BAYESIAN DIAGNOSIS AND PROGNOSIS USING INSTRUMENT UNCERTAINTY

John W. Sheppard, PhD
ARINC Incorporated
2551 Riva Road
Annapolis, MD 21401
(410) 266-2099
john.sheppard@arinc.com

Mark A. Kaufman
NSWC Corona Division
PO Box 5000
Corona, CA 91718
(951) 273-5725
mark.kaufman@navy.mil

Abstract–How can diagnosis and prognosis systems be improved in the presence of uncertainty in test results? How can these uncertainties be identified and modeled? Can diagnosis be improved as a result of understanding these uncertainties? These questions represent the core problems to be explored in this paper. Specifically, we explore the process by which instrument uncertainty can be used to determine conditional probabilities of potential diagnoses given test results generated by these instruments. We then use that information to construct "Bayesian belief networks" with specific goal of maximizing diagnostic accuracy while minimizing construction complexity, and computational complexity. We then extend the ideas presented for Bayesian diagnosis to the prognostic, or predictive diagnostic, problem.

INTRODUCTION

Significant emphasis is being placed on developing new approaches to analyzing test results and isolating faults by procurers of complex systems. The field of Bayesian reasoning has been providing successful approaches to solving complex problems such as diagnosis since the eighteenth century; however, the computational complexity and data requirements associated with Bayesian techniques have tended to stifle their use in system diagnosis.

Our objective is to describe a *general* approach to applying Bayesian techniques to the system diagnosis problem. Key to this approach is a) developing technology-independent techniques for collecting the data required for creating Bayesian models (i.e., Bayesian networks) and b) identifying efficient techniques for reasoning with these models. For most complex systems, sufficient data to model probabilities to high levels of accuracy do not exist, either because the systems have limited field data available or the systems implement new designs. In addition, general reasoning in Bayesian networks is known to be intractable (technically *NP*-hard); therefore, if the desire is to achieve "exact" reasoning with Bayesian networks, then it is necessary to constrain the models to fit particular topological forms for which efficient computation is possible.

In this paper, we propose an approach to capturing probability information from data collected on test instruments during the calibration process, combined with data required to set decision thresholds when designing the tests that use these instruments. We then describe a general diagnostic model that is efficient to process computationally. We review the fundamentals of Bayes decision theory to demonstrate how to use these models in a computational setting and illustrate the approach with a numerical example. We end by extending the model to address issues in predictive diagnosis (i.e., prognosis) as well.

ASSUMPTIONS

Test equipment uncertainty is believed to be a significant predictable source of uncertainty in our model. We differentiate between the power source uncertainty (i.e., the probability that power levels applied as input to the UUT are at the intended levels) and instrument uncertainty (i.e., the probability that measurements and stimuli are within some specified tolerance). When considering power source uncertainty, we assume that if the power system integrity check passes (with some confidence level associated with it), then either we can factor power uncertainty out of the model, or the resulting confidence can be used as input to an overall probability of error. We will assume

power uncertainty is negligible due to the fact the power source is verified or calibrated prior to any measurement being made [2].

For instrument uncertainty, the calibration process is designed to a) minimize error in the instrument, as well as b) determine the level of uncertainty (by reading as opposed to by scale) of the instrument. Thus, assuming the instrument is calibrated, probability of error can be determined based on this specified uncertainty. Assuming that the causes for power source uncertainty and instrument uncertainty are independent and the level of uncertainty due to power source integrity is *not* negligible, the resulting probabilities can simply be multiplied together.

We are assuming that loading effects due to the test set and the switching matrix, up to the connectors on the test set side of the interface device (ID), are factored out as a result of metrology specifications relative to the test equipment and that the measurement uncertainty associated with a particular measurement device is specified at the interface rather than at the device itself. Thus we only need to consider loading effects arising from the pathways between the ID and the point of measurement on the UUT. Further, we assume that any loading effects will result in a simple shift in the measurment. As with any other factor, there will be some level of uncertainty associated with this offset. We assume that the resulting probability of error for the offest is independent of error due to measurement and is also negligible. If the error is significant, then we can include its probability as another multiplier in the model.

Finally, we analyze the effects on setting the PASS/FAIL threshold. The task of setting the threshold is one of optimizing risk, where frequently this is further subdivided into consumer and producer risk. Consumer risk refers to the effects of missed detection, and producer risk refers to the effects of false alarm.

INSTRUMENT UNCERTAINTY

In support of diagnosis, our interest is in determining the probability of some observation, given the underlying state of the system. For example, $\Pr(o(P) \mid P)$ and $\Pr(o(F) \mid F)$ would tell us the probability of observing PASS given the unit is fault free and the probability of observing a FAIL given the unit is faulty respectively[1]. Note that these probabilities work from the discrete random variables of unit state and test outcome; however, the test outcome, ultimately, depends on the underlying measurement. So consider the case where a measurement y is taken and $y \in \pm z$ (i.e., the measurement indicates a PASS observation). The probability of the test actually PASSing given the measurement y is given as,

$$\Pr(P \mid y) = \Pr(P \mid o(P)) \Pr(o(P) \mid y) + \Pr(P \mid o(F)) \Pr(o(F) \mid y). \quad (1)$$

However, when conditioning on observing a PASS, we note that the second term drops out since $\Pr(o(F) \mid y) = 0$. Also, since $\Pr(o(P) \mid y) = 1$, we have,

$$\Pr(P \mid y) = \Pr(P \mid o(P)). \quad (2)$$

Similarly,

$$\Pr(F \mid y) = \Pr(F \mid o(F)). \quad (3)$$

Continuing in this line of reasoning, observe that the probability of taking a particular measurement y given the unit passes can be computed by considering the distribution around y limited to the nominal range.[2] In other words,

$$\Pr(y \mid P) = \frac{1}{u_c\sqrt{2\pi}} \int_{-z}^{z} \exp\left[-\tfrac{1}{2}\left(\tfrac{x-y}{u_c}\right)^2\right] dx. \quad (4)$$

From the principles of probability, we also know that $\Pr(y \mid P) = 1 - \Pr(y \mid F)$.

Note that these equations assume the availability of a specific measured value (i.e., y) that would vary the associated probabilities; however, fixed probabilities are usually used in a Bayesian model. One approach is to use the expected value for the appropriate range. Another (and the one

[1] In our mathematical formulation, we make assumptions on the interpretation of events P/F and $o(P)/o(F)$ that may be different from traditional "measurement-based" assumptions. Specifically, we assume a P or F event corresponds to the Boolean event of the unit being fault-free or faulty respectively and do not tie this event to any specific "performance-based" value. In addition, we assume an $o(P)$ or $o(F)$ event corresponds to the Boolean event of mapping a specific, known measurement value (which we denote by the variable y) to a test outcome of PASS or FAIL respectively.

[2] Usually, this would be restricted to the dynamic range of the instrument; however, we assume the tests have been defined properly such that the measurements of a particular test are of consistent accuracy. Therefore, the number of intolerance measurements having different accuracy would be insignificant. This allows us to "ignore" the tails of the distribution as being negligible in cumulative probability.

we adopt) is to select a "worst-case" value. In this case, such a value would correspond to one that is coincident with one of the limits[3], z (since this would maximize probability of false alarm (false detection) or non-detection).[4] Without loss of generality, focus on $\Pr(P \mid y) = \Pr(P \mid o(P))$ (from Equation 2). If we apply Bayes rule and reduce, we get,

$$\Pr(P \mid o(P)) = (1 - \Pr(F))\left[\frac{1}{u_c\sqrt{2\pi}}\int_{-z}^{z}\exp\left[-\frac{1}{2}\left(\frac{x-z}{u_c}\right)^2\right]dx\right]. \quad (5)$$

Following a similar line of reasoning, we can then derive $\Pr(F \mid o(F))$ as

$$\Pr(F \mid o(F)) = \Pr(F)\left[1 - \frac{1}{u_c\sqrt{2\pi}}\int_{-z}^{z}\exp\left[-\frac{1}{2}\left(\frac{x-z}{u_c}\right)^2\right]dx\right]. \quad (6)$$

In Equations 5 and 6, we note that $\Pr(F)$ is derived from the failure rate of the unit and is assumed to be driven by an exponential distribution. Given this, we now have all of the pieces required to determine test confidence for Bayesian diagnosis.

An alternative approach to deriving these probabilities is to note that the probabilities of the state of the UUT and the measurement of the test both follow normal distributions that can be modeled jointly using the bivariate normal distribution [5], [9].

If we consider the distribution around x_1 to be the expected value of the state of the UUT under nominal conditions, and the distribution around x_2 to be the expected value of the instrument regardless of which UUT is being measured, then we note that these distributions are independent. Thus, $P(x_1, x_2) = P(x_1)\,P(x_2)$ and the bivariate distribution reduces to the form of Equations 5 and 6.

BAYESIAN NETWORKS

Previously, we provided a detailed derivation of a simple model for Bayesian diagnosis [8] and [9]. Here we summarize these results by pointing to two specific issues in Bayesian networks. First, determining the appropriate structure of the network can be difficult in that it requires a detailed understanding of the random variables of the problem to be solved and the conditional probability relationships between those random variables. In fault diagnosis, the random variables correspond to the tests and diagnoses, so the first step of identifying the random variables is relatively straightforward.

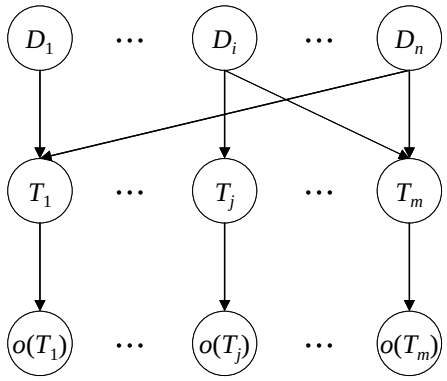

Fig. 1. Diagnostic Bayesian network structure.

Determining the appropriate conditional probabilities has been more problematic. Note that we assume we are able to determine which fault is detected by which test and that the tests were designed with such detection in mind. Thus we do not need to determine the actual relationships. We only need to consider the probabilities on those relationships. But that is exactly what the discussion on instrument uncertainty above provided. Applying the method described above for determining probability of false alarm or probability of missed detection, we can use these probabilities directly.

The second problem to be addressed is the computation required for performing diagnosis with these networks. In general, exact inference in Bayesian networks is *NP*-hard, meaning that no efficient algorithm is known to exist. However, we proposed a specific architecture to the network that reduces the computational problem with polynomial complexity. Specifically, we assume the random variables in **D** (i.e., the diagnoses) are independent, as are the random variables in **T** (i.e., the tests). Now the characteristics of conditional independence allow for simple propagation of the probabilities from the tests to the diagnoses.[5]

[3] Technically, since we do not know *a priori* which limit would be exceeded, we should take the expected value, i.e., the average in this case, of the two limit values. With symmetric distributions, however, this is not necessary, and we can work with one of the limits chosen *without loss of generality*.

[4] Should historical data be available, it might be possible to refine these probabilities based on that history, but such issues of diagnostic "maturation" are beyond the scope of this paper.

[5] In the event additional dependence relationships need to be defined, then propagation of probabilities along these dependence links must also be accounted for. The theory of Bayesian networks allows for this; however, such additional dependencies may lead to large internal cycles, which are

Given the conditional independence of the diagnoses, we can then compute the posterior probabilities of each of the diagnoses given the test results as follows. First, we will assume that we are using the network form presented in Fig. 1 and partition the random variables into three sets: **D** (the diagnoses), **T** (the true test states), and **O** (the test observations). The evidence variables will be restricted to **O**.

$$\begin{aligned} \Pr(D_i \mid \mathbf{O}) &= \alpha \Pr(\mathbf{O} \mid D_i) \Pr(D_i) \\ &= \alpha \Pr(D_i) \sum_{T_j \in \mathbf{T}} \Pr(o(T_j) \mid T_j) \Pr(T_j \mid D_i) \end{aligned} \quad (7)$$

Here, α is a normalizer over the set **D**, equal to

$$\alpha = \sum_{D_i \in \mathbf{D}} \Pr(D_i) \sum_{T_j \in \mathbf{T}} \Pr(o(T_j) \mid T_j) \Pr(T_j \mid D_i)\cdot$$

Assuming we are able to generate the probability distributions for nominal and faulty behavior, we consider the effects of locating the decision boundaries. For this discussion, we will draw on results from Bayes decision theory and its derivative, signal detection theory [4]. In particular, we consider each diagnosis to be a separate classification. In this case, diagnosis reduces to assigning the class label corresponding to the maximum *a posteriori* probability. More formally,

$$h_{MAP} = \arg\max_{\omega \in \Omega} [\lambda_\omega \Pr(x \mid \omega) \Pr(\omega)] \quad (8)$$

where ω represents the "actual state" of the unit being tested and λ_ω is the loss associated with ω being the incorrect classification. The resulting classification is referred to as the *maximum a posteriori hypothesis*. In other words, the diagnosis yielding the highest posterior probability is proposed as the most probable fault.

Observe that $\Pr(T_j \mid D_i) \in \{0, 1\}$ as described earlier, so the members of the sum are restricted only to those tests that observe D_i. Then we only need to consider $\Pr(D_i)$, which corresponds to the prior probability for D_i based on failure rate, and $\Pr(o(T_j) \mid T_j)$, which corresponds to the confidence value assigned to the observed test result. Using the Baye's maximum *a posteriori* hypothesis, we determine the most likely diagnosis simply as

conditions under which the NP-hardness of Bayesian inference becomes problematic.

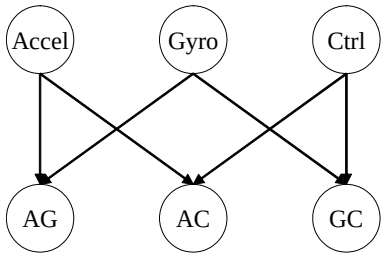

Fig. 2. Simple stability augmentation system BBN.

$$D_{MAP} = \arg\max_{D_i \in \mathbf{D}} \{\Pr(D_i \mid \mathbf{O})\}\cdot \quad (9)$$

In other words, we provide the most probable diagnosis as a means of minimizing expected error (i.e., risk) in the diagnostic process.

A NUMERICAL EXAMPLE

In the following, we present a small numerical example. Suppose, for the sake of discussion, that we are considering the Built-in Test (BIT) from the stability augmentation system (SAS) of a helicopter. Stability augmentation systems provide stability control for the three axes of the aircraft, namely roll, pitch, and yaw. Without loss of generality, we will consider just the roll axis. In evaluating the performance of the roll stability control in the SAS, we consider the health of at least three components: the roll control unit, the roll gyro, and an accelerometer.

For our test scenario, we note that if the expected output of the control unit agrees with the actual, derived roll outputs from the accelerometer and roll gyro, then the system is functioning properly. On the other hand, if any two of these three elements disagree, a fault exists in one of the two units involved in the disagreement. This scenario can be represented with the Bayesian network shown in Fig. 2. Note that we are using a simplified form of the network based on the assumption the tests are designed correctly to evaluate the indicated faults.

For our network, Accel, Gyro, and Ctrl correspond to the diagnoses of whether the accelerometer, gyro, or control unit is faulty respectively. AG represents the observation associated with comparing the accelerometer output with the gyro output. AC compares the accelerometer output with the control output, and GC compares the gyro output with the control output.

Table 1. SAS conditional probability table

Pr(o(AC) | Accel, Ctrl)	0.975	Pr(o(AC) | ¬Accel, Ctrl)	0.867
Pr(o(AC) | Accel, ¬Ctrl)	0.867	Pr(o(AC) | ¬Accel, ¬Ctrl)	0.292
Pr(o(AG) | Accel, Gyro)	0.823	Pr(o(AG) | ¬Accel, Gyro)	0.756
Pr(o(AG) | Accel, ¬Gyro)	0.756	Pr(o(AG) | ¬Accel, ¬Gyro)	0.664
Pr(o(CG) | Ctrl, Gyro)	0.905	Pr(o(CG) | ¬Ctrl, Gyro)	0.787
Pr(o(CG) | Ctrl, ¬Gyro)	0.787	Pr(o(CG) | ¬Ctrl, ¬Gyro)	0.521

Given this network structure, the next step is to define the conditional probability tables on the dependence links and the prior probabilities on the diagnoses. The prior probabilities are based upon the failure probabilities of the corresponding units. Suppose the SAS has been operating without failure for 250 hours and the failure rates (assuming per thousand hours) for the accelerometer, gyro, and control unit are 30, 10, and 2 respectively. Since we must also include the absence of a fault in our probability calculations, we assign a "failure rate" of no-fault to be 958. Then, at this particular time, the relative failure probabilities [9] would be 0.301872, 0.277246, 0.118843, and 0.302039 respectively.[6]

To determine the conditional probabilities, we must first consider the instrument uncertainty. For the sake of simplicity, we will only derive the conditional probability table for AG. Without loss of generality, let o(AG) represent an observation that AG fails and ¬o(AG) represent an observation that AG passes. Similarly, let Accel and Gyro represent the logic states that the accelerometer or gyro have failed, respectively. Thus, ¬Accel would correspond to the accelerometer being fault free (likewise for ¬Gyro).

Given the absence of a measurement when first constructing the model, we assume the measurement occurs at the value of a decision limit for the test (i.e., z). Thus, we set $y = +z$ (since we "normalized" the distribution to have zero-mean). What is interesting about our assumption, however, is that the mean of the distribution is shifted to the limit, so all probabilities of failure given the corresponding single faults reduce to 0.5. Then, when considering the combined fault, we have

$$\Pr(o(\text{AG}) \mid \text{Accel}) = \Pr(o(\text{AG}) \mid \text{Gyro}) = 0.579$$

$$\Pr(o(\text{AG}) \mid \text{Accel, Gyro}) = 1 - (1 - \Pr(o(\text{AG} \mid \text{Accel}))(1 - \Pr(o(\text{AG}) \mid \text{Gyro})) = 0.823.$$

These probabilities would need to be adjusted based on the measurement uncertainty (defining the shape of the distribution) if assumptions other than the limits were chosen. Using the above approach, we can construct the conditional probability table given in Table 1.

Using this approach of setting an expected measured value on the relevant side of the test limit, it is interesting to note that the probabilities of failure for each of the units (given no evidence) is {Accel: 0.302; Ctrl: 0.277, Gyro: 0.119; NF: 0.216}, which correspond to the failure probabilities (except for No Fault—for the Bayesian network, the probability for No Fault is given as Pr(NF | Accel, Ctrl, Gyro) and is derived from the fact that this probability is zero if any fault exists. Thus the probability derived from the "failure rate" for No Fault is not required, except to determine the prior probabilities for the faults in the system. Suppose, we indicate that AC and AG both fail but CG passes. Logically, we would expect Accel to be faulty, and indeed, we find revised probabilities of {Accel: 0.539; Ctrl: 0.069, Gyro: 0.230; NF: 0.000}. Thus we would conclude from the tests that Accel is the most likely to have failed.

DYNAMIC BAYESIAN NETWORKS

The traditional approach to fault diagnosis assumes tests are applied at a specific point in time from which one can infer the condition of the system under test and make a diagnosis. The problem of prognosis, while essentially an extension of diagnosis, is complicated by the fact that time becomes a significant factor in the analysis. In fact, one can represent the prognosis problem as a time series prediction problem in which one attempts to infer a future state from some sequence of past states.

[6] The high number of significant digits is provided to demonstrate the subtle differences in probabilities at this point in the analysis.

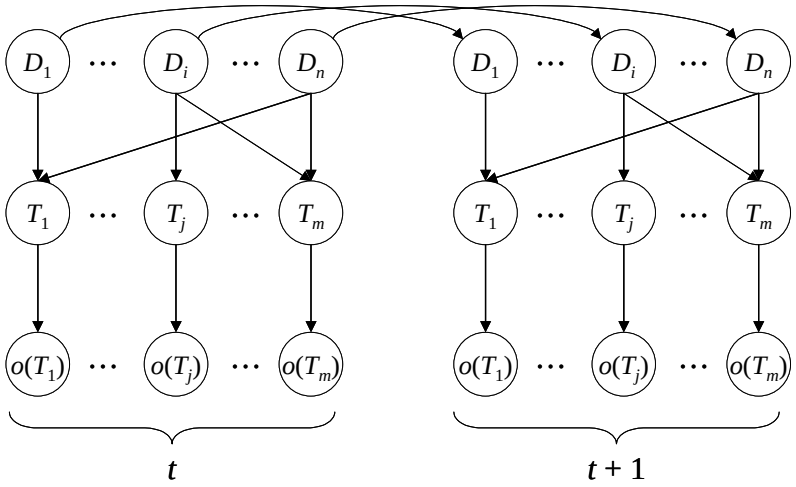

Fig. 3. Dynamic Bayesian network for prognosis.

It turns out that the Bayesian approach to diagnosis can be generalized in a straightforward way to address prognosis as well. In the most basic case, consider the state of the system as if it can be represented at some time t as a single random variable s_t. Assume, further, that the state at time $t + 1$ depends only upon the system state at time t. Then we can represent the time series corresponding to the system state progression as a first-order Markov chain.

Missing in our model is the fact that we do not have direct knowledge of the underlying state of the system. Specifically, we perform tests to observe conditions of the system, from which we infer the system state. Consequently, the basic Markov chain is not sufficient for our purpose—we need to differentiate between observable random variables and "hidden" (or unobservable) random variables. This leads the concept of a hidden Markov model (HMM) [7].

Formally, an HMM = $\langle N, M, A, B_j, \boldsymbol{\pi} \rangle$, where N is the number of states in the model (denote the states as $S = \{s_1, \ldots, s_N\}$), M is the number of distinct observation symbols per state (denote the symbols as $V = \{v_1, \ldots, v_M\}$), A is the state transition probability distribution $A = \{a_{ij}\} = \Pr(q_{t+1} = s_j \mid q_t = s_i)$, B_j is the observation probability distribution in state s_j, $B_j = b_j(k) = \Pr(v_k \text{ at } t \mid q_t = s_j)$, and $\boldsymbol{\pi}$ is the initial state distribution, $\boldsymbol{\pi} = \{\pi_I\} = \Pr(q_0 = s_i)$ [7].

The Markov chain and the HMM can be formulated as special cases of a graphical model first formalized by T. Dean and K. Kanazawa called the "dynamic Bayesian network" [3]. DBNs have been studied further by Kevin Murphy who provided alternatives for representation, inference, and learning [6]. The purpose of a DBN is to model probability distributions over semi-infinite collections of random variables, Z_i, that progress according to some temporal model. Typically, the random variables are partitioned into three subsets indexed by time—$Z_t = (U_t, X_t, Y_t)$ where U_t is the set of inputs at time t, X_t is the set of hidden (i.e., unobservable) variables at time t, and Y_t is the set of outputs at time t. Then, given the set Z, a DBN is defined to be a pair $\langle B_1, B_{\rightarrow} \rangle$, where B_1 is a Bayesian network defining the prior distribution $\Pr(Z_1)$, and $B_{\rightarrow}$ is a "two-slice" temporal Bayesian network defining the distribution $\Pr(Z_t \mid Z_{t-1})$ such that

$$\Pr(Z_t \mid Z_{t-1}) = \prod_{i=1}^{N} \Pr(Z_t^i \mid \mathrm{Pa}(Z_t^i)) \qquad (10)$$

where Z_t^i is the i^{th} node at time t, which could be a component of any of the partitions, and $\mathrm{Pa}(Z_t^i)$ are the parents of Z_t^i in the network. Of interest is the fact that the parents of a node, $\mathrm{Pa}(Z_t^i)$, can either be from the same time slice or from the previous time slice (i.e., the resulting model is restricted to being a first-order Markov model). Of course, the general formulation of DBNs allows for higher-order models simply by expanding the allowable set of parents to previous time slices.

To put this definition in the context of prognosis, we can construct a DBN for prognosis by "chaining" successive BBNs together. Under the first-order Markov assumption, we only need to represent two slices of the DBN and then "unroll" as

necessary in processing the model. For example, Fig. 3 shows how to link the BBNs in sequence. Note that only the diagnoses are linked through time since they change state directly. Changes in observation state are derived from the underlying state changes in the system. This approach is distinct from the HMM that links observations together.

To perform inference with the DBN (and thereby predict future states), first, infer the current state (i.e., the state in the current time slice) from the test observations. Next, "unroll" the DBN to the desired number of time slices (assuming the state progressions occur in discrete time steps—DBNs can handle continuous time, but the computation is more complex). Then, propagate beliefs through time by observing that

$$\begin{aligned}\Pr(D_i^{t+1}) = \Pr(D_i^{t+1} \mid D_i^t)\Pr(D_i^t) \\ + \Pr(D_i^{t+1} \mid \neg D_i^t)\Pr(\neg D_i^t)\end{aligned} . \qquad (11)$$

In fact, given the assumption that only diagnoses progress in state through time and that a diagnosis only depends upon itself in the previous time step, this part of the model reduces to a simple Markov chain, which can be either discrete time or continuous time.

Key to constructing the DBN is defining the temporal transition probabilities. In the simplest case, failure probabilities estimated from the failure rates can be used. When better information is available (e.g,. based on historical data), probabilities derived from this information can be used. The point is that the DBN is fully general and can be adapted to available knowledge about the system being analyzed. Theoretically, causal relationships between faults (i.e., a fault at time step t causes another fault to occur at time step $t + 1$) can be represented directly with the DBN as well (even though such models are rarely useful).

SUMMARY

In this paper, we developed a diagnostic approach based on Bayesian belief networks that incorporates information on failure probability, instrument uncertainty, and the predictions for false indication. Prognosis is performed using an extension of the Bayesian belief network, called a dynamic Bayesian network to model changes over time. The advantage to the discussed method is that it provides a formally consistent and theoretically sound approach to diagnosis and prognosis that can be adapted and matured as better estimates of the associated probabilities become available. By applying Bayesian inference based on reliability information and instrument uncertainty, resulting diagnoses accurately reflect the current state of the underlying system.

REFERENCES

[1] Bolch, G., S. Greiner, H. de Meer, and K. Trivedi, *Queuing Networks and Markov Chains*, New York: John Wiley & Sons, 1998.

[2] *Calibration Philosophy in Practice*, 2nd edition, Fluke Corporation, Everett, WA, 1994.

[3] Dean, T. & K. Kanazawa, "A Model for Reasoning About Persistence and Causation," *Artificial Intelligence*, 93(1–2):1–27, 1989.

[4] Duda, Richard O., Peter E. Hart, and David G. Stork, *Pattern Classification*, New York: John Wiley & Sons, 2001.

[5] Jackson, Dennis, "Measurement Risk Analysis Methods," *Proceedings of the Measurement Science Conference*, Anaheim, January 2005.

[6] Murphy, K. *Dynamic Bayesian Networks: Representation, Inference, and Learning*, PhD Dissertation, Department of Computer Science, University of California, Berkeley, 2002.

[7] Rabiner, L., "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of the IEEE*, Vol. 77, No. 2, pp. 257–286, 1989.

[8] Sheppard, John W., "Accounting for False Indication in a Bayesian Diagnostics Framework," *AUTOTESTCON 2003 Conference Record*, IEEE Press, September 2003.

[9] Sheppard, John W. and Mark Kaufman, "A Bayesian Approach to Diagnosis and Prognosis Using Built In Test," *IEEE Transactions on Instrumentation and Measurement*, Special Section on Built-In Test, Vol. 54, No. 3, June 2005.

[10] Sheppard, John W. and William R. Simpson, "Managing Conflict in System Diagnosis," in *Research Perspectives and Case Studies in System Test and Diagnosis*, Kluwer Academic Publishers, 1998