2011 American Control Conference
on O'Farrell Street, San Francisco, CA, USA
June 29 - July 01, 2011

Computing Detection Delays in Industrial Alarm Systems

Naseeb Ahmed Adnan, Iman Izadi and Tongwen Chen

Abstract— **False and nuisance alarms are major problems in the process industry. Techniques like deadbands, delay-timers, and filtering can significantly reduce these false and nuisance alarms. The down-side, however, is that using these techniques introduces some delay in raising the alarm (detection delay). In this paper, detection delays caused by deadband and delay-timer techniques are calculated using Markov processes. A design procedure is also proposed that compromises between detection delay, false alarm rate and missed alarm rate for an optimal configuration.**

Keywords:Alarm systems; Alarm management; Fault detection; Detection delay; Deadband; Delay-timer.

I. INTRODUCTION

Modern industries are monitored by hundreds and thousands of sensors. These sensors are installed in different areas and they communicate through a medium to monitor physical or environmental conditions of the plant. Operators are informed of any sensor measurement problem by alarms indicating abnormal behavior of the plant. To ensure cost efficiency, safety of the work force and plant, and quality of products, faults must be identified promptly and appropriate actions should be taken as soon as possible. There are several ways for fault detection as discussed in [1], [2], [3]. These detection techniques can be broadly classified into two categories: model-based, and signal processing based. Compared to signal processing based, model-based fault detection is a more active field in the area of control theory and engineering [4]. However, for most practical systems it is very difficult to obtain precisely known mathematical models [1] or they are highly nonlinear and not feasible for implementation from economic point of view. Therefore the application of the model-based scheme is limited.

The most common and frequently used fault detection method in industry is the simple limit checking method of a directly measured variable [1], [5]. This method can be referred to as signal processing based fault detection. This method has the advantages of simplicity and easy implementability. However a problem with this simple technique is to properly select the threshold which directly affects the number of false and missed alarms. A $false\ alarm$ is an alarm that is raised without presence of any abnormality in the process. A $missed\ alarm$ is an alarm that is not raised in the presence of a fault. Due to incorrect threshold setting, normal fluctuations of measured variables may result in a large number of false alarms under fault-free operation or increase the missed alarms significantly once fault has occurred.

In case of a fault occurrence, alarms may not be raised instantly due to different delays in the system. The delay can be caused by various reasons including network delays, bad implementation, hardware problems, sensor failure, and data loss. Also the alarm configuration (deadband, delay-timer, etc) can cause delay in raising the alarm. The difference in time between the actual moment of fault occurrence and the moment an alarm is activated is defined as the *detection delay*. For a reliable and effective alarm system, the false alarm rate, missed alarm rate and detection delay should be considered as three performance specifications.

Since the limit checking technique is widely exercised in industries, in this paper a major design constraint of this technique namely detection delay will be analyzed and considering other constraints such as false alarm rate and missed alarm rate [6], [7], a recommendation on tuning method will be discussed. The detection delay has been discussed to some extent in [8] for some change detection algorithms, e.g. CUSUM-type algorithm. But in this work alarm attributes (e.g. threshold limit, deadbands, delay-timers) described in the ISA 18.2 [9] or EEMUA 191 [10] standards for basic alarm design are considered only to analyze for detection delay; these were not addressed elsewhere earlier. This knowledge is important to include preventive measures in system design to compensate for activation delay.

In deadbands, two different limits are used for alarm raising and clearing. A higher threshold is set for raising the alarm, and a lower one for clearing the alarm. Another very effective technique in alarm systems is delay-timers. With delay-timers configured, the system requires consecutive few samples to cross the threshold before activating the alarm [7]. Both these technique (deadbands and delay-timers) introduce some delay in triggering the alarm. Markov processes are used in this paper to model the alarm system and estimate activation delay for these two techniques.

In Section II, an overview of Markov processes with assumptions presented in the paper are given. In Section III, detection delays are discussed in simplest case of threshold limit comparison. Section IV and V discuss detection delays due to deadbands and delay-timers, respectively. Section VI presents an analysis for finding the optimum set level of threshold and a design procedure. In Section VII, concluding remarks and future work are discussed.

Naseeb Ahmed Adnan and Tongwen Chen are with Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Alberta, Canada, T6G 2V4. naseeb@ece.ualberta.ca, tchen@ece.ualberta.ca

Iman Izadi is with Matrikon Inc.,# 1800, 10405 Jasper Ave. Edmonton, Alberta, Canada, T5J 3N4. iman@ualberta.ca

978-1-4577-0079-8/11/$26.00 ©2011 AACC

II. MARKOV PROCESS

A Markov process is an independent process where outcome at any time instant depends only on the outcome that precedes it and none before that [11]. In this paper Markov processes are used to estimate the detection delay for deadbands and delay-timers. Consider a Markov process with a limited number of states. Assume that the transitional probability, p_{ij}, is the probability of going from state e_i at time t to state e_j at time $t+1$ and define

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1j} & \cdots \\ p_{21} & p_{22} & \cdots & p_{2j} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{i1} & p_{i2} & \cdots & p_{ij} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix},$$

The matrix P is known as transition probability matrix. A probability vector π is called invariant for the Markov process if $\pi = \pi P$. In other words, π is a left eigenvector of P with eigenvalue 1 [11]. It is a known fact that an invariant vector π exists if the Markov process satisfies the following two conditions:

1) The sum of all entries in each row of P is 1 ($\sum_j p_{ij} = 1$).
2) All the entries of P are non-negative ($p_{ij} \geq 0$).

To satisfy these conditions and the definition of Markov process, we make the following assumptions on the process data:

1) Process data is independent and identically distributed (I.I.D.), i.e. at each sampling instant, the process data (random variable) has the same probability distribution as the other instants and all are mutually independent.
2) Probability density functions of the fault free and faulty data are known. These distributions can be estimated from historical data.

III. DETECTION DELAY

If a process variable moves from fault-free region of operation into faulty region of operation at time t_f and alarm is raised at time t_a, then the detection delay (DD) is given by the number of samples in the interval $t_a - t_f$. In the ideal case fault should be detected instantly at the moment of occurrence, which is hardly seen due to different delays. In practical condition, the problem is to detect the occurrence of the change as soon as possible [8]. Techniques like delay-timers, deadbands, and filters are widely used in alarm systems and enhance the effectiveness of limit checking method but increase the detection delay. However, even if no delay-timer, deadband and filter is configured, there may still be some detection delays as it is directly related the position of alarm threshold limit. In this section, we discuss detection delay for the simple threshold comparison alarm configuration.

In the fault free operating region assume the probability of one sample exceeding alarm limit is p_1 and the probability of one sample falling within the alarm limit is p_2. Similarly under the faulty region of operation, q_1 is the probability of one sample falling within the alarm limit and q_2 is the probability of one sample exceeding limit. Therefore $p_2 = 1 - p_1$ and $q_2 = 1 - q_1$ as shown in Fig. 1.

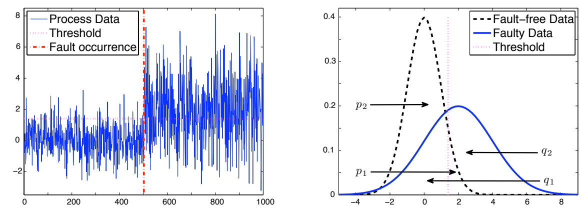

Fig. 1. Process data with threshold and fault occurrence instance [Left]; Corresponding probability density functions of fault-free and faulty data [Right]

Probability of zero detection delay is given by the probability of an alarm being raised instantly at the time of fault; probability of detection delay one means alarm activation is delayed by one sample from the fault instance. Similarly detection delay of z samples denotes alarm is raised z samples later from the actual instance of fault. Assume that, abnormality occurred at time $t = t_f$, A denotes an $alarm$ state and NA denotes a $no\ alarm$ state. Probability of detection delays are

$$\begin{aligned} P(DD=0) =& P(\text{A at } t=t_f) = q_2 \\ P(DD=1) =& P(\text{A at } t=t_f+1 \text{ \& NA at } t=t_f) = q_2.q_1 \\ &\vdots \\ P(DD=z) =& P(\text{A at } t=t_f+z \text{ \& NA at } t=t_f+z-1 \text{ \& } \ldots \\ & \text{\& NA at } t=t_f+1 \text{ \& NA at } t=t_f) \\ =& q_2.q_1^z \end{aligned}$$

Here the detection delays are in terms of samples, though the detection delay is normally known as a measure of time. But it will not affect the real scenario as in practice the sampling time is constant and these values can be converted to actual time measurements.

From the above equations it can be seen that in the simple case the probabilities of detection delay only depend on the probabilistic distribution of the faulty data. Fault-free region of operation does not have any effect on raising or clearing of alarm. The expected value of detection delay is

$$E(DD) = \sum_{z=0}^{\infty} z.P(DD=z) = \sum_{z=0}^{\infty} z.q_2.q_1^z = q_1/q_2$$

Expected value of detection delay (or average detection delay) is an important parameter in alarm design as it indicates the average time it takes to raise an alarm once there is an abnormality in the system. Therefore, it is always desired to reduce average detection delay to ensure more reliable operation of the plant. A method of threshold limit design will be discussed in details in Section VI.

IV. DETECTION DELAY FOR DEADBANDS

Deadbands are widely used in industry to eliminate repeating oscillations or chattering alarms. With deadband configured, alarms are raised and cleared according to two different limits, instead of the same limit in regular cases. For example, for high alarms a limit is set, as usual, for raising the alarm. However, once an alarm is activated it will not be cleared even if the variable falls below the limit. To clear the alarm, the variable must go below a lower threshold.

When a variable transits from the fault-free state to the faulty state, due to presence of noise, it crosses the alarm limit a few times before settling in the abnormal state. This oscillation results in subsequent raising and clearing of the alarm causing the chattering effect. Using deadband (i.e., separating raising and clearing limits) is, then, helpful here in preventing alarm chattering [12].

Deadbands should typically be configured based on the normal operating range of the process variables, measurement noise, and type of the process variables [9], [12]. There are certain standards for setting deadbands, e.g. in ISA 18.2 [9] or EEMUA [10]. The correct configuration of a deadband (setting limits appropriately) is essential in maximizing the benefits of the deadband.

When deadband is configured, as it can be seen in Fig. 2, the probabilities of one sample going over the raising limit and below the clearing limit do not add up to one, i.e., $p_1 + p_2 \neq 1,\ q_1 + q_2 \neq 1$; where p_1, p_2, q_1, q_2 are defined as before in Section III.

To calculate the probabilities of false alarm and missed alarm, notice that a process operating in either the fault-free or faulty region, can be in two states from alarm point of view: alarm state (A) and no alarm state (NA). These states can be modeled with a Markov chain [7]. A Markov process for deadband is shown in Fig. 3 for fault-free operating region. The Markov model in faulty region is similar. Transitional probabilities from one state to other are represented by transition probability matrix P_n, in the fault-free region of operation and by P_f in the faulty region of operation:

$$P_n = \begin{bmatrix} 1-p_1 & p_1 \\ p_2 & 1-p_2 \end{bmatrix}, P_f = \begin{bmatrix} 1-q_2 & q_2 \\ q_1 & 1-q_1 \end{bmatrix}$$

If the process remains in the fault-free operating region, after a transient time the Markov process reaches its steady state and the vector of state probabilities converges to the invariant vector. The steady state vector of probabilities (invariant vector) for P_n is [7]

$$\pi_n = \begin{bmatrix} \dfrac{p_2}{p_1+p_2} & \dfrac{p_1}{p_1+p_2} \end{bmatrix}$$

When a fault occurs, the Markov process changes from fault-free model (represented by P_n) to the faulty model (represented by P_f). Therefore, the steady-state probabilities for the fault-free operation (i.e., π_n) should be used as the initial state probabilities for the faulty operation. The Markov process considered here is an ergodic Markov process; aperiodic and positive recurrent. For an ergodic process, the steady-state invariant vector is unique [11]. Therefore the Markov process in the faulty operation always have this unique steady-state invariant vector as initial condition.

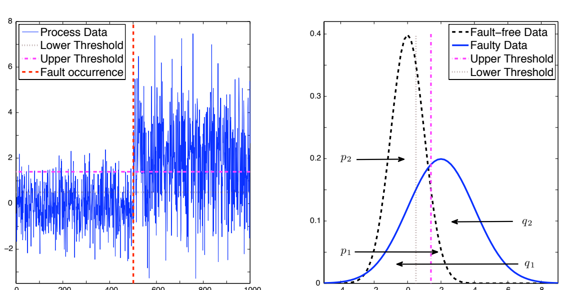

Fig. 2. Process data with deadbands and fault occurrence instance [Left]; Corresponding probability density functions of fault-free and faulty data [Right]

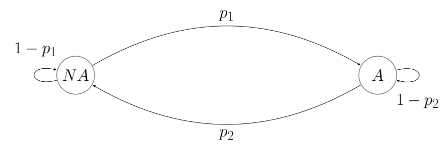

Fig. 3. Markov diagram of a system with deadband in fault-free region of operation

If the system transfers from fault-free to faulty state at time t_f, using the forward Chapman-Kolmogorv equations [11], probabilities of alarm and no-alarm states can be calculated as

$$\begin{aligned}[P_{NA}(t_f)\ \ P_A(t_f)] &= [P_{NA}(t_f-1)\ \ P_A(t_f-1)]\ P_f \\ &= \pi_n P_f = \begin{bmatrix} \dfrac{p_1q_1 + p_2(1-q_2)}{p_1+p_2} & \dfrac{p_2q_2 + p_1(1-q_1)}{p_1+p_2} \end{bmatrix}\end{aligned}$$

Therefore the probability of detection delay zero (probability of alarm being raised immediately after transition from normal to abnormal) is

$$\begin{aligned}P(DD=0) &= P(\text{A at } t = t_f) \\ &= P_A(t_f) = \frac{p_2q_2 + p_1(1-q_1)}{p_1+p_2}\end{aligned}$$

Probabilities of higher detection delays can be expressed in terms of conditional probabilities as

$$\begin{aligned}P(DD=1) =& P(\text{A at } t = t_f+1 \text{ \& NA at } t = t_f) = \frac{p_1q_1 + p_2(1-q_2)}{p_1+p_2}.q_2 \\ P(DD=2) =& P(\text{A at } t = t_f+2 \text{ \& NA at } t = t_f+1, \text{ NA at } t = t_f) \\ =& \frac{p_1q_1 + p_2(1-q_2)}{p_1+p_2}.q_2.(1-q_2) \\ &\vdots \\ P(DD=z) =& P(\text{A at } t = t_f+z \text{ \& NA at } t = t_f+z-1, \ldots, \\ & \text{NA at } t = t_f) \\ =& \frac{p_1q_1 + p_2(1-q_2)}{p_1+p_2}.q_2.(1-q_2)^{z-1}, \quad z \geq 1\end{aligned}$$

The average detection delay (i.e., the expected value of the detection delay) is then given by

$$\begin{aligned}E(DD) &= \sum_{z=0}^{\infty} z.P(DD=z) \\ &= \sum_{z=1}^{\infty} z.q_2(1-q_2)^{z-1}.\frac{p_1q_1 + p_2(1-q_2)}{p_1+p_2} = \frac{p_1q_1 + p_2(1-q_2)}{q_2(p_1+p_2)}\end{aligned}$$

V. DETECTION DELAY FOR DELAY-TIMERS

A delay-timer is a simple yet effective technique that can reduce the number of false and nuisance alarms significantly. By their intuitive nature, human beings prefer to wait for a while before reacting to an abnormality to avoid any temporary overshoot or undershoot. Delay-timers use the same concept in alarm generation. If a delay-timer is configured on a variable, the alarm is raised if n consecutive samples cross the alarm limit. This case is known as on-delay. Similarly, once the alarm is raised, it will only be cleared if m consecutive samples go below the limit, known as off-delay. For on-delay, if the system goes back to normal operating state during the intermediate states, alarm is not activated and vice-versa for the off-delay case. Alarm standards (EEMUA

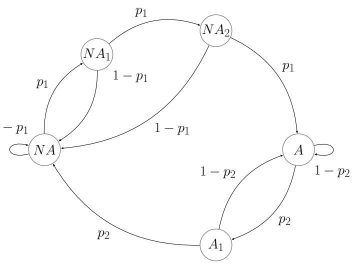

Fig. 4. Markov diagram of a system with delay-timer $n = 3$ and $m = 2$ in fault-free region of operation

[10] and ISA [9]) recommend some values for delay-timers based on the nature of the process variable. Similar to the deadband case, we can model the alarm/no-alarm states of a process variable with delay-timers by Markov chains [7]. Fig. 4 shows the Markov model of a system in its fault-free operation with $n = 3$ samples on-delay and $m = 2$ samples off-delay. A similar Markov model with probabilities q_1 and q_2 can be constructed for the faulty region of operation.

In Fig. 4, assuming NA or no alarm state is the initial state for the process, $1 - p_1$ denotes the probability that it will remain in the same state. If the next sample exceeds alarm threshold with probability p_1 then it moves to state NA_1, which is the first intermediate state for $3-$samples delay case before going to alarm state. Exceeding threshold by consecutive $3-$samples with probability p_1 will take the system to the alarm state. If in between the states NA and NA_2, any sample falls below threshold with probability $1 - p_1$, then system will go back to no alarm state. Only consecutive 3-sample crossing over threshold can raise the alarm, taking the system to the alarm state. Once the system goes to alarm state, to clear the alarm same principal is followed as raising. Consecutive 2 samples falling below threshold with probability p_2 (where $p_2 = 1 - p_1$) can only clear the alarm. In intermediate states as output, the system will always provide either alarm or no alarm, depending on from where these intermediate states started. Transitional probability matrices for $n-$samples on-delay and $m-$samples off-delay for both the fault-free (P_n) and faulty (P_f) region of operations are given by

$$P_n = \left[\begin{array}{c|c} P_{n11} & 0_{n\times(m-1)} \\ \hline P_{n21} & P_{n22} \end{array}\right], \; P_f = \left[\begin{array}{c|c} P_{f11} & 0_{n\times(m-1)} \\ \hline P_{f21} & P_{f22} \end{array}\right]$$

where,

$$P_{n11} = \begin{pmatrix} 1-p_1 & p_1 & 0 & \cdots & 0 \\ 1-p_1 & 0 & p_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1-p_1 & 0 & 0 & \cdots & p_1 \end{pmatrix},$$

$$P_{n21} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 \\ p_2 & 0 & \cdots & 0 \end{pmatrix}, P_{n22} = \begin{pmatrix} 1-p_2 & p_2 & 0 & \cdots & 0 \\ 1-p_2 & 0 & p_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1-p_2 & 0 & 0 & \cdots & p_2 \\ 1-p_2 & 0 & 0 & \cdots & 0 \end{pmatrix},$$

here, horizontal line in P_n is used to indicate, number of columns in P_{n11} and P_{n21} or in $0_{n\times(m-1)}$ and P_{n22} are not equal. Same is true for P_f. Furthermore P_{f11}, P_{f21} and P_{f22} have the same structure as P_{n11}, P_{n21} and P_{n22} respectively, only p_1, p_2 are replaced by q_2, q_1. P_{n11} and P_{f11} have dimension $n \times (n+1)$, P_{n21} and P_{f21} are $m \times n$ and P_{n22} and P_{f22} are $m \times m$. Hence, P_n and P_f are of dimension $(n+m) \times (n+m)$.

Similar to the deadband case, at the moment of fault occurrence, the Markov model of the system switches from the fault-free model (represented by P_n) to the faulty model (represented by P_f). Therefore, to calculate the probabilities of the state after fault occurrence, P_n is assumed to reach its steady state; and the steady state probabilities of the fault-free model, should be used as initial states for the faulty model. The steady state vector of probabilities for the fault-free state of operation (e.g., the invariant vector for P_n) is [7]

$$\pi_n = \frac{1}{p_2^m \sum_{i=0}^{n-1} p_1^i + p_1^n \sum_{j=0}^{m-1} p_2^j} \times \begin{bmatrix} p_2^m & p_1 p_2^m & \cdots & p_1^{n-1} p_2^m & p_1^n & p_2 p_1^n & \cdots & p_2^{m-1} p_1^n \end{bmatrix}$$

To calculate the detection delay for delay-timers, we use the concept of *hitting time* [11]. In our context, hitting time is the minimum time required for system currently in the no-alarm state to switch to the alarm state for the first time.

We divide the whole state space into two subspaces. The first subspace, denoted by $\mathcal{D}$, contains all the no-alarm state(s) $(NA, ..., NA_{n-1})$. The second subspace contains all the alarm state(s) $(A, ..., A_{m-1})$ and is denoted by $\mathcal{E}$. With these definitions, the detection delay is the same as the hitting time: the time required to switch to states $\mathcal{E}$, assuming the system is initially started in states $\mathcal{D}$.

Let Q be the matrix of transitional probabilities from $\mathcal{D}$ to itself in the faulty region. Q is obtained from P_f by keeping all the probabilities corresponding to no-alarm states and replacing all other probabilities with zero. For the Markov process in Fig. 4, Q is

$$Q = \left[\begin{array}{cc} P_{f11} & 0_{n\times(m-1)} \\ \hline 0_{m\times n} & 0_{m\times m} \end{array}\right]$$

It can be shown that probability of $z-$sample detection delay for delay-timer (e.g., hitting time for switching from no-alarm states $\mathcal{D}$, to alarm states $\mathcal{E}$) is given by

$$P(DD = z) = \pi_n . P_f . Q^z \left[0 \cdots 0 \; 1 \cdots 1\right]^T$$

Here in the column vector there are n zeros and m ones. The expected detection delay for delay-timer can then be expressed as

$$E(DD) = \sum_{z=0}^{\infty} z.P(DD = z) = \pi_n . P_f . Q.(I - Q)^{-2}. \left[0 \cdots 0 \; 1 \cdots 1\right]^T$$
$$= \frac{p_2^{m-1}\left(p_1^n q_1 \sum_{i=0}^{n-1} q_2^i + p_2 \left(\sum_{j=0}^{n-1} p_1^j \sum_{k=0}^{n-j-1} q_2^k - q_2^n \sum_{i=0}^{n-1} p_1^i\right)\right)}{q_2^n \left(p_2^m \sum_{i=0}^{n-1} p_1^i + p_1^n \sum_{i=0}^{m-1} p_2^i\right)}$$

Since Q is a $substochastic\ matrix$, i.e. a matrix with nonnegative entries whose row sums are less than or equal to 1 and $Q^n \to 0$ as $n \to \infty$; therefore, all eigenvalues of Q have absolute values strictly less than 1; and the series in the summation converges [11].

Detection delays for delay-timers are calculated so far for $n-$samples on-delay and $m-$samples off-delay. When $n = 1$

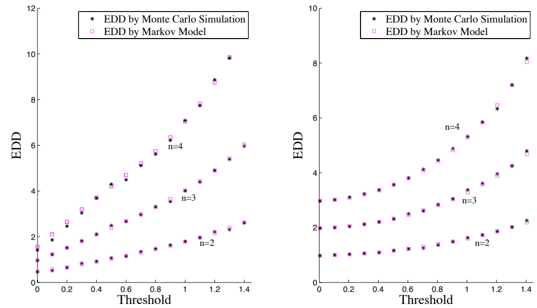

Fig. 5. Verification of the EDD formula by Monte-Carlo simulation. (Left) assuming Gaussian data distribution; (Right) assuming Gamma data distribution

and $m = 1$, the expected detection delay is then simplified to

$$E(DD) = \frac{p_1 q_1 + p_2(1 - q_2)}{q_2(p_1 + p_2)}$$

which is consistent with the result we obtained in Section IV for deadbands. Furthermore if we assume there is no deadband, then $p_2 = 1 - p_1$ and $q_2 = 1 - q_1$ and the result is further simplified to $E(DD) = q_1(1 - q_1)^{-1}$, which is again consistent with the result obtained in Section III for the simple case (no deadband, no delay-timer).

In Fig. 5, Monte Carlo simulations to verify the expected detection delay (EDD) expression are presented. The expected detection delay is plotted for different delay-timers (assuming $m = n$) as a function of the threshold. Monte Carlo simulation is shown for two different distributions, $Gaussian$ and $Gamma$. Fault-free data has mean 0 and variance 1; faulty data has mean 2 and variance 2. The threshold is changed from 0 to 1.4 with an increment of 0.1. The data was simulated for 2000 iterations and the mean EDD was estimated for each value of the threshold. The proposed EDD equation is also plotted with Monte Carlo simulation. It can be seen that, there is very negligible differences between the Monte Carlo simulations and the calculated EDD by our formula.

VI. DESIGN OF ALARM SYSTEMS

An alarm design procedure is described in [7] based on the receiver operating characteristic (ROC) curve. The ROC curve is the plot of the probability of missed alarms versus the probability of false alarms when the trip point changes from -∞ to +∞. As it can be seen in Fig. 1, lowering the trip point decreases the probability of false alarms, but the probability of missed alarms will increase. The ROC curve shows this trade-off between false alarm and missed alarm rates when the trip point changes. A typical ROC curve is shown in Fig. 6 [right] for the corresponding fault-free and faulty data. However, setting the threshold merely based on these two facts may not be desirable as it does not consider detection delay. If the same weight is given to false alarm and missed alarm rates, then the optimum point on the ROC curve (corresponding to the optimum trip point) will be the point closest to the origin. In that case, if the threshold is set higher than the optimum point, false alarms will decrease at the cost of increased missed alarms. On the other hand, setting the threshold lower than the optimum point will result in smaller number of missed alarms but more false alarms. In the rest of this section, we focus our design on delay-timers only. For a given set of fault-free and faulty data, the design parameters are then the threshold (t), the number of on-delay samples (n) and number of off-delay samples (m). For simplicity we assume $m = n$. An acceptable design should not only minimize false and missed alarm rates, but also guarantee a small detection delay. Therefore, for design of an alarm system, three performance specifications are to be considered: false alarm rate (FAR), missed alarm rate (MAR) and expected detection delay (EDD). For practical design, however, the false and missed alarm rates are usually combined in a function that measures how far a point on the ROC curve is from the ideal point (zero false alarm and zero missed alarm). In most cases the ROC curve is symmetrical, and it can be assumed that at the optimal point FAR and MAR are approximately equal. Therefore for simplicity we assume the optimal point is approximately the point where FAR $=$ MAR in design. For simulations, assume the fault-free part of the data has a Gaussian distribution with average of 0 and variance 1. The faulty part of the data has also Gaussian distribution with average 2 and variance 2. To design a system with requirements FAR $\leq 4\%$, MAR $\leq 3\%$ and EDD ≤ 6 samples, a four-step procedure is followed. In Fig. 6, delay timer is changed to different values and corresponding ROC curves are plotted; it can be seen that as the delay timer is increasing, the ROC curve is moving closer to the origin [7]. On the other hand this increase of delay timer increases the expected detection delay as well. Fig. 8 shows expected detection delay for different delay timers for the threshold where FAR $=$ MAR; these equal thresholds were estimated from Fig. 7, where FAR and MAR are plotted with corresponding threshold limits and the intersecting points provide the required estimations.

Step 1

In this step, alarm thresholds that satisfy the optimality condition FAR $=$ MAR are estimated. These thresholds can be estimated from Fig. 7. For the given process data, thresholds corresponding to the point of equal FAR and MAR are shown by dots, for different values of delay timers. The estimated thresholds are given in Table I.

Step 2

In step 2, the smallest delay-timer n_1 is selected from Fig. 7 or Table I such that FAR $=$ MAR $\leq 3\%$; where

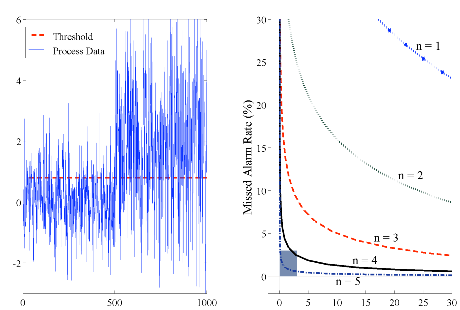

Fig. 6. Process data [Left]; ROC curve changing delay-timer [Right]

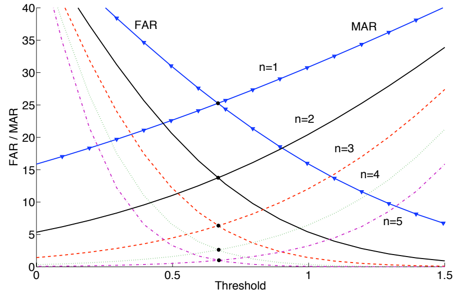

Fig. 7. Estimation of threshold when FAR = MAR

$n \geq n_1$. The corresponding area for FAR and MAR (for $\leq 3\%$) is shown by the shaded area in Fig. 6. The smallest delay-timer that satisfies the condition is $n_1 = 4$. Though design requirement was for FAR $\leq 4\%$, n_1 is selected for the one with lower percentage ($\leq 3\%$) requirements among FAR and MAR. Since n_1 is selected for more conservative range, the selected delay-timer does not violate the original requirements and is expected to provide better performance.

Step 3

The EDD is taken into account in step 3 for the design of delay-timer. The largest value of delay-timer n_2 is selected such that EDD ≤ 6; where $n \leq n_2$. From Fig. 8, $n_2 = 4$.

Step 4

Once n_1 and n_2 are estimated, the next step is to select the range of delay timers to finalize the design process. If $n_2 \geq n_1$, any n satisfying $n_1 \leq n \leq n_2$ is a solution of the delay-timer. Here the only delay-timer that satisfies the condition is, $n = 4$; from Table I the shaded row is the solution of given design problem. $n = 3$ satisfies the condition of EDD but does not satisfy the requirement of FAR / MAR; other delay-timers also do not satisfy requirements. The optimum threshold of operation is 0.67 and delay-timer is 4 samples, and it will take 5.04 samples to raise the alarm. If such a range of delay-timer cannot be found; or in other words if no such n exists to satisfy $n_1 \leq n \leq n_2$; design requirements need adjustments. Less demanding design requirements are recommended for such cases.

A Monte Carlo simulation is performed to check consistency of the calculated values. Setting the threshold to 0.67, associated false alarm rates, missed alarm rates and detection delays were calculated. From a Monte Carlo simulation of 5000 iterations (for Gaussian distributed fault-free data with mean 0, variance 1 and faulty data with mean 2 and variance 2), the calculated FAR is 2.58%, MAR is 2.84 %, and

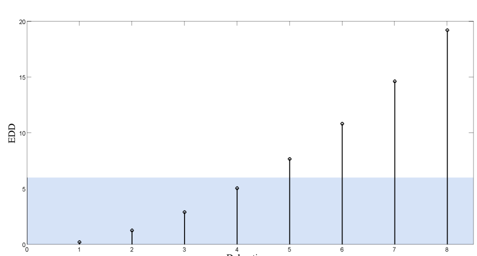

Fig. 8. Effect on EDD for different delay-timers

TABLE I

DESIGN PARAMETERS SELECTION CHART

Delay-timer (n)	Threshold (t)	FAR = MAR (%)	EDD
1	0.67	25.26	0.21
2	0.67	13.76	1.26
3	0.67	6.37	2.89
4	0.67	2.63	5.04
5	0.67	1.01	7.66

detection delay is 4.92 samples, which are consistent with the calculated values.

VII. CONCLUSIONS AND FUTURE WORKS

The expected detection delay, as a measure of the time it takes for the alarm system to respond to a fault, is an important parameter in the design of alarm systems. In this paper, the expected detection delay is calculated for two common techniques in alarm systems, namely, deadbands and delay-timers. We also presented a simple design procedure, based on three important performance measures of an alarm system: false alarm rate, missed alarm rate and the expected detection delay. As a future extension, a more systematic approach to the design of parameters of the alarm system (on-delay timer, off-delay timer and the trip-point) is under investigation.

VIII. ACKNOWLEDGMENTS

The authors would like to thank Dr. B. Schmuland, Department of Mathematical and Statistical Sciences, University of Alberta, for helpful discussions.

REFERENCES

[1] R. Isermann, *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*, Springer-Verlag, Berlin, Heidelberg, 2006.

[2] J. Chen and R. Patton, *Robust Model-Based Fault Diagnosis for Dynamic Systems*, Kluwer Academic Publishers, Boston, 1999.

[3] J.J. Gertler, *Fault Detection and Diagnosis in Engineering Systems*, Marcel Dekker, New York, 1998.

[4] S. X. Ding, *Model-based Fault Diagnosis Techniques: Design Schemes, Algorithms, and Tools*, Springer-Verlag, Berlin, Heidelberg, 2008.

[5] J. Ahnlund, T. Bergquist, and L. Spaanenburg, *Rule-based reduction of alarm signals in industrial control*, Journal of Intelligent and Fuzzy Systems, IOS press, vol. 14, no. 2, pp. 73-84, 2003.

[6] I. Izadi, S.L. Shah, D. S. Shook, and T. Chen, *An introduction to alarm analysis and design*, Proceedings of the 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes, pp. 645-650, June 30 - July 3, 2009, Barcelona, Spain.

[7] I. Izadi, S. L. Shah, D. S. Shook, S. R. Kondaveeti and T. Chen, "Optimal alarm design", *Proceedings of the 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes*, pp. 651-656, June 30 - July 3, 2009, Barcelona, Spain.

[8] M. Basseville, I.V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*, PTR Prentice-Hall Inc. 1993.

[9] International Society of Automation, *Management of Alarm Systems for the Process Industries*, ANSI/ISA18.22009.

[10] Engineering Equipment and Materials Users Association (EEMUA), *Alarm Systems-A Guide to Design, Management and Procurement*, EEMUA Publication 191, 2007.

[11] G. F. Lawler, *Introduction to Stochastic Processes*, 2nd Edition, Chapman & Hall/CRC, 2006.

[12] A. Hugo, "Estimation of Alarm Deadbands",*Proceedings of the 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes*, pp. 663-667, June 30 - July 3, 2009, Barcelona, Spain.