. Incoming information will be denoted by e to connote *evidence* and will be represented by nodes whose values are held constant. Let $\mathbf{W}$ stand for the set of all variables considered, including those in e. Any assignment of values to the variables in $\mathbf{W}$ consistent with e will be called an *extension* or an *interpretation* of e. Our problem is to find an extension $\mathbf{w}^*$ which maximizes the conditional probability $P(\mathbf{w} \mid e)$. In other words, $\mathbf{W} = \mathbf{w}^*$ is the most likely interpretation of the evidence at hand if

$$P(\mathbf{w}^* \mid e) = \max_{\mathbf{w}} P(\mathbf{w} \mid e) \qquad (2)$$

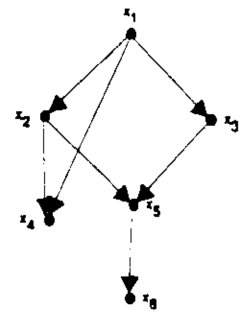


Figure 1

The task of finding $\mathbf{w}^*$ will be executed locally, by letting each variable X compute the function

$$BEL^*(x) = \max_{\mathbf{w}'_X} P(x, \mathbf{w}'_X \mid e) \qquad (3)$$

where $\mathbf{W}'_X = \mathbf{W} - X$. Thus, $BEL^*(x)$ stands for the probability of the most likely extension of e which is also consistent with the hypothetical assignment $X = x$.

The propagation scheme presented below is based on the following philosophy: For every value x of a singleton variable X, there is a best extension of the complementary variables $\mathbf{W}'_X$. Due to the many independence relationships embedded in the network, the problem of finding the best extension of $X = x$ can often be decomposed into that of finding the best complementary extension to each of the neighboring variables, then using this information to choose the best value of X. This process of decomposition (resembling the principle of optimality in dynamic programming) can be applied recursively until, at the network's periphery, we meet evidence variables whose values are predetermined, and the process halts.

The objective of achieving a globally-optimal solution by such recursive local computations can be realized if the network is singly-connected, namely, if there is only one undirected path between any pair of nodes. These include causal trees, where each node has a single parent, as well as networks with multi-parent nodes, representing events with several causal factors. We shall first review the propagation scheme in singly-connected networks and then discuss how it can be applied to general networks containing loops.

Consider an arbitrary variable X having n parents, $U_1, U_2, \ldots, U_n$, and m children, $Y_1, Y_2, \ldots, Y_m$, and imagine that node X receives from its parents the messages $\pi_X^*(u_i)$, $i = 1, \ldots, n$, and $\lambda_{Y_j}^*(x)$, $j = 1, \ldots, m$, from its children (See Figure 2).

Assume that these messages were precalculated to convey the following information:

- $\pi_X^*(u_i)$ stands for the probability of the most likely subextension of the proposition $U_i = u_i$ comprised of variables in the subnetwork on the tail side of the link $U_i \longrightarrow X$. This subextension is sometimes called an "explanation," or a "causal argument."
- $\lambda_Y^*(x)$ stands for the probability of the most likely subextension of the proposition $X = x$ comprised of variables in the subnetwork on the head side of the link $X \longrightarrow Y_j$. This subextension is sometimes called a "prognosis" or a "forecast."

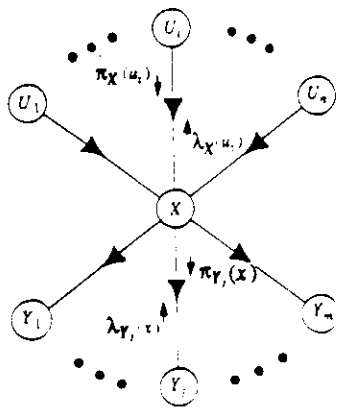


Figure 2

Using these $n+m$ messages together with the fixed probability $P(x \mid u_1, ..., u_n)$, X can identify its best value and further propagate these messages using the following computations:

Step 1: When node X is activated to update its parameters, it simultaneously inspects the $\pi_X^*(u_i)$ and $\lambda_{Y_j}^*(x)$ messages communicated by each of its parents and children and forms the product

$$F(x, u_1, ..., u_n) = \prod_{j=1}^{m} \lambda_{Y_j}^*(x) P(x \mid u_1, ..., u_n) \prod_{i=1}^{n} \pi_X^*(u_i) \quad (4)$$

This F function enables X to compute its $BEL^*(x)$ function and, simultaneously, identify the best value x^* from the domain of X:

$$x^* = \max_x{}^{-1} BEL^*(x) \quad (5)$$

where

$$BEL^*(x) = \alpha \max_{u_k:1 \le k \le n} F(x, u_1, ..., u_n) = \alpha \lambda^*(x) \pi^*(x)$$

and

$$\lambda^*(x) = \prod_{j=1}^{m} \lambda_{Y_j}^*(x)$$

$$\pi^*(x) = \max_{u_k:1 \le k \le n} P(x \mid u_1, ..., u_n) \prod_{i=1}^{n} \pi_X^*(u_i) \quad (6)$$

α is a normalizing constant, rendering $\sum_x BEL^*(x) = 1$, which need not be computed in practice.†

Step 2: Using the $BEL^*(x)$ function computed in step 1, node X computes the children-bound messages:

$$\pi_{Y_j}^*(x) = \alpha \frac{BEL^*(x)}{\lambda_{Y_j}^*(x)} \quad (7)$$

and posts these on the links to $Y_1, ..., Y_m$.

Step 3: Using the F function computed in step 1, node X computes the parent-bound messages by performing n vector maximizations, one for each parent:

$$\lambda_X^*(u_i) = \max_{x, u_k: k \ne i} [F(x, u_1, ..., u_n) / \pi_X^*(u_i)] \quad i = 1, ..., n \quad (8)$$

This concurrent message-passing process is both initiated and terminated at the peripheral nodes of the network, subject to the following boundary conditions:

† Note that x^* may not be the value which maximizes the posterior distribution $P(x \mid e)$ of the individual variable X and cannot be computed from the latter. For example, in diagnosing faulty circuits, each component is individually more likely to be operational, while the x^* assignment will categorically identify some components as faulty.

1. ***Anticipatory Node:*** representing an uninstantiated variable with no successors. For such a node, X, we set $\lambda_{Y_j}^*(x) = (1, 1, ..., 1)$.

2. ***Evidence Node:*** representing a variable with instantiated value. If variable X assumes the value x', we introduce a dummy child z with
$$\lambda_Z^*(x) = \begin{cases} 1 & \text{if } x = x' \\ 0 & \textit{otherwise} \end{cases}$$
This implies that, if X has children, $Y_1, ..., Y_m$, each child should receive the same message $\pi_{Y_j}^*(x) = \lambda_Z^*(x)$ from X.

3. ***Root Node:*** representing a variable with no parents. For each root variable X, we permanently replace $P(x, u_1, ..., u_n) \prod_{i=1}^{n} \pi_X^*(u_i)$ in Eq.(4) with the prior probability of X, $P(x)$.

In [Pearl 1986b], it is shown that, in singly-connected networks, the semantics of the messages produced via Eqs. (7 & 8) are preserved. Further, the propagation process can be activated concurrently, it subsides in time proportional to the network diameter and, at equilibrium, the x^* values chosen via Eq.(5) constitutes the optimal extension $\mathbf{w}^*$.

If the network is multiply-connected, the loops can be broken by the method of *conditioning*, also called *reasoning by assumptions*:

1) A set of nodes (called a cycle cutset) is instantiated, which renders the network singly-connected.
2) The propagation scheme is triggered to find the best extension of that instantiation and to merit it accordingly.
3) Another instantiation of the cycle cutset is assumed, and the process is repeated until the assumption with the best extension is identified.

The identification process can be performed either by enumeration or, better yet, by invoking branch-and-bound or heuristic search in the space of cutset assumptions. Note that, in the worst case, the search space is only exponential in the size of the cycle cutset and, in sparse networks, would be fairly low.

3. Example

To illustrate the scheme just described, we will consider in detail an example treated in [de Kleer 86], [Davis 84] and [Genesereth 84]. The problem is, given the digital circuit depicted in Figure 3, to find the set of malfunctioning components which most likely would have caused the observed behavior: $F=10$, $G=12$. The blocks named M_1, M_2 and M_3 are multipliers, while M_4 and M_5 are adders. The inputs appear on the left, the outputs on the right. The numbers in brackets are the expected values at potentially-observable points under no-fault conditions.

The Bayesian net corresponding to this circuit is shown in Fig.4. The nodes of this net represent both observable points and status of components. The circuit components appear as root variables constraining the relationship between input and output.

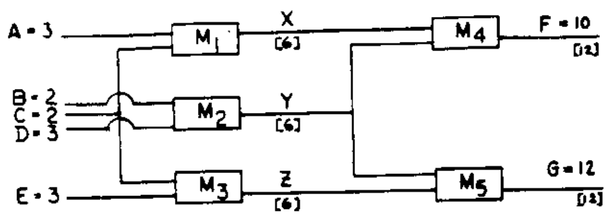


Figure 3

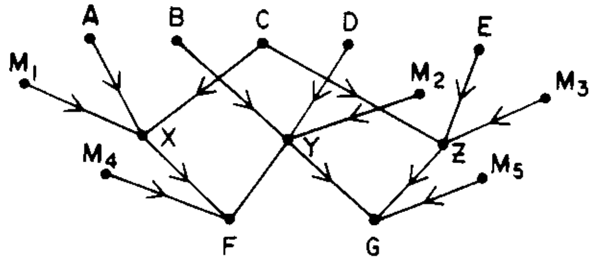


Figure 4

The link probabilities for the mapped circuit fragments depicted in Figure 5 will be given by:

$$P(x \mid I_1, I_2, M) = \begin{cases} 1 & \text{if } M = good \ \& \ x = f(I_1, I_2) \\ 0 & \text{if } M = good \ \& \ x \neq f(I_1, I_2) \\ \dfrac{1}{R_X} & \text{if } M = bad \ \& \ x \ any\ value \end{cases}$$

where R_X stands for a large constant representing the *range* of possible X's values, and $f(\cdot)$ stands for the function computed by M. This mapping conforms to the assumption normally made in circuit diagnosis: "good" behavior does not guarantee "good" components [de Kleer 1986b].

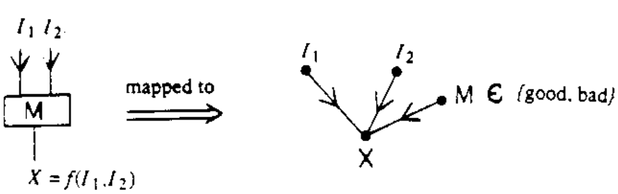


Figure 5

Initialization

Initially, the value of the input is known, and no other observation has been made. Therefore, for every input, $I \in \{A, B, C, D, E\}$, we have:

$$\pi^*(I = measured\ value) = 1$$
$$\pi^*(I \neq measured\ value) = 0.$$

The status of the components, however, is not known, but we assume *a priori*:

$$\pi^*(M_i = bad) = P_i\,,$$
$$\pi^*(M_i = good) = Q_i = 1 - P_i$$

where P_i is some small value standing for the prior probability of failure of the i-th component, and Q_i is simply an abbreviation for $1 - P_i$.

For the purpose of this example, we shall make the reasonable assumption that the prior probability of failure for components of a certain type is the same, while the probability of failure for multipliers is greater than for adders. In other words, we assume:

$$P_1 = P_2 = P_3 > P_4 = P_5$$
and
$$R_X = R_Y = R_Z = R_F = R_G\,. \tag{9}$$

We start by propagating down the effects of the priors. For example, $\pi^*(x)$ is computed from Eq.(6) of section 2:

$$\pi^*(x) = \max_{A,C,M_1} P(x \mid A, C, M_1)\, \pi_X^*(A)\, \pi_X^*(C)\, \pi_X^*(M_1).$$

Since A and C are fixed at $(A = 3, C = 2)$, $\pi^*(x)$ can be rewritten as:

$$\pi^*(x) = \max_{M_1} P(x \mid A = 3, C = 2, M_1)\, \pi^*(A = 3)\, \pi^*(C = 2)\, \pi^*(M_1)$$

where we have replaced the $\pi_X^*(\cdot)$ with $\pi^*(\cdot)$. For $x = 6$, and assuming $Q_1 >> P_1 / R_X$, the maximum will be achieved with $M_1 = good$. For $x \neq 6$, the maximum is achieved with $M_1 = bad$ because, according to the link probabilities, it is impossible to have a multiplier working correctly with inputs 3 and 2 and output different from 6.

Using similar arguments, we compute the following parameters:

$$\pi^*(x) = \begin{cases} Q_1 & x = 6 \\ P_1 / R_X & x \neq 6 \end{cases}$$

$$\pi^*(y) = \begin{cases} Q_2 & y = 6 \\ P_2 / R_Y & y \neq 6 \end{cases}$$

$$\pi^*(z) = \begin{cases} Q_3 & z = 6 \\ P_3 / R_Z & z \neq 6 \end{cases}$$

The rest of the messages are computed as follows. Since F and G are anticipatory nodes, we have:

$$\lambda^*(F_i) = \lambda^*(G_i) = 1,$$

where F_i and G_i range over all the possible values of F and G, respectively.

The message $\lambda_G^*(y)$ can be computed from Eq.(8):

$$\lambda_G^*(y) = \max_{G,z,M_5} P(G \mid y, z, M_5)\, \pi_G^*(M_5)\, \pi_G^*(z).$$

Independently of the value of y, the maximum will always be achieved by choosing $G = y + z$, $z = 6$ and $M_5 = good$. This yields

$$\lambda_G^*(y) = Q_5 Q_3\,,$$

which is equivalent to

$$\lambda_G^*(y_i) = 1\,,$$

for y_i ranging over all the possible values of y. The same holds for $\lambda_G^*(z)$, $\lambda_F^*(y)$ and $\lambda_F^*(x)$.

Returning to the top-down propagation, we can now compute $\pi^*(F)$:

$$\pi^*(F=12) = \max_{x,y,M_4} P(F=12 \mid x, y, M_4)\, \pi_F^*(x)\, \pi_F^*(y)\, \pi_F^*(M_4)$$
$$= \max_{x,y,M_4} P(F=12 \mid x, y, M_4)\, \pi^*(x)\, \pi^*(y)\, \lambda_G^*(y)\, \pi^*(M_4),$$

component status s	$\pi^*(s)$ priors	$\lambda^*(s)$ after observing $F=10$ and $G=12$	$\lambda^*(s)$ after observing $F, G, X=6, Y=4$ and $Z=6$
$M_1 = good$	Q_1	$\frac{P_4}{R_F} Q_2 Q_3 Q_5$	$\frac{P_2 P_5}{R_Y R_G} Q_3 Q_4$
$M_1 = bad$	P_1	$\frac{1}{R_X} Q_2 Q_3 Q_4 Q_5$	$\frac{1}{R_X} \frac{P_2 P_5}{R_Y R_G} Q_3 Q_4$
$M_2 = good$	Q_2	$\frac{P_1}{R_X} Q_3 Q_4 Q_5$	0
$M_2 = bad$	P_2	$\frac{1}{R_Y} \frac{P_3}{R_Z} Q_1 Q_4 Q_5$	1
$M_3 = good$	Q_3	$\frac{P_1}{R_X} Q_2 Q_4 Q_5$	$\frac{P_2 P_5}{R_Y R_G} Q_1 Q_4$
$M_3 = bad$	P_3	$\frac{1}{R_Z} \frac{P_2}{R_Y} Q_1 Q_4 Q_5$	$\frac{1}{R_Z} \frac{P_2 P_5}{R_Y R_G} Q_1 Q_4$
$M_4 = good$	Q_4	$\frac{P_1}{R_X} Q_2 Q_3 Q_5$	$\frac{P_2 P_5}{R_Y R_G} Q_1 Q_3$
$M_4 = bad$	P_4	$\frac{1}{R_F} Q_1 Q_2 Q_3 Q_5$	$\frac{1}{R_F} \frac{P_2 P_5}{R_Y R_G} Q_1 Q_3$
$M_5 = good$	Q_5	$\frac{P_1}{R_X} Q_2 Q_3 Q_4$	0
$M_5 = bad$	P_5	$\frac{1}{R_G} \frac{P_2}{R_Y} Q_1 Q_3 Q_4$	1

Figure 7. Pattern of messages at component-nodes.

For completeness, the rightmost column shows the λ^*'s resulting from the additional observations $x=6$, $y=4$ and $z=6$. The π^*'s messages do not change. The reader may verify that with the new information, the best diagnosis establishes that components M_2 and M_5 are faulty with certainty.

4 The Degenerate Case

When there are several optimal extensions, the following problem may develop. Consider the circuit shown in Figure 8 below, comprised of two buffers, D_1 and D_2. When operating properly, the output reproduces the input. When malfunctioning, the output can be of any value, each with very low probability.

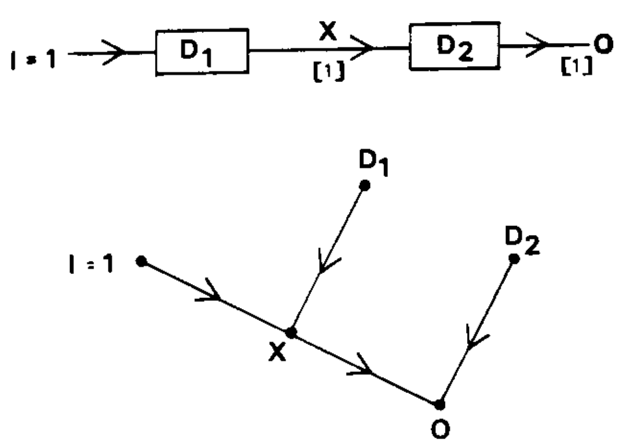


Figure 8. The Circuit and its Bayesian Net Representation

Suppose the output O is found to be of value 2, indicating at least one faulty component. Executing the propagation scheme yields the following set of final messages:

$$\lambda^*(D_1 = good) = P_2/R_O \qquad \pi^*(D_1 = good) = 1 - P_1$$

$$\lambda^*(D_1 = bad) = 1 - P_2/R_X \qquad \pi^*(D_1 = bad) = P_1$$

$$\lambda^*(D_2 = good) = P_1/R_X \qquad \pi^*(D_2 = good) = 1 - P_2$$

$$\lambda^*(D_2 = bad) = 1 - P_1/R_O \qquad \pi^*(D_2 = bad) = P_2$$

Assuming $P_1 = P_2 = P$ and $R_X = R_O = R$, we obtain:

$$BEL^*(D_1 = good) = BEL^*(D_1 = bad) =$$
$$BEL^*(D_2 = good) = BEL^*(D_2 = bad) = \alpha \frac{1}{R} P\,(1-P),$$

where α is a constant. What, then, are the values corresponding to D_1^* and D_2^*? In other words, what is the best interpretation? The fact that there are two equally-likely interpretations, $\{D_1^* = good, D_2^* = bad\}$ and $\{D_1^* = bad, D_2^* = good\}$, makes it impossible to unequivocally assign a label to either D_1^* or D_2^*. Breaking the ties arbitrarily might result in a wrong extension being chosen, e.g., $\{D_1^* = good, D_2^* = good\}$ or $\{D_1^* = bad, D_2^* = bad\}$. In general, by "breaking the ties" arbitrarily, we may end up choosing instantiations that, while belonging to equally-meritorious extensions, do not belong to the *same* one!

To avoid such bad choices, we must devise a mechanism to enforce a selection of values within the *same* optimal extension; tie-breaking should also take into account all instantiations made previously. Recalling that

$$BEL^*(x) = \alpha\, \lambda^*(x)\, \pi^*(x)$$

$$\lambda^*(x) = \max_{w_X^-} P(w_X^- \mid x),$$

$$\pi^*(x) = \max_{w_X^+} P(x, w_X^+),$$

where w_X^- and w_X^+ are the subextensions including X's parents and children, respectively, we see that it is only necessary to uncover the w_X^- and w_X^+ involved in the computations of $\lambda^*(x)$ and $\pi^*(x)$. This can be achieved by saving, for each node X in the net, only those neighbors' values at which the maximization is achieved. In other words, if node U has a neighborhood as the one depicted in Fig.9 (a), and

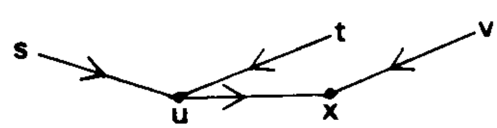
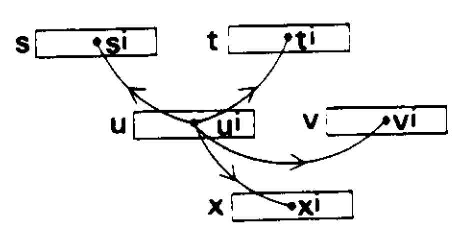


Figure 9. *(a) A Net Fragment*



(b) The Pointer Structure Saved

$$\lambda_X^*(u_i) = \max_{v, x} P(x \mid v, u_i)\, \pi_X^*(v)\, \lambda^*(x)$$
$$= P(x^i \mid v^i, u_i)\, \pi_X^*(v^i)\, \lambda^*(x^i)$$

where all the parameters are known. The maximum is at the nominal values of X, Y and M_4:

$$\pi^*(F=12) =$$

$$= P(F=12 \mid x=6, y=6, M_4=good)\, \pi^*(x=6)\, \pi^*(y=6)\, \pi^*(M_4=good)$$

$$= Q_1 Q_2 Q_4$$

For $F \neq 12$ we get:

$$\pi^*(F \neq 12) = \max \left[Q_1 \frac{P_2}{R_Y} Q_4 , \frac{P_1}{R_X} Q_2 Q_4 , Q_1 Q_2 \frac{P_4}{R_F} \right] ,$$

where the three alternatives correspond to the failure of M_2, M_1 and M_4, respectively. Since we are assuming $P_4 < P_1 = P_2$ and $R_X = R_Y = R_F$, we can eliminate the third alternative. A similar procedure is used to compute $\pi^*(G)$. To find out which extension is the best at this point, we compute $BEL^*(N)$ for each node N and select for it the value N^* at which $BEL^*(N)$ peaks.† Of course, the answer, under no observed failure, will be that all components are operating properly.

Fault Interpretation

When the first two tests are made, finding $F = 10$ and $G = 12$, new messages begin to propagate concurrently along the links. For simplicity, we will follow only those messages that will affect the optimal label for M_4, i.e., those passing through the path darkened in Figure 6.

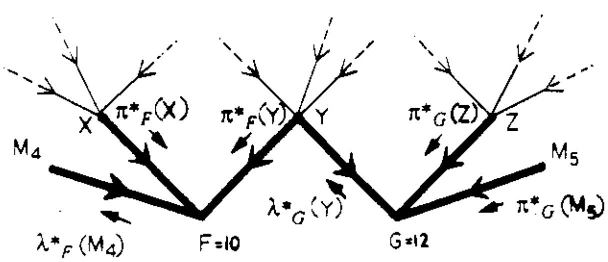


Fig. 6. Messages affecting the label of M_4, after G and F are observed.

The message $\lambda_G^*(y)$ is computed from:

$$\lambda_G^*(y) = \max_{M_5, z} P(G = 12 \mid y, z, M_5)\, \pi_G^*(z)\, \pi_G^*(M_5).$$

For $y = 6$, the maximum is achieved at $z = 6$ and $M_5 = good$, resulting in:

$$\lambda_G^*(y=6) = Q_3 Q_5 .$$

The case $y \neq 6$ requires $z \neq 6$ or $M_5 = bad$. Since $\pi_G^*(z \neq 6) = P_3$, $\pi_G^*(M_5 = bad) = P_5$ and $P_3 > P_5$, we get:

$$\lambda_G^*(y \neq 6) = P_3 Q_5 .$$

Now Y computes messages for its neighbors and sends to F the message $\pi_F^*(y)$ computed as

$$\pi_F^*(y) = \pi^*(y)\, \lambda_G^*(y) .$$

The message that M_4 receives is computed from:

$$\lambda_F^*(M_4) = \max_{x,y} P(F = 10 \mid x, y, M_4)\, \pi_F^*(x)\, \pi_F^*(y)$$

$$= \max_{x,y} P(F = 10 \mid x, y, M_4)\, \pi^*(x)\, \pi^*(y)\, \lambda_G^*(y)$$

† — For ways of dealing with multiple solutions, see section 4.

For $M_4 = good$, the maximum must occur when either $x = 4$ and $y = 6$, or $x = 6$ and $y = 4$, i.e.,

$$\lambda_F^*(M_4 = good) = \max \{ \pi^*(x=4)\, \pi^*(y=6)\, \lambda_G^*(y=6),$$

$$\pi^*(x=6)\, \pi^*(y=4)\, \lambda_G^*(y=4) \}$$

$$= \max \left\{ \frac{P_1}{R_X} Q_2 Q_3 Q_5 , Q_1 \frac{P_2 P_3}{R_Y R_Z} Q_5 \right\}$$

which, under the assumptions in (9), becomes:

$$\lambda_F^*(M_4 = good) = \frac{P_1}{R_X} Q_2 Q_3 Q_5 .$$

This reflects the fact that the most likely interpretation constrained by $M_4 = good$ singles out M_1 as the only faulty component.

For $M_4 = bad$, the maximum is achieved with $x = y = 6$, resulting in:

$$\lambda_F^*(M_4 = bad) = \frac{1}{R_F} Q_1 Q_2 Q_3 Q_5 ,$$

since $M_4 = bad$ alone explains the observed behavior.

At this point, we can compute $BEL^*(M_4)$ to find M_4^*, i.e., the believed status of M_4:

$$BEL^*(M_4) = [\, BEL^*(M_4 = good), BEL^*(M_4 = bad)]$$

$$= \alpha [\, \lambda^*(M_4 = good)\, \pi^*(M_4 = good),$$

$$\lambda^*(M_4 = bad)\, \pi^*(M_4 = bad) \,].$$

therefore, M_4^* is obtained as:

$$M_4^* = \max_{M_4}^{-1} BEL^*(M_4)$$

$$= \max_{M_4}^{-1} \begin{cases} \dfrac{P_1}{R_X} Q_2 Q_3 Q_4 Q_5 & \text{if } M_4 = good \\ \dfrac{P_4}{R_F} Q_1 Q_2 Q_3 Q_5 & \text{if } M_4 = bad \end{cases}$$

$$= good$$

since P_1 was assumed to be greater than P_4.

It is remarkable to note that, even at this early stage of the propagation, we can already label M_4 as *good*, and be confident that this label will remain part of the globally-optimal diagnosis. Apparently, the λ^* message arriving at M_4 already contains a summary of global information (gathered during the initialization phase) which is sufficient to alert M_4 to the existence of a more likely culprit -- the multiplier M_1.

Pursuing the propagation through the rest of the network, the optimal status of all the other components is determined. The resulting pattern of messages for all component-nodes is depicted in the three leftmost columns of Figure 7.

The optimal status of any component-node can be determined by simply comparing the $\lambda^* \cdot \pi^*$ product for each of its possible status. For example, to determine the optimal status of M_1, we compare the product $\frac{P_4}{R_F} Q_2 Q_3 Q_5 Q_1$ with $\frac{1}{R_X} Q_2 Q_3 Q_4 Q_5 P_1$, and conclude that the optimal status of M_1, M_1^*, is *bad*, since the first term is greater than the second under the current assumptions involving P's and R's values.

$$\pi^*(u_i) = \max_{s,t} P(u_i \mid s, t)\, \pi_U^*(s)\, \pi_U^*(t)$$
$$= P(u_i \mid s^i, t^i)\, \pi_U^*(s^i)\, \pi_U^*(t^i),$$

we would save the pointer structure depicted in Figure 9 (b), where the arrows $a_i \rightarrow b_j$ mean that b_j participates in the best extension constrained by $a = a_i$. Since $\lambda_X^*(u_i)$ is not computed at node U but at node X, to keep the computations local, X will additionally send to U the message $[(x^i, v^i) \rightarrow u_i]$ for each state u_i of U.

Having this local information available at every node not only solves the ambiguity problem, but also provides a mechanism for *retrieving* an optimal extension constrained by any instantiation of the individual variables. In particular, to get the overall optimal extension, we need solve only a single maximization for $BEL^*(x)$ for some X and then recursively follow the pointers attached to x^*. It is also possible to obtain, without extra effort, the second-best overall extension [Geffner & Pearl 1986].

A second "plus" of this mechanism is facilitating sensitivity analysis; to analyze the merit of observing an unknown variable, we can simply follow the links attached to each of its possible instantiations and examine its impact on other propositions in the system.

5 Conclusions

We have illustrated the applicability of the scheme proposed in [Pearl 1986b] to the problem of diagnosing multiple faults. The scheme uses the conditional independence properties embedded in the system to decompose the task of finding a best overall interpretation into smaller sub-tasks of finding the best interpretations for subparts of the net, then combining them together. This decomposition yields a global optimal diagnosis by local and concurrent computations using a message-passing scheme. Contrasting the exponential complexity associated with unaided diagnostic methods, the proposed scheme offers a drastic reduction to linear complexity for singly-connected networks and $\exp(\mid cycle\text{-}cutset \mid)$ in multiply-connected networks.

In this paper we do not address ways in which the proposed scheme can be used effectively in the selection of tests or to determine when an acceptable diagnosis has been found. For these and related topics, the reader might refer to [Geffner & Pearl 1986].

REFERENCES

Ben Bassat, M., Carlson, R. W., Puri, V. K., Lipnick, E., Portigal, L. D. & Weil, M. H. [1980]. "Pattern-Based Interactive Diagnosis of Multiple Disorders: The MEDAS System," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, PAMI-2:148-160.

Buchanan, B. G. & Shortliffe, E. H., (Eds.) [1984]. *Rule-Based Expert Systems*, Addison-Wesley, Reading, Massachusetts.

Cooper, G. F., [1984]. "NESTOR: a Computer-Based Medical Diagnostic Aid that Integrates Causal and Probabilistic Knowledge," Ph.D. Dissertation, Department of Computer Science, Stanford University, Stanford, CA.

Davis, R., [1984]. "Diagnostic Reasoning Based on Structure and Behavior," *Artificial Intelligence Journal*, 24.

de Kleer, J. & Williams, B. C., [1986a]. "Reasoning about Multiple-Faults," *Proceedings, AAAI-86*, 5th Nat'l. Conf. on AI, Phila., PA., pp. 132-139.

de Kleer, J. & Williams, B. C., [1986b]. "Diagnosing Multiple Faults," *Artificial Intelligence*, forthcoming.

Geffner, H. & Pearl, J., [1986]. "A Distributed Approach to Diagnosis," UCLA CSD Cognitive Systems Laboratory *Technical Report (R-66)*.

Genesereth, M. R., [1984]. "The Use of Design Descraiptions in Automated Diagnosis," *Artificial Intelligence Journal*, 24.

Pearl, J., [1986a]. "Fusion, Propagation and Structuring in Belief Networks," UCLA CSD *Technical Report 850022 (R-42)*; also in *Artificial Intelligence*, September 1986, Vol. 29, No. 3, 241-288.

Pearl, J., [1986b]. "Distributed Revision of Belief Commitment in Multi-Hypotheses Interpretation," *Proceedings*, 2nd AAAI Workshop on Uncertainty in Artificial Intelligence, August 1986, Philadelphia, PA., pp. 201-209; also in *Artificial Intelligence*, forthcoming.

Peng, Y. & Reggia, J., [1986]. "Plausibility of Diagnostic Hypotheses," *Proceedings, AAAI-86*, 5th Nat'l. Conf. on AI, Philadelphia, PA., Aug. 11-15, 1986, pp. 140-145.

Pople, H. [1982]. "Heuristic Methods for Imposing Structures on Ill-Structured Problems," in Solovits, P. (Ed.), *AI in Medicine*, Westview Press, Boulder, Colorado, pp. 119-190.

Reggia, J. A., Nau, D. S. & Wang, Y. [1983]. "Diagnostic Expert Systems Based on a Set-Covering Model," *Int'l. Journal of Man-Machine Studies*, 19, pp. 437-460.

Reiter, R., [1985]. "A Theory of Diagnosis from First Principles," *Technical Report TR-187/86*, Computer Science Department, University of Toronto.