A Probabilistic Method for Certification of Analytically Redundant Systems

Bin Hu and Peter Seiler

Abstract— **Analytical fault detection algorithms have the potential to reduce the size, power and weight of fault tolerant safety-critical aerospace systems. One obstacle is the need for appropriate tools to certify the reliability of these systems. To complement high fidelity Monte Carlo simulations, this paper presents a theoretical method to assess the probabilistic performance of analytically redundant systems. Specifically, this paper considers a dual-redundant fault tolerant system that uses a fault detection algorithm to switch between the hardware components. The exact system failure rate per hour is computed using the law of total probability. The analysis assumes known failure models for the hardware components as well as knowledge of the probabilistic performance of the fault detection logic. A numerical example is provided to demonstrate the proposed method.**

I. Introduction

Commercial flight control electronics typically satisfy reliability and safety requirements of no more than 10^{-9} catastrophic failures per flight hour [3], [5]. Therefore, fault tolerance is introduced to enable continued operation in the event of a component failure. Fault tolerance is currently achieved mainly through the use of physically redundant components. For example, the Boeing 777 flight control electronics consists of multiple redundant computing modules, actuators, and sensors [21]. Physically redundant architectures are reliable but they increase the system size, weight, power, and cost. In addition, these architectures are impractical for smaller unmanned aerial vehicles which cannot carry the associated payload. As a result, there have been efforts to develop analytical redundancy as an alternative approach to achieve fault tolerance, e.g. the oscillatory monitors on the Airbus A380 [8]. Model-based fault detection and isolation (FDI) is one method to realize analytical redundancy. This technique has applications which span most disciplines of engineering [12] and a thorough treatment can be found in standard references [4], [11], [6]. The recent AddSafe project in Europe [1] dealt with the future green aircraft and assessed the suitability of these more advanced fault detection methods for optimizing the aircraft design.

There are several issues that must be addressed before analytical redundancy finds general acceptance for aerospace applications. One issue is the need to certify the reliability of an analytically redundant system with aviation authorities, e.g. the Federal Aviation Administration or European Aviation Safety Agency. In particular, it must be possible to assess and certify the system reliability. In a physically redundant configuration, a failed component is detected by directly comparing the behavior of each redundant component. Hence, these architectures detect faults accurately and their performance can be certified using fault trees and known hardware component failure rates [14], [13]. Systems that use analytical redundancy, on the other hand, depend on the fault detection algorithm as well as the hardware component failure rates. Thus different tools are required to assess the reliability of analytically redundant systems.

The extended fault tree technique has been proposed to assess the reliability of an analytically redundant system [2], [9]. In this work, the fault detection performance involves missed detections and false alarms that occur at the system sample rate. The system failure rate per sample frame is computed by characterizing false alarms and missed detections as basic events that are incorporated into a fault tree. However, the safety requirements are typically specified over longer time periods, e.g. per hour [3], [5]. The possible failure of the entire system at different time steps introduces time correlations which should be addressed properly.

The main contribution of this paper is a mathematical framework to efficiently compute the system failure rate per hour of an analytically redundant system. The framework proposed here builds on the prior work in [2], [9] by incorporating various time scales. The method is described for a simple dual-redundant sensor configuration with a fault detection scheme, as formulated in Section II. The system failure rate per hour is exactly computed using probabilistic models of the fault detection performance and the hardware component failures (Section III). Finally, a numerical example is presented to demonstrate the utility of the proposed approach (Section IV). The proposed approach is complementary to Monte Carlo simulations. In particular, the Monte Carlo method is a practical solution to assess system performance via simulations on a high fidelity model [17]. A potential drawback is that the failure rate for safety critical systems is designed to be very low. Thus a large number of Monte Carlo simulations may be required to draw statistically meaningful conclusions. The proposed mathematical analysis provides an efficient method to exactly compute the system reliability. In addition, the analysis provides additional insight into the various design choices. However the analysis requires specific assumptions about the failure models, operating conditions, etc. The use of both theoretical analysis and high fidelity simulations thus provides complementary benefits. This is similar to the current practice for flight control law validation [10], [16] which uses a mixture of high fidelity nonlinear simulations and exact analyses, e.g. gain/phase margins based on approximate linearized models.

Bin Hu and Peter Seiler are with the Aerospace Engineering and Mechanics Department, University of Minnesota, Email: huxxx221@umn.edu; seiler@aem.umn.edu.

II. Duplex Sensor System

Consider a dual-redundant sensor system operating in discrete-time (Figure 1). At each sample time k, the duplex system attempts to generate a "correct" measurement $\hat{m}(k)$ of a particular signal $s(k)$ for use by a flight control algorithm. Fault tolerance is achieved by the combination of two sensors and a fault detection scheme. At each sample time the two sensors generate measurements $m_1(k)$ and $m_2(k)$ of the signal $s(k)$. The measurement from the primary sensor is used in the absence of a detected fault. The system switches to the backup sensor once a fault is detected. The fault detection scheme is assumed to be an analytical method that relies on measurements that are independent of those generated by the primary and backup sensors. The objective is to assess the reliability of this duplex system. The duplex system shown in Figure 1 is simplified but captures the essential features of a more realistic redundant architecture.

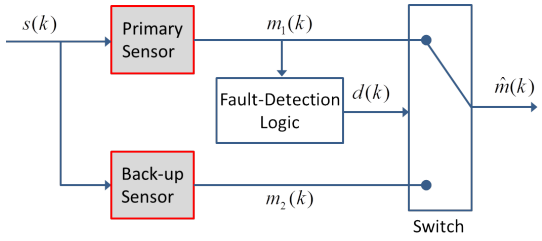

Fig. 1. Duplex Sensor System

A. Problem Formulation

A definition of reliability was established by the Technical Committee on Fault Detection, Supervision and Safety of Technical Processes [12]. Reliability is the ability of a system to perform a required function under stated conditions, within a given scope, and during a given period of time. Two aspects of this definition should be clarified for the duplex sensor system. First, the analysis in this paper is formulated in discrete-time. Hence the given period of time is a window of length N. Typical aerospace requirements are specified per hour and hence N may be large, e.g. $N = 3.6 \times 10^5$ samples per hour for a system with a 100 Hz sample rate. Second, the required function for the duplex system is to generate a "correct" measurement for use by a control law. The control laws and aircraft dynamics have low pass characteristics and thus a single "bad" sample may not lead to system failure. However, the continued use of incorrect data over multiple (N_0) time frames will eventually cause a failure. To summarize, the duplex system performs its required function as long as it does not generate "bad" data for N_0 consecutive steps. $P_{S,N}$ is defined as the probability that the system fails to perform this required function over an N-step window.

The analysis requires models of the sensor components. Let $\theta_i(k) \in \{0, 1\}$ denote the status of the i^{th} sensor ($i = 1, 2$) at time k: $\theta_i(k) = 0$ if the i^{th} sensor is operational at time k and $\theta_i(k) = 1$ if it has failed. It is assumed that once a sensor fails then it remains failed, i.e. intermittent failures are neglected. Due to this assumption it is possible to define a unique failure time T_i for the i^{th} sensor ($i = 1, 2$) as:

$$T_i = \begin{cases} k & \text{if } \theta_i(k-1) = 0 \text{ and } \theta_i(k) = 1 \\ N+1 & \text{if } \theta_i(k) = 0 \ \forall k \leq N \end{cases} \quad (1)$$

The notation $T_i = N + 1$ corresponds to the case where the sensor remains functional during the entire N-step window. Reliability theory can be used to model the failure time of the sensors [18], [15]. In many applications, the mean time between failure (MTBF) can be estimated from field data. The analysis in this paper assumes the probability mass function $P[T_i = k]$ is known for both sensors $i = 1, 2$ and for all time $k \leq N+1$. Finally, it is assumed that T_1 and T_2 are independent. This assumption implies dissimilar sensors are used and hence common failure modes are neglected. This simplifies the notation and computation required for analysis. A similar approach to that presented in this paper can be used to incorporate the joint probability mass function of T_1 and T_2 for systems with correlated sensor failures.

The probability of system failure $P_{S,N}$ also depends on the fault detection logic. The FDI scheme has a logic signal $d(k)$ that indicates the status of the primary sensor at time k: $d(k) = 1$ if a fault has been detected and $d(k) = 0$ otherwise. The FDI logic switches immediately to the backup sensor once a fault is detected. Thus the logic selects the primary sensor, $\hat{m}(k) = m_1(k)$, if $d(k) = 0$ and it selects the backup sensor if $d(k) = 1$. It is assumed that once the fault detection logic switches to the backup sensor then it will continue using the backup. Logic that intermittently switches between sensors is not considered. Again, this assumption implies that it is possible to define a unique switching time T_S as:

$$T_S = \begin{cases} k & \text{if } d(k-1) = 0 \text{ and } d(k) = 1 \\ N+1 & \text{if } d(k) = 0 \ \forall k \leq N \end{cases} \quad (2)$$

$T_S = N + 1$ denotes the case where no fault is detected throughout the entire N-step window.

The system can be in one of four states depending on the primary sensor status and the fault detection signal. These four states can be arranged in a confusion matrix [7] as shown in Table I. The entries of the confusion matrix depend on both the hardware and the FDI logic. The performance of the FDI logic alone is typically quantified by (single-frame) conditional probabilities of false alarm and detection. Specifically in [6], [20], the probability of false alarm at time k is defined as $P[d(k) = 1 \mid \theta_1(k) = 0]$. Similarly, the probability of detection at time k is defined as $P[d(k) = 1 \mid \theta_1(k) = 1]$. As shown in Section III, these single frame conditional probabilities are not sufficient to compute the system failure probability. Instead, computation of $P_{S,N}$ requires the FDI performance to be characterized across multiple time steps. The first FDI performance metric is $P[T_S \leq N \mid T_1 = N+1]$. This is the conditional probability of a false alarm at some point in the N-step window given that the primary sensor remains operational. The second FDI performance metric is $P[T_S \geq k + N_0 \mid T_1 = k]$ defined for $1 \leq k \leq N$. This is the conditional probability that the fault detection logic continues to use the primary sensor for at least N_0 steps after a failure at time k.

	$\theta_1(k) = 1$	$\theta_1(k) = 0$
$d(k) = 1$	True Positive	False Positive
$d(k) = 0$	False Negative	True Negative

TABLE I

CONFUSION MATRIX FOR FAULT DETECTION LOGIC

In the notation defined above, the duplex system produces bad data at time k if the primary sensor is selected and failed ($d(k) = 0$ and $\theta_1(k) = 1$) or the backup sensor is selected and failed ($d(k) = 1$ and $\theta_2(k) = 1$). Thus the system failure probability $P_{S,N}$ can be formally defined as:

Definition 1: $P_{S,N}$ is the probability that there exists $k_0 \leq N$ such that for each $k \in \{k_0, k_0 + 1, \ldots, k_0 + N_0 - 1\}$ one of the following is true:

1) $d(k) = 0$ and $\theta_1(k) = 1$
2) $d(k) = 1$ and $\theta_2(k) = 1$

and the sensor i selected at time $k_0 + N_0 - 1$ has a failure time within the N-step window ($T_i \leq N$).

By this definition, the system fails if it produces bad data for N_0 consecutive steps due to failures in the primary and/or backup sensor that occur within the N-step window. A system failure may occur due to a sequence of bad data beginning within the window ($k_0 \leq N$) and ending outside the window ($k_0 + N_0 - 1 > N$). The required detection time N_0 is typically much smaller than the analysis window N. Hence the choice of whether or not to include these boundary events should have negligible effect on $P_{S,N}$. Different assumptions regarding such boundary events can be handled with essentially notational changes.

B. Specific Example

As discussed above, the analysis in Section III only requires the following information:

1) Sensor Failure Model: $P[T_i = k]$ specified for $i = 1, 2$ and $1 \leq k \leq N$.
2) FDI False Alarm: $P[T_S \leq N \mid T_1 = N + 1]$.
3) FDI Missed Detection: $P[T_S \geq k + N_0 \mid T_1 = k]$ defined for $1 \leq k \leq N$.

This section briefly illustrates the notation in the context of a specific example. The example assumes sensor failures are governed by a geometric distribution and the FDI switching logic is independent and identically distributed (IID) in time.

First, assume the failure time of each sensor has an identical continuous-time exponential distribution with parameter $\lambda = \frac{1}{MTBF}$ [15]. The continuous-time exponential distribution can be approximated using a discrete-time geometric distribution with parameter $q := 1 - e^{-\lambda \Delta_t}$ where Δ_t is the sample time [19]. If the sensor is operational at $k = 0$ then it follows from the geometric distribution that the probability mass function for the sensor failures is given by:

$$P[T_i = k] = \begin{cases} (1-q)^{k-1} q & \text{if } 1 \leq k \leq N \\ (1-q)^N & \text{if } k = N + 1 \end{cases} \quad (3)$$

Let $P_F := P[d(k) = 1 \mid \theta_1(k) = 0]$ and $P_D := P[d(k) = 1 \mid \theta_1(k) = 1]$ denote the (single-frame) probability of false alarm and detection. The multiple-frame FDI performance probabilities can be related to these single-frame probabilities due to the assumption of FDI logic being IID. First, $P[T_S \leq N \mid T_1 = N+1]$ is the conditional probability that a fault is declared in the N step window given that the primary sensor remains operational. The set of sequences $\{d(k)\}_{k=1}^{N}$ where $d(k) = 1$ for at least one k is complementary to the sequence where $d(k) = 0$ for $1 \leq k \leq N$. Thus the multiple-frame false alarm probability can expressed in terms of the single frame probabilities as:

$$P[T_S \leq N \mid T_1 = N + 1] = 1 - (1 - P_F)^N \quad (4)$$

Next, $P[T_S \geq k+N_0 \mid T_1 = k]$ is the conditional probability that a fault is not declared in the first $k + N_0 - 1$ time steps given that sensor 1 failed at time k. This corresponds to a true negative for the first $k - 1$ steps followed by N_0 steps of false negatives. Thus this probability is expressed as:

$$P[T_S \geq k + N_0 \mid T_1 = k] = (1 - P_F)^{k-1}(1 - P_D)^{N_0} \quad (5)$$

III. PROBABILISTIC ANALYSIS

This section provides an exact expression for $P_{S,N}$. The analysis relies on basic probability theory with the law of total probability as the main tool. An application of this law is the following statement: Let the events $\{T_1 = k\}_{k=1}^{N+1}$ form a disjoint partition of the sample space. Then the probability of any other event $\mathcal{A}$ can be expressed as:

$$P[\mathcal{A}] = \Sigma_{k=1}^{N+1} P[\mathcal{A} \mid T_i = k] P[T_i = k] \quad (6)$$

A. General Theory

The dual redundant system fails to perform its required function if it generates "bad" data for N_0 consecutive steps. $P_{S,N}$ is the probability of the system failing to perform this function in an N-step window. There are four mutually exclusive events that lead to system failure:

1) Event M_N: The primary sensor fails at some time $k \leq N$ and the fault detection logic fails to switch within N_0 frames. This is a missed detection, denoted M_N.
2) Event F_N: The primary sensor remains operational during the entire N-step window. The fault detection logic has a false alarm and switches to the backup sensor but the backup sensor fails within the N-step window. This event is a false alarm, denoted F_N.
3) Event D_N: The primary sensor fails at some time $k \leq N$. The fault detection logic detects the failure within N_0 frames of the failure and correctly switches to the backup sensor. The backup sensor fails within the N-step window (either before or after the detected failure in the primary sensor). This event is a proper detection, denoted D_N, but results from failure in both sensors.
4) Event E_N: The primary fails at some time $k \leq N$. The fault detection logic raises a false alarm prior to time k and switches to the backup sensor but the backup sensor fails within the N-step window. This event is an early false alarm, denoted E_N.

The four events are mutually exclusive and hence:

$$P_{S,N} = P[M_N] + P[F_N] + P[D_N] + P[E_N] \quad (7)$$

The remainder of the section provides expressions for these four failure events. The first event is the missed detection M_N. The probability of a missed detection event can be expressed as $P[M_N] = P\left[\{T_1 \leq N\} \cap \{T_S \geq T_1 + N_0\}\right]$. Apply the law of total probability (Equation 6) to obtain:

$$P[M_N] = \Sigma_{k=1}^{N} P[T_S \geq k + N_0 \mid T_1 = k] P[T_1 = k] \quad (8)$$

The second event is the false alarm F_N. The false alarm event can be specified as $P[F_N] = P[\{T_1 = N+1\} \cap \{T_S \leq N\} \cap \{T_2 \leq N\}]$. The sensor failures are independent from each other. Moreover, the switching logic is independent of the backup sensor. Hence this probability is:

$$\begin{aligned} P[F_N] =& P[T_S \leq N | T_1 = N+1] \\ & P[T_1 = N+1] \; P[T_2 \leq N] \end{aligned} \quad (9)$$

The third event D_N involves a primary sensor failure and a true detection that causes a switch to the backup sensor. A failure of the backup sensor then leads to a system failure. Thus $P[D_N] = P[\{T_1 \leq N\} \cap \{T_1 \leq T_S < T_1 + N_0\} \cap \{T_2 \leq N\}]$. Similarly, the fourth event E_N also involves a primary sensor failure but in this case a false alarm causes a switch to the backup sensor prior to the primary sensor failure. The probability of this event can be expressed as $P[E_N] = P[\{T_1 \leq N\} \cap \{T_S < T_1\} \cap \{T_2 \leq N\}]$. The events D_N and E_N are mutually exclusive and combined as:

$$\begin{aligned} & P[D_N] + P[E_N] = \\ & P[\{T_1 \leq N\} \cap \{T_S < T_1 + N_0\} \cap \{T_2 \leq N\}] \end{aligned} \quad (10)$$

Apply the law of total probability to rewrite this as:

$$\begin{aligned} & P[D_N] + P[E_N] = \\ & \Sigma_{k=1}^{N} P[\{T_1 = k\} \cap \{T_S < T_1 + N_0\} \cap \{T_2 \leq N\}] \end{aligned} \quad (11)$$

The sensor failures and the the switching logic are independent and hence this can be expressed as:

$$\begin{aligned} & P[D_N] + P[E_N] = \\ & \Sigma_{k=1}^{N} P[T_S < k + N_0 \mid T_1 = k] P[T_1 = k] P[T_2 \leq N] \end{aligned} \quad (12)$$

Finally, we can compute the total system failure probability (Equation 7) by combining the probabilities for the basic failure events (Equations 8, 9, and 12). This yields the following expression for the system failure probability:

$$\begin{aligned} P_{S,N} = & \Sigma_{k=1}^{N} P[T_S \geq k + N_0 \mid T_1 = k] P[T_1 = k] \\ & + P[T_S \leq N \mid T_1 = N+1] P[T_1 = N+1] P[T_2 \leq N] \\ & + \Sigma_{k=1}^{N} P[T_S < k + N_0 \mid T_1 = k] P[T_1 = k] P[T_2 \leq N] \end{aligned} \quad (13)$$

Computing this system failure probability only requires the information specified in Section II. Specifically, the system failure probability can be computed from Equation 13 as long as the sensor failure $P[T_i = k]$, FDI false alarm $P[T_S \leq N \mid T_1 = N+1]$ and FDI missed detection $P[T_S \geq k + N_0 \mid T_1 = k]$ probabilities are all known. Moreover, it is not necessary for the FDI residuals to be Gaussian in order to compute these probabilities.

B. Specific Example

This section demonstrates the calculation of $P_{S,N}$ using the probabilities for the sensor and FDI performance (Equations 3-5) for the example in Section II-B. For this example the probabilities for the missed detection (Equation 8), false alarm (Equation 9), and combined detection / early false alarm (Equation 12) events can be computed as follows:

$$Pr[M_N] = q(1-P_D)^{N_0} \frac{1-(1-P_F)^N (1-q)^N}{1-(1-P_F)(1-q)} \quad (14)$$

$$Pr[F_N] = \left(1-(1-P_F)^N\right)(1-q)^N Pr[T_2 \leq N] \quad (15)$$

$$Pr[D_N] + Pr[E_N] = \left(Pr[T_1 \leq N] - Pr[M_N]\right) Pr[T_2 \leq N] \quad (16)$$

where $Pr[T_i \leq N] = 1-(1-q)^N$ $(i = 1, 2)$ based on the geometric sensor failure model (Equation 3). The exact system failure probability $P_{S,N}$ is given by the sum of Equations 14-16. This result can be simplified if further assumptions are made. If $Nq \ll 1$ and $NP_F \ll 1$ both hold then the event probabilities in Equations 14-16 can be approximated as:

$$Pr[M_N] \approx Nq(1-P_D)^{N_0} \quad (17)$$

$$Pr[F_N] \approx N^2 P_F q(1-Nq) \quad (18)$$

$$Pr[D_N] + Pr[E_N] \approx \left(1-(1-P_D)^{N_0}\right)(Nq)^2 \quad (19)$$

Next make the following definitions: $\hat{q} := Nq$, $\hat{P}_F := NP_F$ and $\hat{P}_D := 1-(1-P_D)^{N_0}$. Each of these definitions has a clear meaning. $\hat{q}$ and $\hat{P}_F$ are the approximate sensor failure and false alarm probabilities per hour, respectively. $\hat{P}_D$ is the approximate conditional probability of detection of a fault within the N_0 step detection window. With this notation the system failure probability is approximated as:

$$P_{S,N} \approx \hat{q}(1-\hat{P}_D) + \hat{P}_D \hat{q}^2 + \hat{P}_F \hat{q}(1-\hat{q}) \quad (20)$$

This approximation provides an intuition for the basic causes of system failure. The first term $\hat{q}(1-\hat{P}_D)$ is due to a a missed detection of a failed primary sensor. $\hat{P}_D \hat{q}^2$ accounts for the case where the FDI scheme detects a failed primary sensor but the backup sensor also fails. Finally, $\hat{P}_F \hat{q}(1-\hat{q})$ refers to the case where the primary sensor is functioning, the FDI scheme triggers a false alarm and then the backup sensor fails. The approximation in Equation 20 can be used to incorporate missed detections and false alarms as basic events in the extended fault tree analysis as described in [2], [9]. These approximations are intuitive but only justified for geometric failure models, IID fault detection logic, $Nq \ll 1$, and $NP_F \ll 1$. If any of these assumptions fail then Equations 14-16 should instead be used to compute $P_{S,N}$.

IV. Applications

This section provides a numerical example to demonstrate the proposed analysis method. The dual-redundant system is assumed to run at a 100Hz sample rate ($\Delta_t = 0.01$sec) with primary and back-up sensors that have a mean time between

failure of 1000 hours. This failure rate is approximated using a discrete-time geometric distribution with $q = 2.78 \times 10^{-9}$. The system fails if it produces bad data for at least $N_0 = 20$ consecutive samples. The FDI logic is described in further detail below. The objective is to compute the probability of failure for the dual-redundant system, $P_{S,N}$, using a window of length $N = 3.6 \times 10^5$. This corresponds to the per-hour system failure probability at the specified 100Hz sample rate. For comparison, note that the system failure rate per hour is 10^{-3} for a single-sensor architecture. For a triple-redundant architecture with simple hardware voting, the system will fail if any two of the three sensors fail. In this case the system failure rate per hour is approximately 3×10^{-6}. The system failure probability for the dual-redundant architecture with analytical FDI will be compared to these two extreme cases.

The logic for monitoring the primary sensor is assumed to be a model-based FDI algorithm. A typical model-based FDI scheme is compromised of two parts: a filter that generates a residual $r(k)$ and a decision function which determines the logic signal $d(k)$ that indicates the status of the primary sensor. There are many approaches to design the FDI filter, e.g. observers, parity equations, parameter estimators, and robust filters [4], [11], [6]. The filter output, $r(k)$, is a random variable and the objective is to design the filter to achieve a decoupling property: $r(k)$ has zero mean when $\theta_1(k) = 0$ and non-zero mean when $\theta_1(k) \neq 0$. For some filter designs, e.g. Kalman filters, the residual is uncorrelated in time. For these approaches it is reasonable to model $r(k)$ as:

$$r(k) = n(k) + \theta_1(k) f \tag{21}$$

where $n(k)$ is an IID noise process and f is assumed to be an additive bias fault that occurs when the sensor fails ($\theta_1(k) = 1$). $n(k)$ is assumed to be Gaussian with zero mean and variance σ^2. The decision logic generates the status signal $d(k)$ based on $r(k)$. Again, there are many different approaches to design the decision function, e.g. thresholding, statistical tests, and fuzzy logic [11], [6]. For simplicity, this example considers the use of a constant thresholding logic:

$$d(k) := \begin{cases} 1 & \text{if } |r(j)| > H \text{ for some } j \leq k \\ 0 & \text{else} \end{cases} \tag{22}$$

In other words, a fault is declared when the residual magnitude exceeds the threshold H. Note that this decision logic does not have intermittent switching, i.e. $d(k)$ remains at 1 once the residual exceeds the threshold. This fault detection logic is IID in time due to the given assumptions on $r(k)$ and hence the system failure probability $P_{S,N}$ can be computed using the results in Section III-B. Recall the definition of the single-frame false alarm and detection probabilities: $P_F := P[d(k) = 1 \mid \theta_1(k) = 0]$ and $P_D := P[d(k) = 1 \mid \theta_1(k) = 1]$. The residual is Gaussian at each time and hence:

$$P_F = 1 - \int_{-H}^{H} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{r^2}{2\sigma^2}} dr \tag{23}$$

$$P_D = 1 - \int_{-H}^{H} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(r-f)^2}{2\sigma^2}} dr \tag{24}$$

These single-frame probabilities can be accurately and efficiently computed using the error function `erf` in `Matlab`.

The system failure probability can be computed from the results in Section III for specific values of the residual variance σ^2, fault level f, and threshold H. First, the single-frame false alarm and detection probabilities are computed using Equations 23 and 24. The exact probabilities for the basic failure events can be computed from Equations 14-16 using P_F, P_N, q, N_0, and N. There is no need to use the approximations (Equations 17-19) as the exact equations can be efficiently evaluated. Finally, the exact system probability is given by the sum of the basic failure event probabilities (Equation 7). These steps are equivalent to evaluating the general result in Equation 13. Finally, note that the single-frame FDI probabilities appear to depend independently on σ^2, f, and H. Equations 23 and 24 can be non-dimensionalized so that only the ratios $\frac{H}{\sigma}$ and $\frac{f}{\sigma}$ appear in the integrals. The remainder of the section considers the effect of $\frac{H}{\sigma}$ and $\frac{f}{\sigma}$ on $P_{S,N}$.

Figure 2 shows $P_{S,N}$ as a function of the normalized threshold $\frac{H}{\sigma}$ for three values of the normalized fault level $\frac{f}{\sigma}$ =1, 6, and 10. The vertical axis is a log-scale to highlight the changes in system performance as a function of the threshold. For small thresholds the system will rarely have a missed detection but will often trigger a false alarm. As a result, for sufficiently small thresholds the system has $P_{S,N} \approx 10^{-3}$ for all fault levels, i.e. the duplex system has similar reliability to the single sensor architecture. For large thresholds the system will rarely have a false alarm but it will also frequently have missed detections when failures occur. Thus the duplex system also has similar reliability as the single sensor system for large thresholds. For intermediate values of the threshold, the system failure probability depends on the ratio of the fault to noise level. For large fault levels ($\frac{f}{\sigma} = 10$) the threshold can be chosen to achieve a system failure probability near 10^{-6}. This probabilistic performance is even better than that achieved by the triplex system. However, the analysis of the duplex system neglects some effects, e.g. failure modes in any additional sensors used to compute the residual, and hence this ideal performance is unlikely be achievable in practice. For smaller fault levels ($\frac{f}{\sigma} = 6$) the system failure probability is higher and there is a smaller range of thresholds that achieve $P_{S,N}$ near 10^{-6}. Finally, for small fault sizes relative to the noise ($\frac{f}{\sigma} = 1$) the system has $P_{S,N} > 10^{-3}$ for some thresholds, i.e. the performance of the duplex system is even worst than that achieved by a single sensor. The results shown in Figure 2 were generated with the exact formulas for $P_{S,N}$ but the approximations discussed in Section III-B provide some intuition for the $\frac{f}{\sigma} = 1$ curve. Specifically, the approximation in Equation 20 can be expressed as:

$$P_{S,N} \approx \hat{q} + (\hat{P}_F - \hat{P}_D)\hat{q}(1 - \hat{q}) \tag{25}$$

If $\hat{P}_F \geq \hat{P}_D$ then $P_{S,N} \geq \hat{q}$. Thus the dual redundant system fails more often than a single sensor if the false alarm probability exceeds the detection probability.

The results in Figure 2 indicate the importance of proper threshold selection. For a given fault level $\frac{f}{\sigma}$, let $H^*(\frac{f}{\sigma})$

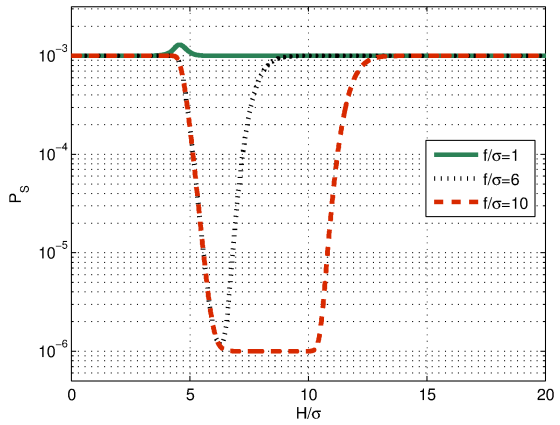

Fig. 2. $P_{S,N}$ vs. $\frac{H}{\sigma}$ for $N = 3.6 \times 10^5$

denote the threshold that minimizies $P_{S,N}$. Figure 3 shows $P_{S,N}$ as a function of the fault level for the optimal threshold H^* and for $H = 10\sigma$. As expected, for any fault level the system failure probability is lower with H^*. For both curves $P_{S,N}$ decreases monotonically with increasing fault level. Figure 3 also shows the limits of performance for the specific residual-based FDI scheme. In particular, for small fault levels ($f \leq 2\sigma$) the failure probability of the duplex system is similar to that of a single sensor system even if the optimal threshold is chosen. More advanced decision functions, e.g. likelihood ratio test, are required if the fault level is small relative to the noise.

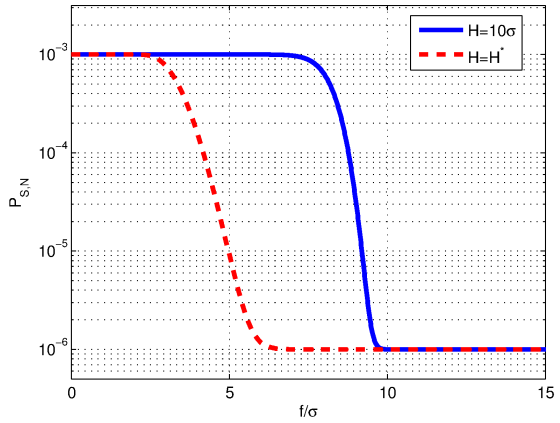

Fig. 3. $P_{S,N}$ vs. $\frac{f}{\sigma}$ for $N = 3.6 \times 10^5$

V. Conclusion

This paper analyzed the reliability of a dual-redundant system with analytical fault detection logic. The system failure probability can be exactly computed provided that probabilistic information is known for sensor failures and fault detection performance. The proposed approach can be combined with Monte Carlo simulations to assess system reliability. An example was given to demonstrate the approach. This example assumed the FDI decisions are independent in time. Future work will consider intermittent faults and time correlations that arise due to fault detection filters.

Acknowledgments

This work was supported by the National Science Foundation under Grant No. 0931931 entitled "CPS: Embedded Fault Detection for Low-Cost, Safety-Critical Systems". Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. This work was also supported by NASA under Grant No. NRA NNX12AM55A entitled "Analytical Validation Tools for Safety Critical Systems Under Loss-of-Control Conditions". The authors also acknowledge Brian Taylor for helpful discussions on assessing the reliability of avionics.

References

[1] ADDSAFE: Advanced fault diagnosis for sustainable flight guidance and control. http://addsafe.deimos-space.com/, 2012. European 7th Framework Program.

[2] Jan Aslund, 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.

[3] R.J. Bleeg. Commercial jet transport fly-by-wire architecture considerations. In *AIAA/IEEE DASC*, pages 399–406, 1988.

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

[5] R.P.G. Collinson. *Introduction to Avionic Systems*. Kluwer, 2003.

[6] S.X. Ding. *Model-Based Fault Diagnosis Techniques: Design Schemes, Algorithms, and Tools*. Springer-Verlag, 2008.

[7] T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27:861–874, 2006.

[8] P. Goupil. Oscillatory failure case detection in the A380 electrical flight control system by analytical redundancy. *Control Engineering Practice*, 18(9):1110–1119, 2010.

[9] F. Gustafsson, J. Aslund, E. Frisk, M. Krysander, and L. Nielsen. On threshold optimization in fault-tolerant systems. In *IFAC World Congress*, 2008.

[10] M. Heller, R. Niewoehner, and P. K. Lawson. On the validation of safety critical aircraft systems, part i: An overview of analytical and simulation method. In *AIAA GNC Conf.*, 2003.

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

[12] R. Isermann and P. Ballé. Trends in the application of model-based fault detection and diagnosis of technical processes. *Control Engineering Practice*, 5(5):709–719, 1997.

[13] M. Krasich. Use of fault tree analysis for evaluation of system-reliability improvements in design phase. *IEEE Proc. Reliability and Maintainability Symposium*, pages 1–7, 2000.

[14] W.S. Lee, D.L. Grosh, A.F. Tillman, and C.H. Lie. Fault tree analysis, methods, and applications: a review. *IEEE Trans. on Reliability*, 34(3):194–203, 1985.

[15] M. Rausand and A. Hoyland. *System Reliability Theory: Models, Statistical Methods, and Applications*. Wiley-Interscience, 2004.

[16] J. Renfrow, S. Liebler, and J. Denham. F-14 flight control law design, verification, and validation using computer aided engineering tools. In *IEEE Conference on Control Applications*, pages 359–364, 1994.

[17] C.P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, 2004.

[18] N. D. Singpurwalla. *Reliability and Risk: A Bayesian Perspective*. John Wiley & Sons, 2006.

[19] T. J. Wheeler, P. Seiler, A. K. Packard, and G. J. Balas. Performance analysis of fault detection systems based on analytically redundant linear time-invariant dynamics. In *Proceedings of the American Control Conference*, pages 214–219, 2011.

[20] A. S. Willsky and H. L. Jones. A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems. *IEEE Transactions on Automatic Control*, 21:108–112, 1976.

[21] Y.C. Yeh. Safety critical avionics for the 777 primary flight controls system. In *Proc. of the 20th DASC*, pages 1.C.2.1–1.C.2.11, 2001.