Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

On Threshold Optimization in Fault Tolerant Systems ⋆

Fredrik Gustafsson, Jan Åslund, Erik Frisk, Mattias Krysander, and Lars Nielsen

Department of Electrical Engineering
Linköping University, Sweden
Email: {fredrik,jaasl,frisk,matkr,lars}@isy.liu.se

Abstract: Fault tolerant systems are considered, where a nominal system is monitored by a fault detection algorithm, and the nominal system is switched to a backup system in case of a detected fault. Conventional fault detection is in the classical setting a trade-off between detection probability and false alarm probability. For the considered fault tolerant system, a system failure occurs either when the nominal system gets a fault that is not detected, or when the fault detector signals an alarm and the backup system breaks down. This means that the trade-off for threshold setting is different and depends on the overall conditions, and the characterization and understanding of this trade-off is important. It is shown that the probability of system failure can be expressed in a general form based on the probability of false alarm and detection power, and based on this form the influence ratio is introduced. This ratio includes all information about the supervised system and the backup system that is needed for the threshold optimization problem. It is shown that the influence ratio has a geometrical interpretation as the gradient of the receiver operating characteristics (ROC) curve at the optimal point, and furthermore, it is the threshold for the optimal test quantity in important cases.

Keywords: fault detection, fault tolerant control, threshold selection, system failure

1. INTRODUCTION

Safety is of major concern in many applications [Vil92], and because of that fault tolerance can be introduced by means of a back-up system. Then the key mechanism is situation classification, [BKLS03], followed by a decision to switch to the back-up system. When designing such a system, it is very important from an application point of view to understand the safety implications of design choices like threshold selection [ÅBF+07]. Such safety analysis is the topic of this paper, aiming at obtaining useful and explicit characterizations.

Consider the design of a fault tolerant system in Figure 1. A fault detector monitors the performance of a nominal system. This nominal system can be a certain sensor, a computer that generates a control signal, or (part of) a plant. The idea is that it is possible to switch to a backup system whenever appropriate. Here, the backup system is used after an alarm is issued by the fault detector. The risk after such a decision is that, before maintenance is practically possible, the backup system breaks down after which there is no remedy.

The navigation system in the Swedish fighter Gripen is designed as in Figure 1. A standard inertial navigation system, comprising an inertial measurement unit (IMU) with support sensors, is the nominal system 1. The steering system has its own set of IMU and support sensors, from which an AHRS (adaptive heading reference system)

⋆ This work has been performed within the MOVIII excellence center funded by the Swedish Strategic Research Council

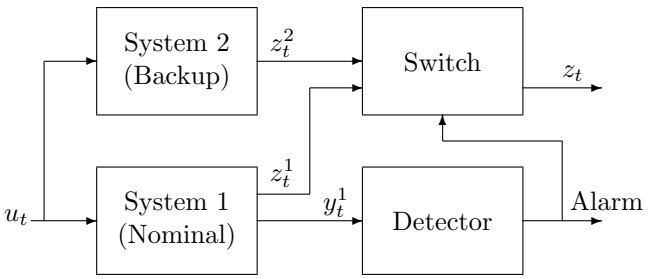

Fig. 1. Fault tolerant system, where the nominal system 1 is switched to the backup system 2 after the fault detector signals for an alarm.

provides an independent backup navigation system. The detector is here called XMON (cross monitoring), and it compares the two navigations systems using statistical decision algorithms.

One first fundamental question is of course if the total system becomes safer when a diagnosis function and a back-up system are introduced, and if so, by how much? Another question is how to formulate specification requirements on the diagnosis algorithms so that overall system safety is as good as possible [ÅBF+07]. This also naturally leads to the question of how to select internal design parameters in the diagnosis algorithms, like that of selection of a threshold that balances the rates of missed detection and false alarm.

To get a handle on these questions it is necessary to have a quantitative method, and this will also be a requirement from government, e.g. when declaring air worthiness for aircraft. It should be noted that the main concern here is the interplay between safety and algorithms, and that this should not be confused with the more studied problem on safety of software. It is here assumed that the software is a correct coding of the specified algorithm following the procedures for implementation of safety critical systems.

The purpose of this paper is to analyze fault tolerant systems, like the one illustrated in Figure 1, to obtain as much useful insight as possible. This means that once the problem is formulated, different paths of solution are investigated, which gives alternative characterizations of the optimal design. Section 2 introduces fault detection and recapitulates the main performance measures. Section 3 gives the basic formulation and expands the probability of system failure into an expression including false alarm probability and detection probability. Based on that expression, an influence factor is defined. It is shown in Section 4 and 5 how the influence factor characterizes overall system characteristics and the threshold selection. Finally, the conclusions are drawn in Section 6.

2. FAULT DETECTION AND PERFORMANCE MEASURES

In noisy environments, a typical fault detector consist of a test statistic, also called a test quantity, and a corresponding threshold. The test quantity, $T(\text{observations})$ is a function of the observations and the detector alarms if the test statistic exceeds the threshold h, i.e. if

$$T(\text{observations}) > h \tag{1}$$

The idea is that the test statistic is close to zero in the fault free case, but not in the faulty case [Ber85]. Thus if $T > h$ the fault detector alarms and a fault has been detected.

Two commonly used performance measures for fault detectors are the probabilities of false alarm

$$P_{FA}(h) = P(T > h|\text{no fault}) \tag{2}$$

and the detection power of a fault f of size f_0

$$P_D(h) = P(T > h|f = f_0). \tag{3}$$

The fault size f_0 will be a representative size of the faults causing failures. The dependence of the fault size on the detection power will therefore not be explicit in the continuation. Ideally, the false alarm probability should be zero and the probability of detection one.

3. SYSTEM FAILURE AND FAULT DETECTION

The objective is to analyze fault tolerant systems that include a fault detection system and the system in Figure 1 will be used to illustrate basic concepts. A discussion on more general systems is found in Section 5. For the system in Figure 1, the nominal system is automatically switched to the backup system when the fault detector alarms. The alarm may be correct or false and the question is how to select the threshold of the fault detector to minimize the probability of system failure (SF). Since the detection system consists of a test quantity that alarms if it exceeds a given threshold h, the studied optimization problem is then

$$\min_h P(\text{SF}) \tag{4}$$

The event SF is a logical consequence of other events. For the example system in Figure 1 there are two mutually exclusive cases where the system fails:

- Nominal system breaks down ($\neg$NOM) without being detected ($\neg$ALARM).
- The fault detector alarms (ALARM) after which the backup system breaks down ($\neg$BACKUP).

This example will be used in the following to illustrate properties that hold in a more general setting. The probability for system failure is thus

$$P(\text{SF}) = P(\neg\text{ALARM} \wedge \neg\text{NOM}) + P(\text{ALARM} \wedge \neg\text{BACKUP}) \tag{5}$$

Note that all probabilities are conditioned on situation dependent factors, such as operation time, operational regulations, maintenance programs, etc.

Expanding the first term in (5) gives

$$\begin{aligned} P(\neg\text{ALARM} \wedge \neg\text{NOM}) &= P(\neg\text{ALARM}|\neg\text{NOM})P(\neg\text{NOM}) \\ &= (1 - P_D)P(\neg\text{NOM}) \end{aligned}$$

Assume that alarm and failure of the backup system are independent events, then the second term in (5) is

$$P(\text{ALARM} \wedge \neg\text{BACKUP}) = P(\text{ALARM})P(\neg\text{BACKUP})$$

where

$$\begin{aligned} P(\text{ALARM}) &= P(\text{ALARM}|\neg\text{NOM})P(\neg\text{NOM}) \\ &\quad + P(\text{ALARM}|\text{NOM})P(\text{NOM}) \\ &= P_D P(\neg\text{NOM}) + P_{FA} P(\text{NOM}) \end{aligned}$$

Summing up, the probability for system failure in the example is given by

$$P(SF) = \alpha P_{FA} - \beta P_D + \gamma \tag{6}$$

where

$$\begin{aligned} \alpha &= P(\text{NOM})P(\neg\text{BACKUP}) \\ \beta &= P(\neg\text{NOM})P(\text{BACKUP}) \\ \gamma &= P(\neg\text{NOM}) \end{aligned}$$

One can note that $P(SF)$ is a linear function of the probabilities P_{FA} and P_D in the example. This property also holds in more general cases where the fault tolerant system switches between different configurations depending on alarm state. This is studied further in Section 5.

The expression (6) will be used in the following section to analyze the optimization problem (4). The influence ratio λ is defined as the ratio of the influence factors for the probabilities P_{FA} and P_D in (6), i.e.

$$\lambda = \frac{\alpha}{\beta} \tag{7}$$

To solve the optimization problem (4), in order to find the optimal threshold h, is equivalent to solving the problem

$$\min_h \ \lambda P_{FA}(h) - P_D(h)$$

This means that the ratio λ includes all information about the supervised system, the backup system, and the fault tolerant control strategy that is needed for the threshold optimization problem.

4. THRESHOLD SELECTION

This section will describe how the parameter λ appears explicitly in two different contexts concerning optimal threshold selection. The first is a geometric interpretation of λ in a tool, called ROC curve, for evaluating performance of a test. The second concerns determining the value of an optimal threshold when an optimal test, a likelihood ratio test, is used.

4.1 The ROC curve and the influence ratio λ

Given a test statistic T, the false alarm probability and detection power depend on the threshold value h as can be seen in (2) and (3). Threshold selection typically is a compromise between obtaining low false alarm probability and a high detection power. This compromise can be seen in a so called receiver operating characteristics (ROC) curve, where $P_{FA}(h)$ is plotted against $P_D(h)$. ROC curves can also be used to compare different test statistics T_i. For example, for a given false alarm probability the best test statistic is the one which gives the highest power $P_D(h)$. That is, the ROC curve is a convenient evaluation tool in fault detection. The ROC curve for a test that will be used in Section 4.2 is plotted in Figure 2.

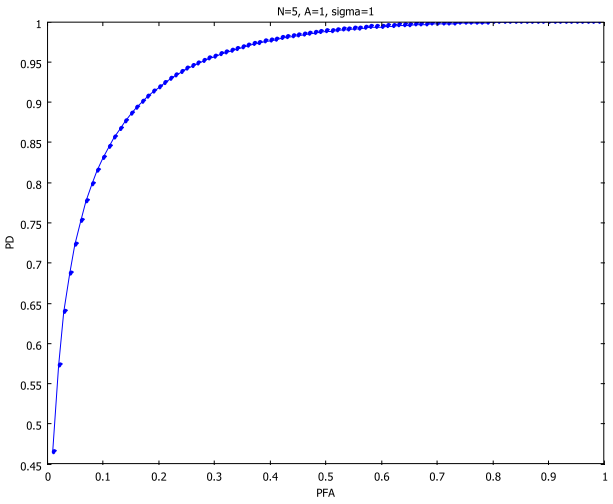

Fig. 2. Example of an ROC curve.

Differentiating (6) with respect to h we get the following condition for the optimal threshold

$$\alpha \frac{dP_{FA}(h)}{dh} = \beta \frac{dP_D(h)}{dh}$$

which gives that

$$\frac{dP_D}{dP_{FA}} = \frac{\alpha}{\beta} = \lambda$$

Thus, for the optimal threshold it holds that the gradient of the ROC curve equals the influence ratio λ introduced in (7). The constant λ is dependent only on the prior probabilities of component failures.

4.2 Optimal threshold selection and the influence ratio λ

An optimal threshold is determined with respect to a given test quantity. Here, a quite generic setup for the detection problem is used and the specific test that is used is an optimal detection test in the detection setup.

A simple but still quite generic model for the fault detection problem is to decide whether a computed residual is zero mean or not. The two corresponding hypotheses are given by

$$H_0 : r_k = e_k, \tag{8a}$$
$$H_1 : r_k = A + e_k. \tag{8b}$$

Here, r_k, $k = 1, \ldots, N$, represent different residuals over time and/or space, A is a constant given by the fault response of a representative fault, and e_k is white Gaussian noise with known variance. Many applications can be recast into this model.

The Neyman-Pearson theorem [Leh91, Kay98] states that an optimal test quantity is given by the likelihood ratio

$$\bar{T}(r_{1:N}) = \frac{P(r_{1:N}|H_1)}{P(r_{1:N}|H_0)} > \bar{h} \tag{9}$$

where $\bar{h}$ is the threshold. Utilizing the whiteness and Gaussian property of the noise we get

$$\begin{aligned}\bar{T}(r_{1:N}) &= \frac{\frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{\sum_{k=1}^{N}(r_k - A)^2}{2\sigma^2}}}{\frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{\sum_{k=1}^{N}(r_k)^2}{2\sigma^2}}} \\ &= e^{-\frac{NA^2 - \sum_{k=1}^{N} 2Ar_k}{2\sigma^2}} > \bar{h}\end{aligned}$$

Taking logarithm and simplifying give the test quantity

$$T(r_{1:N}) = \frac{1}{N}\sum_{k=1}^{N} r_k > \frac{\sigma^2}{NA}\log\bar{h} + \frac{A}{2} \equiv h \tag{10}$$

which is distributed according to

$$T(r_{1:N}) \in \begin{cases} \mathcal{N}\left(0, \frac{\sigma^2}{N}\right) & \text{under } H_0 \\ \mathcal{N}\left(A, \frac{\sigma^2}{N}\right) & \text{under } H_1 \end{cases}$$

Now the optimal test quantity for the detection setup (8) has been determined. Next step is to determine the optimal threshold, i.e. the threshold that minimizes $P(SF)$. For this, let $Q(x)$ denote the normal cumulative distribution function

$$Q(x) = \int_{-\infty}^{x} \varphi(x)\, dx$$

where

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

The probability for false alarm and detection are then

$$P_{FA} = P(T(r_{1:N}) > h|H_0) = 1 - Q\left(\frac{h}{\sqrt{\sigma^2/N}}\right) \tag{11a}$$

$$P_D = P(T(r_{1:N}) > h|H_1) = 1 - Q\left(\frac{h - A}{\sqrt{\sigma^2/N}}\right) \tag{11b}$$

With the expressions for P_{FA} and P_D it is straightforward to derive the expression for the ROC curve by noting that

$$h = \sqrt{\frac{\sigma^2}{N}} Q^{-1}(1 - P_{FA})$$
$$P_D = 1 - Q\left(Q^{-1}(1 - P_{FA}) - \sqrt{\frac{NA^2}{\sigma^2}} \right)$$

The last expression shows that P_D is a function of the desired P_{FA} and SNR only.

The optimal threshold is derived by substituting (11) into (6) and differentiating with respect to h. The condition for an optimal h is then

$$\beta\varphi\left(\frac{h-A}{\sqrt{\sigma^2/N}}\right) = \alpha\varphi\left(\frac{h}{\sqrt{\sigma^2/N}}\right)$$

which is equivalent to

$$\exp\left(-\frac{(h-A)^2}{2\sigma^2/N}\right) \Big/ \exp\left(-\frac{h^2}{2\sigma^2/N}\right) = \lambda$$

Taking logarithm gives

$$\frac{2Ah - A^2}{2\sigma^2/N} = \log\lambda$$

and solving for h gives the solution

$$h = \frac{A}{2} + \frac{\sigma^2}{NA}\log\lambda$$

Comparing the solution with (10) gives that

$$\bar{h} = \lambda$$

thus, the optimal threshold is in fact equal to the influence ratio λ. In the calculations, it is assumed that λ is positive. In the example in Figure 1 this assumption is always fulfilled. For more general systems, this might not always be the case. See Section 5 for further discussions.

4.3 Discussion

In Section 3 the influence ratio λ was introduced. Sections 4.1 and 4.2 have shown that λ is of central importance concerning threshold selection.

Now, the example in Figure 1 will be used to further discuss the results. Let p_1 and p_2 denote the probabilities of failure of the nominal and backup system respectively. Then, the probability of system failure, as a function of detection threshold h, is

$$P(SF(h)) = (1-p_1)p_2P_{FA}(h) - p_1(1-p_2)P_D(h) + p_1 \tag{12}$$

Thus, the influence ratio is

$$\lambda = \frac{(1-p_1)p_2}{(1-p_2)p_1}$$

For small probabilities p_i we have the approximation

$$\lambda \approx \frac{p_2}{p_1}$$

This means that, for any designed test quantity, at the optimal point on the ROC curve it holds that

$$\frac{dP_D}{dP_{FA}} \approx \frac{p_2}{p_1}$$

The geometric interpretation is shown in Figure 3 where the system property gives the influence ratio and the designed test the ROC curve. It can be seen in the figure that a larger influence ratio, i.e. a relatively more unreliable backup system, leads to that the diagnostic system should prioritize false alarms compared to detection performance.

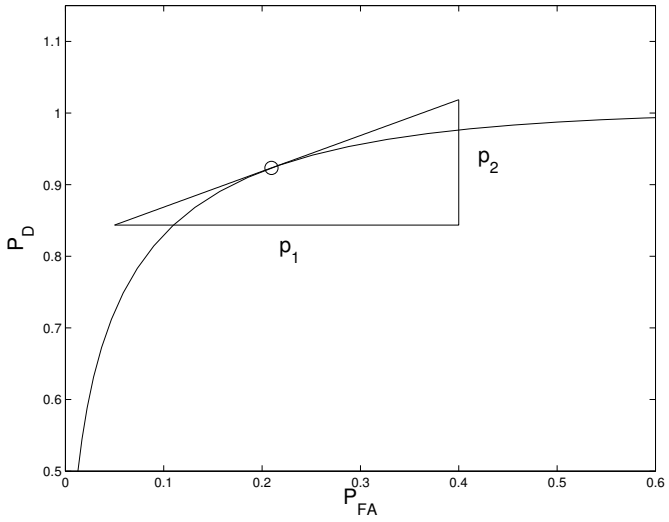

Fig. 3. Geometrical interpretation of the influence ratio in a ROC curve.

When considering a more specific test quantity, like in Section 4.2, the influence ratio is equal to the optimal threshold. For the example $\bar{h} \approx p_2/p_1$ which means that the detection test (9) becomes

$$\frac{P(r_{1:N}|H_1)}{P(r_{1:N}|H_0)} > \frac{p_2}{p_1}$$

The right hand side is a ratio between the probabilities of failure for the backup and nominal systems. The left hand side is, similarly, a ratio between probabilities of the observations given the no fault mode and the faulty mode.

5. GENERAL SYSTEMS

Discussions in previous sections have mainly focused on the system in Figure 1. This section will discuss how more general systems can be handled and show that the results apply directly. The analysis was based on the fact that we had the following expression for the probability for system failure:

$$P(SF) = \alpha P_{FA} - \beta P_D + \gamma$$

A quite general structure of a fault tolerant system is that we switch between different system configurations depending on alarm state. A strategy may for example be to switch to a backup system or to reconfigure the control strategy in case of an alarm.

First we will illustrate the results for a strategy for fault tolerance with a representative example. Then we show that the linearity property of (6) holds in a general setting. Finally, we give some examples and a discussion on cases where the fault detection system does not contribute to fault tolerance.

5.1 Analysis of a fault tolerant system

In this section an example is studied to show how the analysis can be extended to more complex systems. The system relies on three sensors, $s_1, \ldots s_3$ and is designed

such that it fails if sensor s_1 and sensor s_2 or s_3 fails. To improve system reliability, a diagnosis system is implemented that supervises sensor s_1. In case of a fault on s_1 is detected, the quantity that s_1 measures is instead estimated by an observer that uses sensor s_2 as feedback. A fault tree that describes the complete logic for when a system failure occurs can be seen in Figure 4. Events e_i

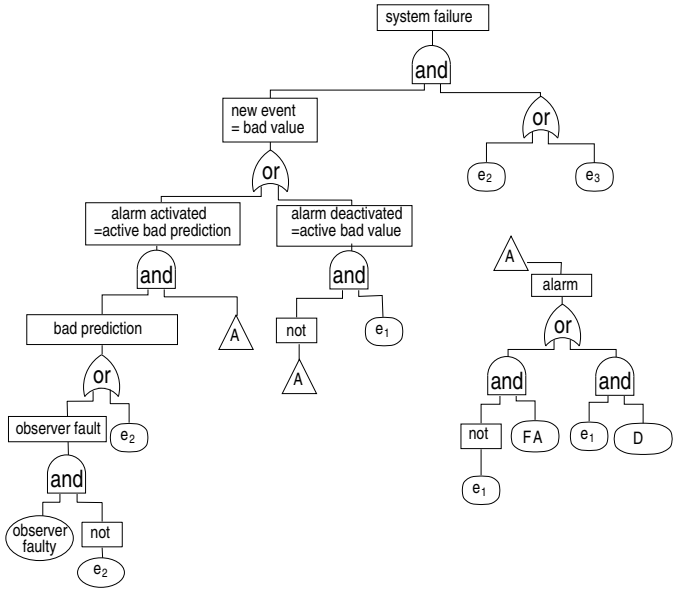

Fig. 4. Fault tree for the example system.

represents failure of sensor s_i, event FA false alarm, event D that the diagnosis system detects a fault, and event "observer faulty" means that the observer delivers faulty estimates even though there is no fault in the system. This may be, for example, due to a poor process model. The fault tree was derived using the approach in [ÅBF+07]. It is assumed that all sensors fail independently.

To be able to make a probabilistic analysis of the system, we introduce the probabilities

$$P(e_i) = P(\text{sensor } s_i \text{ fails}) = p_{si}$$

The fault detection system is characterized by the parameters $\sigma^2 = 1$, $A = 1$, and $N = 5$ from Section 4.2. The probabilities used in the example are set to

$$p_{obs} = 10^{-2},\ p_{s1} = 2 \cdot 10^{-1},\ p_{s2} = 10^{-3},\ p_{s3} = 10^{-3}$$

Now, computing the probability for system failure gives

$$P(SF) = \alpha P_{FA} - \beta P_D + \gamma$$

where

$$\begin{aligned}\alpha &= (1 - p_{s1})(p_{obs}p_{s3} + p_{s2}(1 - p_{obs}p_{s3}))\\ \beta &= p_{s1}p_{s3}(1 - p_{obs})(1 - p_{s2})\\ \gamma &= p_{s1}(p_{s2} + p_{s3} - p_{s2}p_{s3})\end{aligned}$$

Thus, the linearity property of $P(SF)$ holds in this case and it will be proven below that this also holds in a general setting. In Figure 5, the probability $P(SF)$ plotted as a function of the threshold h where the minimum is clearly visible.

The function $P(SF)$ has limit values as $h \to \pm\infty$ according to

$$\lim_{h\to\infty} P(SF) = \gamma, \qquad \lim_{h\to-\infty} P(SF) = \alpha - \beta + \gamma$$

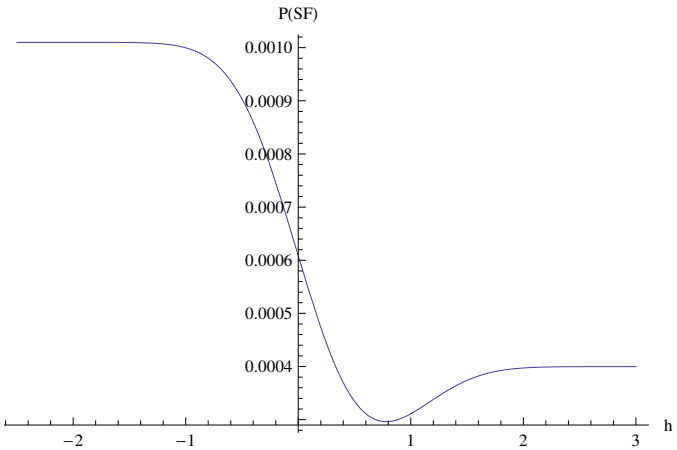

Fig. 5. The probability $P(SF)$ plotted as a function of the threshold h.

These two limits correspond to values 0 and ∞ of the threshold $\bar{h}$ in (9).

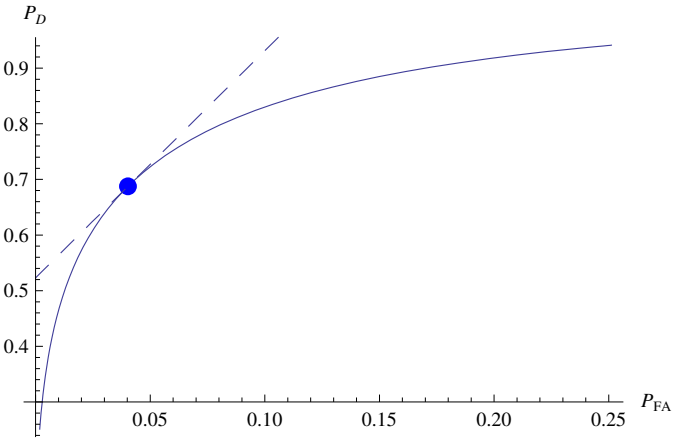

Fig. 6. ROC curve for the example system.

In this case, the test is described in Section 4.2 and the optimal threshold is $\bar{h} = \alpha/\beta$, i.e.

$$\bar{h} = \frac{(1 - p_{s1})(p_{obs}p_{s3} + p_{s2}(1 - p_{obs}p_{s3}))}{p_{s1}p_{s3}(1 - p_{obs})(1 - p_{s2})} = 4.1$$

The interpretation of $\lambda = \alpha/\beta$ in the ROC curve that was described in Section 4.1 is illustrated for this example in Figure 6. In this case the optimal point is characterized by

$$\frac{dP_D}{dP_{FA}} = \lambda = 4.1$$

5.2 Linearity of the system failure probability

Now, we prove that the linearity of $P(SF)$ holds in a more general setting than the case in Figure 1. Depending on the alarm state, the system is configured differently as seen in the previous example.

If we have no alarm and the original configuration is used, then we use the notation C_1 for the case where the configuration is working properly and $\neg C_1$ for the case that it fails. The cases C_2 and $\neg C_2$ are defined analogously for the backup configuration.

Theorem 1. Let S be the supervised event. Then the system failure probability is given by

$$P(SF) = \alpha P_{FA} - \beta P_D + \gamma$$

where

$$\alpha = P(\neg \mathrm{C}_2 \wedge \neg \mathrm{S}) - P(\neg \mathrm{C}_1 \wedge \mathrm{S})$$
$$\beta = P(\neg \mathrm{C}_1 \wedge \neg \mathrm{S}) - P(\neg \mathrm{C}_2 \wedge \mathrm{S})$$
$$\gamma = P(\neg \mathrm{C}_1)$$

Proof. As before, the system failure even can be expanded into mutually exclusive cases as:

$$P(SF) = P(\neg \mathrm{A} \wedge \neg \mathrm{C}_1) + P(\mathrm{A} \wedge \neg \mathrm{C}_2) = P(\neg \mathrm{A} \wedge \neg \mathrm{C}_1 \wedge \mathrm{S}) + P(\neg \mathrm{A} \wedge \neg \mathrm{C}_1 \wedge \neg \mathrm{S}) + P(\mathrm{A} \wedge \neg \mathrm{C}_2 \wedge \mathrm{S}) + P(\mathrm{A} \wedge \neg \mathrm{C}_2 \wedge \neg \mathrm{S})$$

where A is the alarm event. Expansion of each of the four terms gives

$$P(\neg \mathrm{A} \wedge \neg \mathrm{C}_1 \wedge \mathrm{S}) = P(\neg \mathrm{A} | \neg \mathrm{C}_1 \wedge \mathrm{S}) P(\neg \mathrm{C}_1 \wedge \mathrm{S}) = P(\neg \mathrm{A} | \mathrm{S}) P(\neg \mathrm{C}_1 \wedge \mathrm{S}) = (1 - P_{FA}) P(\neg \mathrm{C}_1 \wedge \mathrm{S})$$

$$P(\neg \mathrm{A} \wedge \neg \mathrm{C}_1 \wedge \neg \mathrm{S}) = P(\neg \mathrm{A} | \neg \mathrm{C}_1 \wedge \neg \mathrm{S}) P(\neg \mathrm{C}_1 \wedge \neg \mathrm{S}) = P(\neg \mathrm{A} | \neg \mathrm{S}) P(\neg \mathrm{C}_1 \wedge \neg \mathrm{S}) = (1 - P_D) P(\neg \mathrm{C}_1 \wedge \neg \mathrm{S})$$

$$P(\mathrm{A} \wedge \neg \mathrm{C}_2 \wedge \mathrm{S}) = P(\mathrm{A} | \neg \mathrm{C}_2 \wedge \mathrm{S}) P(\neg \mathrm{C}_2 \wedge \mathrm{S}) = P(\mathrm{A} | \mathrm{S}) P(\neg \mathrm{C}_2 \wedge \mathrm{S}) = P_D P(\neg \mathrm{C}_2 \wedge \mathrm{S})$$

$$P(\mathrm{A} \wedge \neg \mathrm{C}_2 \wedge \neg \mathrm{S}) = P(\mathrm{A} | \neg \mathrm{C}_2 \wedge \neg \mathrm{S}) P(\neg \mathrm{C}_2 \wedge \neg \mathrm{S}) = P(\mathrm{A} | \neg \mathrm{S}) P(\neg \mathrm{C}_2 \wedge \neg \mathrm{S}) = P_{FA} P(\neg \mathrm{C}_2 \wedge \neg \mathrm{S})$$

Collecting the expressions proves that $P(SF)$ is linear in P_{FA} and P_D, i.e.

$$P(SF) = \alpha P_{FA} - \beta P_D + \gamma$$

where the coefficients α, β, and γ are given in the theorem. □

5.3 Example of superfluous fault detectors

In the analysis in previous sections it was assumed that both α and β were positive. This is not always the case as will be illustrated with two examples followed by a discussion and an interpretation.

First, consider the case shown in Figure 7. Let event e_i

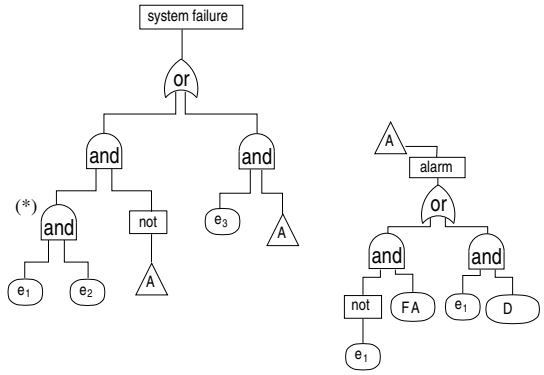

Fig. 7. Example system.

correspond to the failure of a sensor s_i. Here, the first configuration is that we have system failure if event e_1 and e_2 occur. In case of alarm, the system switches to a mode where event e_3 causes system failure. The detection system alarms if event e_1 is detected, i.e. the supervised event S is e_1. In this case it holds that

$$\beta = p_1(p_2 - p_3)$$

which is negative if $p_2 < p_3$. This means that detecting a fault, even if the alarm is correct, will increase the probability of system failure. The reason is that even if we know that sensor s_1 has failed, it is still better not to switch to sensor s_3 since sensor s_2 is more reliable. It is therefore better to never alarm and the detection system is superfluous.

As a second example, consider the same system as in Figure 7 where the AND-gate indicated by (*) is replaced by an OR-gate. Then we have

$$\alpha = (1 - p_1)(p_3 - p_2)$$

which is negative if $p_3 < p_2$. In such a case, a false alarm has a positive influence on the system safety. The reason for this is that even if we know that sensor s_1 has not failed, it is still better to switch to sensor s_3, since it is more reliable than sensor s_2. Also here the detection system is superfluous since it is always better to use sensor s_3.

6. CONCLUSIONS

In this paper we have considered threshold selection in fault tolerant systems. One of the fundamental questions has been to analyze the influence of fault detection properties on the overall system safety. Two important fault detection properties are probability of false alarm and detection power and it has been shown that the probability of system failure can be expressed in the form (6) for a general setting. Based on this form, the influence ratio was introduced in (7). This ratio includes all information about the supervised system and the backup system that is needed for the threshold optimization problem. It has been shown that the influence ratio has a geometrical interpretation as the gradient of the ROC curve at the optimal point. Furthermore, it was shown that the influence ratio is the threshold for the optimal test quantity given by the likelihood ratio in the case studied in Section 4.2. Finally, two examples were given that illustrate cases where the fault detectors are superfluous.

REFERENCES

[ÅBF+07] Jan Åslund, Jonas Biteus, Erik Frisk, Mattias Krysander, and Lars Nielsen. Safety analysis of autonomous systems by extended fault tree analysis. *International Journal of Adaptive Control and Signal Processing*, 21(2-3):287–298, 2007.

[Ber85] James O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer, 1985.

[BKLS03] M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswiecki. *Diagnosis and Fault-Tolerant Control*. Springer, 2003.

[Kay98] S.M. Kay. *Fundamentals of signal processing – detection theory*. Prentice Hall, 1998.

[Leh91] E.L. Lehmann. *Testing statistical hypothesis*. Statistical/Probability series. Wadsworth & Brooks/Cole, 1991.

[Vil92] Alain Villemeur. *Reliability, availability, maintainability and safety assessment*, volume 1 & 2. John Wiley & sons, U.K., 1992.