Int. J. Appl. Math. Comput. Sci., 2015, Vol. 25, No. 1, 103–116
DOI: 10.1515/amcs-2015-0008

amcs

A PROBABILISTIC METHOD FOR CERTIFICATION OF ANALYTICALLY REDUNDANT SYSTEMS

BIN HU [a], PETER SEILER [a,*]

[a] Aerospace Engineering and Mechanics Department
University of Minnesota, 107 Akerman Hall, 110 Union St. SE, Minneapolis, MN, USA
e-mail: {huxxx221,seile017}@umn.edu

Analytical fault detection algorithms have the potential to reduce the size, power and weight of safety-critical aerospace systems. Analytical redundancy has been successfully applied in many non-safety critical applications. However, acceptance for aerospace applications will require new methods to rigorously certify the impact of such algorithms on the overall system reliability. This paper presents a theoretical method to assess the probabilistic performance for an analytically redundant system. Specifically, a fault tolerant actuation system is considered. The system consists of dual-redundant actuators and an analytical fault detection algorithm to switch between the hardware components. The exact system failure rate per hour is computed using the law of total probability. This analysis requires knowledge of the failure rates for the hardware components. In addition, knowledge of specific probabilistic performance metrics for the fault detection logic is needed. Numerical examples are provided to demonstrate the proposed analysis method.

Keywords: avionics, certification, safety-critical systems, reliability, fault detection, fault-tolerant systems.

1. Introduction

Reliability and safety requirements for commercial flight control electronics are typically of the order of no more than 10^{-9} catastrophic failures per flight hour (Bleeg, 1988; Collinson, 2011). Therefore, fault tolerance is introduced to enable this safety-critical system to continue operation in the event of component failures. Fault tolerance is currently achieved mainly through the use of physically redundant components. For example, the Boeing 777 flight control electronics consist of three primary flight computing modules, each containing three dissimilar processors (Yeh, 1996; 2001). The actuators and sensors have similar levels of redundancy.

Physically redundant architectures are very reliable but they increase the system size, weight, power, and cost. As a result, there have been efforts to develop analytical redundancy as an alternative approach to achieve fault tolerance (Goupil, 2011). Recent examples include oscillatory monitors on the Airbus A380 (Goupil, 2010) and the *ADDSAFE* project in Europe (ADDSAFE, 2012). Small unmanned aerial vehicles (UAVs) represent another safety critical system that can benefit from analytical redundancy. The reliability of small UAVs is an emerging issue driven by the desire to integrate and fly such vehicles in conventional airspace (Vanek *et al.*, 2014). In the United States, a recent law (United States Congress, 2012) requires the Federal Aviation Administration (FAA) to "provide for the safe integration of civil unmanned aircraft systems into the national airspace system as soon as practicable, but not later than September 30, 2015." Small UAVs cannot carry the payload associated with physical redundancy and hence analytical redundancy will likely be required to improve their reliability.

There are several issues that must be addressed before analytical redundancy finds general acceptance in aerospace applications. One key issue is the need to rigorously assess the impact of analytical redundancy on the overall system reliability. A related issue is the need to certify the reliability of an analytically redundant system with aviation authorities, e.g., the FAA in the United States or the European Aviation Safety Agency. In particular, the system must not only be highly reliable and safe but it must also be possible to certify the system reliability and safety. In a physically redundant configuration, a failed component is detected by directly comparing the behavior of each redundant component.

*Corresponding author

Hence, these architectures tend to detect faults accurately and quickly. Moreover, their performance can be certified from known hardware component failure rates using a failure mode and effect analysis as well as a fault tree analysis (Lee *et al*., 1985; Krasich, 2000). The reliability of systems that use analytical redundancy, on the other hand, depends on the performance of the detection algorithm as well as the hardware component failure rates. New failure modes are introduced due to the mixed use of analytical algorithms and hardware components. Thus different tools are required to assess the reliability of analytically redundant systems.

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 proposed framework builds on the prior work of Åslund *et al*. (2007) and Gustafsson *et al*. (2008) as discussed below. The proposed analysis method is described for a simple dual-redundant actuator configuration with an analytical fault detection scheme. This problem formulation, described further in Section 2, is similar to the dual-redundant actuator architecture that has been implemented on the Airbus A380 (Goupil, 2010; Efimov *et al*., 2013). Our paper develops a probabilistic method to assess the reliability of the dual-redundant actuator system (Section 3). This method first enumerates all failure modes of the duplex system. Then the system failure rate per hour is exactly computed using the hardware component failure rates and probabilistic models of the fault detection performance. Section 4 applies the proposed framework to a concrete fault detection and isolation (FDI) scheme and briefly discusses techniques for computing the probabilistic FDI performance metrics. A numerical example is presented to demonstrate the utility of the proposed approach (Section 5). Finally, it is noted that this paper expands on the initial results published by the authors in a conference paper (Hu and Seiler, 2013).

Before continuing to the main result, the prior work that is relevant to this paper is briefly reviewed. The problem formulation in this paper includes an analytical fault detection scheme to switch between actuators. Model-based FDI is one method to realize this analytical redundancy. This technique has wide applications which span most disciplines of engineering (Isermann and Ballé, 1997), and a thorough treatment can be found in standard references (Chen and Patton, 1999; Isermann, 2006; Ding, 2008). Data-driven FDI methods provide an alternative means to detect faults. There has been some direct comparisons of model-based and data-driven methods (e.g., Freeman *et al*., 2013), but further work is needed to clarify the advantages of each FDI approach. The analysis framework proposed in this paper is applicable to either FDI approach provided certain probabilistic performance metrics (to be described more precisely in Section 2.1) can be computed for the fault detection logic.

The work most closely related to this paper is the extended fault tree technique given by Åslund *et al*. (2007) and Gustafsson *et al*. (2008). In the extended fault tree analysis, 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 (Bleeg, 1988; Collinson, 2011). The possible failure of the entire system at different time steps introduces time correlations and new failure modes which should be addressed properly. The framework described here builds on the prior work of Åslund *et al*. (2007) and Gustafsson *et al*. (2008) by incorporating events at various time scales.

The proposed approach is complementary to Monte Carlo simulations. In particular, those are commonly used in current industrial practice to assess system performance via simulations on a high fidelity model (Robert and Casella, 2004; Asmussen and Glynn, 2007). 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. As shown later in the paper, the analysis decouples the system failure rate into a certain hardware failure rate and FDI performance metrics. This decomposition also makes further worst case analysis possible when model uncertainty is significant. The main limitation of the analysis method is that it is valid only under specific assumptions about the failure models, operating conditions, etc. Thus the theoretical analysis and high fidelity simulations provide complementary benefits. This is similar to the current practice for flight control law validation (Renfrow *et al*., 1994; Heller *et al*., 2001; Belcastro and Belcastro, 2003) which uses a mixture of high fidelity nonlinear simulations and exact analyses, e.g., gain/phase margins based on approximate linearized models.

2. Duplex actuator system

Consider a dual-redundant actuator system operating in discrete-time (Fig. 1). At each sample time k, the duplex system attempts to move the control surface to a "correct" position based on a particular command signal $u(k)$ given by a flight control algorithm. Fault tolerance is achieved by the combination of two actuators and an FDI logic. At each sample time one of the two actuators is in active

mode and the other is in passive mode. The primary actuator is monitored by the FDI logic, and is used, i.e., is in active mode, in the absence of a detected fault. The FDI logic switches the system to a backup actuator once a fault is detected in the primary actuator. The FDI logic is assumed to be an analytical method, e.g., model-based or data-driven, that relies on the control commands $u(k)$ as well as a measurement $s_1(k)$ of the actual control surface position for the primary actuator. In practice, the FDI scheme can be designed in a variety of different ways. For a concrete example, Goupil (2010) used a model-based parity equation to generate a residual and then applied a spectral adaptive threshold as a decision function for detecting a fault. The duplex system shown in Fig. 1 is a simplified abstraction of the actual architecture on an Airbus A380 (Goupil, 2010). The abstraction captures the essential features of this kind of analytically redundant architecture. The objective is to assess the reliability of this duplex system.

2.1. Problem formulation. The following definition of reliability was established by the Technical Committee on Fault Detection, Supervision and Safety of Technical Processes.

Definition 1. (*Isermann and Ballé, 1997*) *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 actuator 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" control surface position. The control laws and aircraft dynamics typically have low pass characteristics, and thus incorrect operation of the actuation at a single sample time will not lead to system failure. However, the continued use of a "bad" control surface position 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 provide a "bad" control surface position 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.

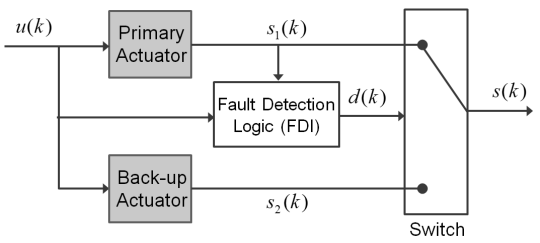


Fig. 1. Duplex actuator system.

The analysis requires models of the actuator components. Denote the primary and backup actuators by $i = 1$ and $i = 2$, respectively. Let $\theta_i(k) \in \{0, 1\}$ denote the status of the i-th actuator ($i = 1, 2$) at time k: $\theta_i(k) = 0$ if the i-th actuator is operational at time k and $\theta_i(k) = 1$ if it has failed. It is assumed that once an actuator 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 actuator ($i = 1, 2$) as

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

The notation $T_i = N + 1$ corresponds to the case where the actuator remains functional during the entire N-step window.

Reliability theory can be used to model the failure time of the actuators (Singpurwalla, 2006; Rausand and Hoyland, 2004). 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 actuators $i = 1, 2$ and for all time $k \le N + 1$. Finally, it is assumed that T_1 and T_2 are independent. This final assumption implies dissimilar actuators are used and hence common failure modes are neglected. The independence assumption can be considered reasonable and approximately true in many cases. For example, some control surfaces on an Airbus A380 are operated by two adjacent dissimilar actuators: an electro-hydrostatic actuator and a conventional hydraulic actuator (Goupil, 2011, Fig. 5). This assumption simplifies the notation and computation required for analysis. The correlated failure case will be briefly discussed later in the paper.

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 actuator at time k: $d(k) = 0$ if the primary actuator is in active mode and $d(k) = 1$ otherwise. Thus the logic uses the primary actuator, $s(k) = s_1(k)$, if $d(k) = 0$ and it uses the backup actuator if $d(k) = 1$. The primary actuator is turned off (passive mode) and the backup actuator is turned to the active mode once a fault is detected in the primary actuator. It is assumed that once the fault detection logic switches to the backup actuator then it will continue using the backup. Logic that intermittently switches between actuators is not considered. Again, this assumption implies that it is possible to define a unique

switching time T_S as

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

with $T_S = N + 1$ denoting 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 actuator status and the fault detection signal. These four states can be arranged in a confusion matrix (Egan, 1975; Fawcett, 2006) as shown in Table 1. 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 the works of Ding (2008) as well as Willsky and Jones (1976), 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 3, 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 using two specific metrics. The first FDI performance metric is $P[T_S \leq N \mid T_1 = N + 1]$. This is the conditional probability that the FDI logic switches to the backup actuator at some point in the N-step window given that the primary actuator remains operational. In other words, this is a false alarm probability over the N-step window. The second FDI performance metric is $P[T_S \geq k + N_0 \mid T_1 = k]$ defined for $k = 1, 2, \ldots, N$. This is the conditional probability that the FDI logic continues to select the primary actuator at least until time step $k + N_0$, given a primary actuator failure at time k. In other words, this is the probability of a missed detection conditioned on a failure at time k.

The dual-redundant system fails if the FDI logic selects a failed actuator. In the notation defined above, the duplex system produces a bad control surface position at time k if the primary actuator is selected and failed ($d(k) = 0$ and $\theta_1(k) = 1$) or the backup actuator 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 follows.

Definition 2. $P_{S,N}$ 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 actuator i selected at time $k_0 + N_0 - 1$ has a failure time within the N-step window ($T_i \leq N$).

Table 1. Confusion matrix for fault detection logic.

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

By this definition, the system fails if it produces a bad control surface position for N_0 consecutive steps due to failures in the primary and/or backup actuator that occur within the N-step window. A system failure may occur due to a sequence of bad control surface positions 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.

The proposed analysis method is developed for the dual redundant system formulated in this section. This dual redundant system is an active-passive architecture in the sense that at all times one actuator is active (on) while the other is passive (off). The proposed analysis can be extended to more cases, e.g., active-active dual redundant systems and triplex (or higher) redundant systems. The analysis for these extensions depends on the precise details of how the physically redundant components interact with the fault detection logic. The active-passive dual redundant architecture analyzed in this paper consists of logics to switch between two modes (use of the primary or back-up actuator). In general, the basic analysis approach can be extended to more complicated architectures if the fault tolerant system switches between several modes with a unique definition of the switching time.

2.2. Specific example. As discussed above, the analysis in Section 3 only requires the following information:

1. actuator 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 $k = 1, 2, \ldots, N$.

This section briefly illustrates the notation in the context of a specific example. The example assumes actuator 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 the i-th actuator has a continuous-time exponential distribution with

parameter $\lambda_i = \text{MTBF}_i{}^{-1}$ (Rausand and Hoyland, 2004). The continuous-time exponential distribution can be approximated using a discrete-time geometric distribution with parameter $q_i := 1 - e^{-\lambda_i \Delta_t}$, where Δ_t is the sample time (Wheeler *et al.*, 2011). If the actuator is operational at $k = 0$, then it follows from the geometric distribution that the probability mass function for the actuator failures is given by

$$P[T_i = k] = \begin{cases} (1-q_i)^{k-1} q_i & \text{if } 1 \le k \le N, \\ (1-q_i)^N & \text{if } k = N+1, \end{cases} \tag{3}$$

It is important to note that the actuator failure rates can be modeled by distributions other than the geometric distribution used here. For example, a discrete Weibull distribution (Murthy *et al.*, 2004; Nakagawa and Osaki, 1975; Stein and Dattero, 1984) can be used to model increasing failure rates as the actuator ages. This specific example uses the geometric distribution, but the proposed approach can accommodate any other discrete failure distribution. The specific choice of distribution needs to be validated based on failure rates of fielded components.

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) probabilities of false alarm and detection, respectively. The multiple-frame FDI performance probabilities can be related to these single-frame probabilities due to the assumption of FDI logic being IID. First, $P[T_S \le N \mid T_1 = N+1]$ is the conditional probability that a fault is declared in the N-step window given that the primary actuator 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 $1 \le k \le N$. Thus the multiple-frame false alarm probability can expressed in terms of the single frame probabilities as

$$P[T_S \le N \mid T_1 = N+1] = 1 - (1 - P_F)^N. \tag{4}$$

Next, $P[T_S \ge k + N_0 \mid T_1 = k]$ is the conditional probability that a fault is not declared in the first $k+N_0-1$ time steps given that the primary actuator ($i = 1$) 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

$$\begin{aligned} &P[T_S \ge k + N_0 \mid T_1 = k] \\ &\quad = (1 - P_F)^{k-1} (1 - P_D)^{N_0}. \end{aligned} \tag{5}$$

3. 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

$$P[\mathcal{A}] = \sum_{k=1}^{N+1} P[\mathcal{A} \mid T_1 = k] P[T_1 = k]. \tag{6}$$

This can also be expressed as

$$P[\mathcal{A}] = \sum_{k=1}^{N+1} P[\mathcal{A} \cap \{T_1 = k\}]. \tag{7}$$

3.1. General theory. The dual redundant system fails to perform its required function if it generates a "bad" control surface position for N_0 consecutive steps. $P_{S,N}$ is the probability of the system failing to perform this function in an N-step window. A failure modes-and-effects analysis should first be performed to identify all mutually exclusive failure modes leading to system failures. There are four mutually exclusive events that lead to system failure:

1. **Event M_N**: The primary actuator fails at some time $k \le N$ and the FDI logic fails to switch within N_0 frames. This is a missed detection, denoted by M_N.

2. **Event F_N**: The primary actuator remains operational during the entire N-step window. The fault detection logic has a false alarm and switches to the backup actuator but the backup actuator fails within the N-step window. This event is a false alarm induced failure, denoted by F_N.

3. **Event D_N**: The primary actuator fails at some time $k \le N$. The fault detection logic detects the failure within N_0 frames of the failure and correctly switches to the backup actuator. The backup actuator fails within the N-step window (either before or after the detected failure in the primary actuator). This event is a proper detection, denoted by D_N, but results from a failure in both actuators.

4. **Event E_N**: The primary actuator fails at some time $k \le N$. The fault detection logic raises a false alarm prior to time k and switches to the backup actuator but the backup actuator fails within the N-step window. This event is an early false alarm, denoted by E_N.

The four events are mutually exclusive and hence

$$P_{S,N} = P[M_N] + P[F_N] + P[D_N] + P[E_N]. \tag{8}$$

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 $P[M_N] =$

$P[\{T_1 \leq N\} \cap \{T_S \geq T_1 + N_0\}]$. Apply the law of total probability (Eqn. (6)) to obtain

$$P[M_N] = \sum_{k=1}^{N} P[T_S \geq k + N_0 \mid T_1 = k] P[T_1 = k] \quad (9)$$

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 actuator failures are independent of each other. Moreover, the switching logic is independent of the backup actuator. Hence this probability is

$$\begin{aligned} P[F_N] =& P[T_S \leq N | T_1 = N+1] \\ & \times P[T_1 = N+1]\, P[T_2 \leq N]. \end{aligned} \quad (10)$$

The third event D_N involves a primary actuator failure and a true detection that causes a switch to the backup actuator. A failure of the backup actuator 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 actuator failure but in this case a false alarm causes a switch to the backup actuator prior to the primary actuator 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

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

Apply the law of total probability to rewrite this as

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

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

$$\begin{aligned} & P[D_N] + P[E_N] \\ & = \sum_{k=1}^{N} P[T_S < k + N_0 \mid T_1 = k] P[T_1 = k] \\ & \quad \times P[T_2 \leq N]. \end{aligned} \quad (13)$$

Finally, we can compute the total system failure probability (Eqn. (8)) by combining the probabilities for the basic failure events (Eqns. (9), (10), and (13)). This yields the following expression for the system failure probability:

$$\begin{aligned} & P_{S,N} \\ & = \sum_{k=1}^{N} P[T_S \geq k + N_0 \mid T_1 = k] P[T_1 = k] \\ & + P[T_S \leq N \mid T_1 = N+1] P[T_1 = N+1] P[T_2 \leq N] \\ & + \sum_{k=1}^{N} P[T_S < k + N_0 \mid T_1 = k] P[T_1 = k] P[T_2 \leq N]. \end{aligned} \quad (14)$$

This equation provides an intuition for the basic causes of system failure. The first term is due to a a missed detection of a failed primary actuator. The second term refers to the case where the primary actuator is functioning, the FDI scheme triggers a false alarm and then the backup actuator fails. Finally, the third term accounts for the case where the primary actuator fails and the FDI scheme triggers an alarm but the backup actuator also fails. Computing this system failure probability only requires the information specified in Section 2 . Specifically, the system failure probability can be computed from Eqn. (14) as long as the probabilities of actuator failure $P[T_i = k]$, FDI false alarm $P[T_S \leq N \mid T_1 = N+1]$ and FDI missed detection $P[T_S \geq k + N_0 \mid T_1 = k]$ are all known.

The system failure probability in Eqn. (14) can be re-arranged into a more useful and intuitive form. Note that $T_S < k + N_0$ and $T_S \geq k + N_0$ are complementary events. This yields the following relation:

$$\begin{aligned} & P[T_S < k + N_0 \mid T_1 = k] \\ & \quad = 1 - P[T_S \geq k + N_0 \mid T_1 = k]. \end{aligned} \quad (15)$$

Substitute this for the last term in Eqn. (14) and regroup the result to obtain

$$\begin{aligned} & P_{S,N} \\ & = P[T_1 \leq N] P[T_2 \leq N] \\ & \quad + P[T_S \leq N \mid T_1 = N+1] P[T_1 = N+1] P[T_2 \leq N] \\ & \quad + \sum_{k=1}^{N} P[T_S \geq k + N_0 \mid T_1 = k] P[T_1 = k] \\ & \quad \times P[T_2 = N+1]. \end{aligned} \quad (16)$$

This equation provides another intuition for the basic causes of the system failure. The first term does not depend on the FDI performance and refers to the case where both the actuators fail. It provides a lower bound for the system failure rate $P_{S,N}$. No matter how well the FDI logic performs, the dual redundant system can not have a failure rate lower than this term. The second term is identical to the second term in Eqn. (14). This term refers to the case where the primary actuator is functioning, the FDI logic triggers a false alarm and then the backup actuator fails. The third term is due to a missed detection of a failed primary actuator given the condition

that the backup actuator does not fail. The three terms are due to three mutually exclusive failure modes. Equation (16) has the advantage that it decouples the causes of the system failure based on hardware component failures (term 1) and FDI performance (terms 2 and 3). This allows the effect of the FDI performance on the total system reliability to be fully separated from the reliability of the hardware components. This further enables the FDI logic to be designed and analyzed based on the false alarm and missed detection probabilities.

As described in Section 2.1, the analysis is based on the assumption that T_1 and T_2 are independent. The approach in this paper can, in theory, be extended to include correlated failures. The final results to compute $P_{S,N}$ with correlated failures can be found in Appendix. This extension requires knowledge of the joint probability mass function for the failure times T_1 and T_2 of the two actuators. Estimating this joint mass function would be impractical in most cases and this limits the utility of these generalizations. Finally, it is important to note that the analysis is exact in theory but sources of error will be introduced in practice. Specifically, the proposed framework requires knowledge of the actuator (hardware) failure probabilities along with the FDI performance metrics (false alarm and missed detection probabilities). Less accurate estimates of these performance metrics (either conservative or optimistic) will thus result in a less accurate estimate of the overall system reliability. Similar issues arise when constructing conventional fault trees that deal only with hardware failures.

3.2. Simplifying approximations. The FDI false alarm metric $P[T_S \leq N \mid T_1 = N+1]$ requires a single calculation. On the other hand, the FDI missed detection metric $P[T_S \geq k + N_0 \mid T_1 = k]$ depends on k and hence N computations are required. In certain circumstances, the following approximation can be used for $k = 1, 2, \ldots, N$:

$$P[T_S \geq k + N_0 \mid T_1 = k] \approx P[T_S \geq 1 + N_0 \mid T_1 = 1]. \quad (17)$$

This approximation enables the FDI missed detection metric to be evaluated for all $k = 1, 2, \ldots, N$ using only one calculation at $k = 1$. $P[T_S \geq 1 + N_0 \mid T_1 = 1]$ is the conditional probability that $d(k) = 0$ for all $k = 1, 2, \ldots, N$ given that the primary actuator fails at the first time step ($T_1 = 1$). This can be viewed as a missed detection probability over a detection window with size N_0. For many model-based FDI systems consisting of residual generation and decision logic, this approximation will hold if the FDI false alarm probability is very small. A rigorous derivation justifying this approximation is omitted since it is not the main focus of this paper.

The formula for the system failure probability $P_{S,N}$ simplifies by using this approximation. First make the following definitions:

$$\hat{q}_i := P[T_i \leq N], \quad (18)$$

$$\hat{P}_F := P[T_S \leq N \mid T_1 = N+1], \quad (19)$$

$$\hat{P}_D := 1 - P[T_S \geq 1 + N_0 \mid T_1 = 1]. \quad (20)$$

Each of these definitions has a clear meaning. Here $\hat{q}_i$ is the i-th actuator failure probability per hour and $\hat{P}_F$ is the false alarm probability per hour. $\hat{P}_D$ is the probability of detection of a fault within the N_0-step detection window conditioned on a primary actuator fault occurring at $k = 1$. The "hat" denotes that these probabilities are valid over multiple time steps, i.e., they are not simply single time frame probabilities.

With this notation and the assumption $P[T_S \geq k + N_0 \mid T_1 = k] \approx 1 - \hat{P}_D$, the system failure probability (Eqn. (16)) is approximated as

$$P_{S,N} \approx \hat{q}_1\hat{q}_2 + \hat{P}_F\hat{q}_2(1-\hat{q}_1) + (1-\hat{P}_D)\hat{q}_1(1-\hat{q}_2). \quad (21)$$

Equation (21) is an approximate form of Eqn. (16) and it provides intuition for the basic causes of system failures. For example, the first term in Eqn. (21) is $\hat{q}_1\hat{q}_2$ and this represents the failure probability due to faults in both actuators. The second term $\hat{P}_F\hat{q}_2(1-\hat{q}_1)$ accounts for the case where the FDI scheme raises a false alarm and then the backup actuator fails. The third term $(1-\hat{P}_D)\hat{q}_1(1-\hat{q}_2)$ is due to a missed detection of a failed primary actuator. A similar approximation can be derived for the system failure probability in the form given by Eqn. (14).

With the simplifying approximation, Eqn. (21) can be used to incorporate missed detections and false alarms as basic events in the extended fault tree analysis as described by Åslund *et al.* (2007) and Gustafsson *et al.* (2008). If the simplifying assumption in Eqn. (17) fails, then the exact formula in Eqn. (16) (or the alternative form in Eqn. (14) should instead be used to compute $P_{S,N}$.

3.3. Specific example. This section demonstrates the calculation of $P_{S,N}$ using the probabilities for the actuator and FDI performance (Eqns. (3)–(5)) for the example in Section 2.2. Recall that the example assumes actuator failures are governed by a geometric distribution with a single frame failure rate of q_i (Eqn. (3)). The probability of an actuator failure over N steps is thus explicitly given by $P[T_i \leq N] = 1 - (1-q_i)^N$ for $i = 1, 2$. The FDI switching logic is assumed to be IID in time with single-frame probabilities of false alarm and detection denoted by P_F and P_D, respectively. For this example, the multiple-step probabilities of missed detection (Eqn. (9)),

false alarm (Eqn. (10)), and combined detection/early false alarm (Eqn. (13)) events can be explicitly computed as

$$P[M_N] = q_1(1-P_D)^{N_0} \times \frac{1-(1-P_F)^N(1-q_1)^N}{1-(1-P_F)(1-q_1)}, \quad (22)$$

$$P[F_N] = \left(1-(1-P_F)^N\right)(1-\hat{q}_1)\hat{q}_2, \quad (23)$$

$$P[D_N] + P[E_N] = (\hat{q}_1 - P[M_N])\,\hat{q}_2, \quad (24)$$

where the notation $\hat{q}_i := P[T_i \leq N] = 1-(1-q_i)^N$ introduced in Eqn. (18) has been used. As derived in Section 3.1, the exact system failure probability $P_{S,N}$ is given by the sum of Eqns. (22)–(23).

The exact system failure probability for this example can be simplified as described in Section 3.2 if the approximation condition holds. For this example, the multiple-step false alarm and detection probabilities defined in Eqns. (19) and (20) are given by $\hat{P}_F = 1-(1-P_F)^N$ and $\hat{P}_D = 1-(1-P_D)^{N_0}$. To verify the approximation condition, first notice that

$$P[T_S \geq k+N_0 \mid T_1 = k] = (1-P_F)^{k-1}(1-\hat{P}_D). \quad (25)$$

Moreover, it is straightforward to show that

$$(1-\hat{P}_F)(1-\hat{P}_D) \leq (1-P_F)^{k-1}(1-\hat{P}_D) \leq 1-\hat{P}_D. \quad (26)$$

Thus $\hat{P}_F \ll 1$ implies that the approximation condition $P[T_S \geq k+N_0 \mid T_1 = k] \approx 1-\hat{P}_D = P[T_S \geq 1+N_0 \mid T_1 = 1]$ is valid. For many FDI schemes, the false alarm metric is low and $\hat{P}_F \ll 1$ holds. The approximation condition implies that $P[M_N] \approx (1-\hat{P}_D)\hat{q}_1$. Thus the total system failure probability $P_{S,N}$ given by the sum of Eqns. (22)–(23), can be written in the simplified form as

$$P_{S,N} \approx \hat{q}_1\hat{q}_2 + \hat{P}_F\hat{q}_2(1-\hat{q}_1) + (1-\hat{P}_D)\hat{q}_1(1-\hat{q}_2). \quad (27)$$

This is identical with the simplified formula in Eqn. (21). The main point is that various terms can be explicitly computed from the single-step probabilities q_1, P_F and P_D. It is also important to stress that this simplified formula is only valid when the approximation assumption holds. If the approximation is invalid then the more complex formula in Eqns. (22)–(23) must be used to compute the exact failure probability.

4. Model-based FDI systems

Section 3.3 described the calculation of the system reliability for a simple but abstract example. The purpose of this section is to provide additional details for a more concrete FDI system. Specifically, the FDI logic in Fig. 1 can be either model based or data driven. Section 4.1 describes the computation of FDI metrics for a specific model-based FDI logic. Section 4.2 then discusses further issues related to more general model-based FDI architectures. Data-driven methods can be evaluated within the general theory of Section 3. However, this requires the calculation of multiple-step false alarm and missed detection performance metrics for the data-driven FDI logic, and this issue is beyond the scope of this paper.

4.1. Residual-based FDI. The FDI logic monitoring the primary actuator is assumed to be a model-based algorithm. A typical model-based FDI scheme is comprised 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 actuator. There are many approaches to design the FDI filter, e.g., observers, parity equations, parameter estimators, and robust filters (Chen and Patton, 1999; Isermann, 2006; Ding, 2008). 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 the primary actuator is functioning properly ($\theta_1(k) = 0$) and non-zero mean when a fault occurs ($\theta_1(k) \neq 0$). 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 (Isermann, 2006; Ding, 2008).

This section considers the concrete FDI logic shown in Fig. 2. Suppose the actuator dynamics are perfectly known. An estimated control surface position can be computed based on the control input $u(k)$ and the actuator model Act_{model}. The real control surface position $s_1(k)$ is directly measured. The residual $r(k)$ is generated from the difference between the measured and estimated control surface positions. Assume any disturbances on the primary actuator are negligible. Moreover, the noise affecting the measurement $s_1(k)$ is modeled by an IID Gaussian process $n(k)$ with zero mean and variance σ^2. Finally, the fault on the primary actuator that occurs when $\theta_1(k) = 1$ is modeled by an additive bias f subject to $s_1(k)$. Given these assumptions, the FDI residual $r(k)$ is modeled as

$$r(k) = n(k) + \theta_1(k)f. \quad (28)$$

The decision logic uses a constant thresholding:

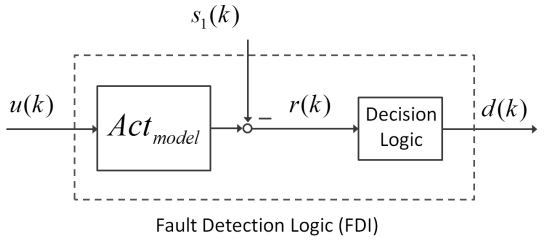


Fig. 2. Fault detection logic (FDI).

$$d(k) := \begin{cases} 1 & \text{if } |r(j)| > H \text{ for some } j \le k, \\ 0 & \text{otherwise.} \end{cases} \tag{29}$$

In other words, a fault is declared when the residual magnitude exceeds the threshold H. Note that this decision logic does not have intermittent switchings, i.e., $d(k)$ remains at 1 once the residual exceeds the threshold. This fault detection logic is IID in time and hence the system failure probability $P_{S,N}$ can be computed using the results in Section 3.3. 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

$$P_F = 1 - \int_{-H}^{H} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{r^2}{2\sigma^2}} \, \mathrm{d}r, \tag{30}$$

$$P_D = 1 - \int_{-H}^{H} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(r-f)^2}{2\sigma^2}} \, \mathrm{d}r. \tag{31}$$

These single-frame probabilities can be accurately and efficiently computed using the error function `erf` in `Matlab`. Then the FDI false alarm and missed detection metrics can be computed by Eqns. (4) and (5). The system failure probability $P_{S,N}$ involves these FDI performance metrics and the hardware component failure rates. $P_{S,N}$ can be computed by the sum of Eqns. (22)–(23) as described in Section 3.3.

4.2. Further issues. As mentioned above, the FDI logic can be designed in many different ways. For different designs the computation of FDI metrics can be more complicated than the case presented in Section 4.1. This section briefly discusses three additional issues in computing the probabilistic performance metrics for the FDI logics. In addition, guidelines are provided for using the proposed analysis framework with more complex FDI architectures.

One difficulty introduced by other model-based FDI design techniques are the time-correlations present in the residuals. The simple architecture considered in the previous subsection resulted in IID residuals. More general filtering architectures will make the residuals correlated in time. Monte Carlo simulations provide one approach for estimating FDI metrics in the presence of such time correlations. A straightforward application of Monte Carlo simulations can be time consuming because false alarms and missed detections occur infrequently. Various rare event simulation techniques, e.g., importance sampling and the splitting technique (Rubino and Tuffin, 2009), can be used to more efficiently compute FDI metrics. The implementation of rare event simulation techniques is problem dependent and requires experience. Alternatively, there are some existing theoretical tools that can be used to compute the multiple-frame FDI performance metrics for systems that have time-correlated residuals. The theoretical tools are more efficient but can be applied only under restricted assumptions. For example, the finite state Markov chain approximation can be used to efficiently compute the FDI metrics for the more specific case that the residual is governed by a first-order process (Brook and Evans, 1972; Lucas and Saccucci, 1990). For non-Gaussian residuals, it is possible to apply extreme value theory (Embrechts *et al.*, 1997) or a Poisson clumping heuristic (Aldous, 1989) to roughly estimate FDI metrics. However the accuracy of these approximations requires proper justification.

A second certification difficulty is that the performance of the model-based FDI is impacted by the accuracy of the models. Hence worst-case analysis is also required to determine the impact of model uncertainty and disturbances on the FDI performance. Worst-case analysis is important for understanding the trade off between robustness against model uncertainty and the good disturbance rejection. For example, the full-state observer-based FDI designed by Patton and Chen (1991) can completely reject the disturbance when the model is known perfectly. However, this sensitivity of the FDI performance to model uncertainty must be analyzed. The framework presented in Section 3.1 enables the worst case analysis of false alarm metrics and missed detection metrics to be separated from the hardware failure rates. Monte Carlo simulations can again be used to obtain sample-based lower bounds on the worst-case FDI metrics. Further research is needed to develop theoretical tools capable of obtaining (rigorous) upper bounds on the worst-case FDI metrics.

The third and final issue of importance is that all analysis tools should be validated using experimental data. Specifically, the available analysis tools (simulations and/or theoretical techniques) rely on simplifying assumptions to some degree. Any new analysis techniques should be compared against real data to ensure the validity of the results. In the aerospace domain, this would correspond to the use of flight data to obtain empirical estimates of FDI missed detection and false alarm probabilities. These empirical estimates should be compared against analysis results in order to gain

confidence in the analysis techniques. The main benefit of this approach is that flight tests are expensive and can be performed only on a limited set of flight conditions and trajectories. Validating the analysis results using the limited flight data would enable confident application of the analysis techniques to many flight conditions where there may not be flight data available. In summary, analysis tools should be justified based on real data so that they can finally meet the needs for practical applications.

5. Numerical example

This section provides a numerical example to demonstrate the proposed analysis method.

5.1. Problem setup. The dual-redundant system is assumed to run at a 100 Hz sample rate ($\Delta_t = 0.01$ sec) with primary and back-up actuators that both have a mean time between failure of MTBF $= 1000$ hours. Hence failure rates are approximated using discrete-time geometric distributions with $q_i = 2.78 \times 10^{-9}$ $(i = 1, 2)$. For simplicity, use the same notation for both actuators and set $q := q_i$ for the single-step failure rate and $\hat{q} := P[T_i \leq N] = 1 - (1 - q)^N \approx 10^{-3}$ for the N-step (per-hour) failure rate. The actuator system fails if it commands a bad surface position for at least $N_0 = 20$ consecutive samples. It is assumed that the actuation system uses the FDI logic described in Section 4.1. 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 100 Hz sample rate.

The reliability of single and triple-redundant actuation systems provides useful benchmarks. An actuation system based on a single actuator with MTBF $=$ 1000 hours has a per-hour failure probability of 10^{-3}. In other words, the reliability of this architecture is given simply by the reliability of the actuator itself. Alternatively, a triple-redundant actuation system could be used to improve the reliability. For example, the rudders on the Boeing 777 use a triple-redundant actuation system (Yeh, 1996; 2001). A triple-redundant architecture will fail if any two of the three actuators fail. In this case, the system failure probability per hour is $3(1 - \hat{q})\hat{q}^2 + \hat{q}^3 \approx 3 \times 10^{-6}$, where $\hat{q}$ is the per-hour failure probability of a single actuator with MTBF $= 1000$ hours. The system failure probability for the dual-redundant architecture with the analytical FDI will be compared with these two extreme cases.

5.2. Effects of thresholds and fault levels. The system failure probability $P_{S,N}$ can be computed from the results in Sections 3 and 4.1 for specific values of the residual variance σ^2, fault level f, and threshold H. The numerical procedure is briefly summarized. First, the single-frame false alarm and detection probabilities, P_F and P_D, are computed using Eqns. (30) and (31). Note that the single-frame FDI probabilities appear to depend independently on σ^2, f, and H. Equations (30) and (31) can be non-dimensionalized so that only the ratios H/σ and f/σ appear in the integrals. Thus the remainder of the analysis only considers the effect of H/σ and f/σ on $P_{S,N}$. Next, the exact probabilities for the basic failure events can be computed from Eqns. (22)–(23) using P_F, P_N, q, N_0, and N. There is no need to use the approximations (Eqn. (21)) as the exact equations can be efficiently evaluated. The exact system probability $P_{S,N}$ is then given by the sum of these basic failure event probabilities (Eqn. (8)). These steps are equivalent to evaluating the general result in Eqn. (14).

Figure 3 shows $P_{S,N}$ as a function of the normalized threshold H/σ for two values of the normalized fault level f/σ =1 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 actuator 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 actuator system for large thresholds.

For intermediate values of the threshold, the system failure probability depends on the ratio of the fault to the noise level. For large fault levels ($f/\sigma = 10$) 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 actuation system. This result can be explained as follows. Roughly one of the actuators (and its corresponding probability of failure) in the triplex system has been replaced by a perfect model in

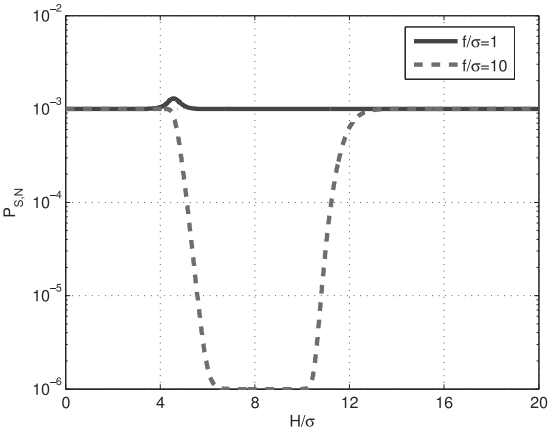


Fig. 3. $P_{S,N}$ vs. H/σ.

the dual-redundant system. In practice, the actuator model used by the analytical FDI logic will have some errors and this model uncertainty will degrade the performance of the analytical dual-redundant system.

For small fault sizes relative to the noise ($f/\sigma = 1$) the system has $P_{S,N} > 10^{-3}$ for some thresholds, i.e., the performance of the duplex system is even worse than that achieved by a single actuator. The results shown in Fig. 3 were generated with the exact formulas for $P_{S,N}$ but the approximations presented in Eqn. (21) provide some insight for the results $f/\sigma = 1$. Specifically, the approximation in Eqn. (27) can be expressed as

$$P_{S,N} \approx \hat{q}^2 + (1 - \hat{P}_D + \hat{P}_F)\hat{q}(1 - \hat{q}), \qquad (32)$$

where $\hat{P}_F = 1 - (1 - P_F)^N$ and $\hat{P}_D = 1 - (1 - P_D)^{N_0}$. 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 actuator if the N-step false alarm probability exceeds the detection probability. This analysis highlights an important distinction between the false alarm and detection probabilities. Specifically, when the per-frame probabilities P_F and P_D are both small, then the N-step probabilities are approximately $\hat{P}_F \approx NP_F$ and $\hat{P}_D \approx N_0 P_D$. The false alarm probability depends on the time scale of the entire analysis (N steps) while the detection probability only depends on the required detection time (N_0 steps). The typical case is $N \gg N_0$, and hence a very low per-frame false alarm rate ($P_F \ll P_D$) is required to ensure a good overall system reliability.

5.3. Optimized thresholds. The results in Fig. 3 indicate the importance of proper threshold selection. As an example, Fig. 3 shows that for $f/\sigma = 10$ the optimal threshold is $H^* = 8.4$ and this yields the optimal performance of $P^*_{S,N} \approx 10^{-6}$ for this fault level. More generally, let $H^*(f/\sigma)$ denote the threshold that minimizes $P_{S,N}$ for a given fault level f/σ. Figure 4 shows the optimal performance $P^*_{S,N}$ and threshold H^* as a function of the fault level. As expected, the optimal performance $P^*_{S,N}$ decreases monotonically with an increasing fault level. Figure 4 also shows the limits of performance for the specific model-based FDI scheme described in Section 4.1. In particular, for small fault levels ($f/\sigma \leq 2$) the failure probability of the duplex system is similar to that of a single actuator system even if the optimal threshold is chosen. This implies that more advanced filter techniques and decision functions are required if the fault level is small relative to the noise.

5.4. Relation to ROC curves. A receiver operating characteristic (ROC) curve (Egan, 1975) is one tool for selecting the threshold in detection systems. The ROC curve graphically illustrates the FDI performance as the threshold H varies. The solid line in Fig. 5 is

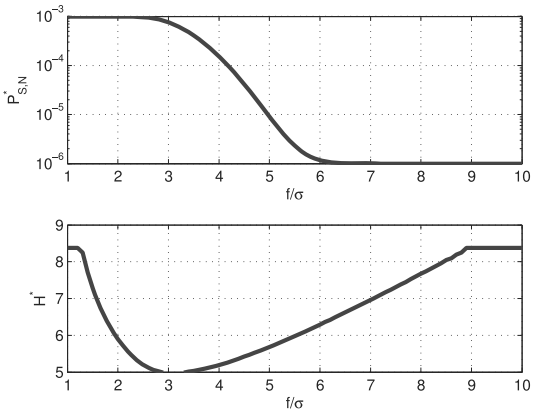


Fig. 4. Optimal performance $P^*_{S,N}$ and threshold H^* vs. f/σ.

a standard single-frame ROC curve for the fault level $f/\sigma = 3.5$. This solid line shows the single-frame detection probability P_D vs. the single-frame false alarm probability P_F for a range of thresholds H. The desired performance is to have low false alarms ($P_F \approx 0$) and high detection rates ($P_D \approx 1$). This corresponds to the upper left corner on the plot.

A key point of the analysis in this paper is that the total system probability $P_{S,N}$ depends on multiple-step false alarm and detection performance. For comparison, Fig. 5 also shows the multiple frame ROC curve for the fault level $f/\sigma = 3.5$. Specifically, the dashed line is a plot of the N-step detection probability $\hat{P}_D$ vs. the N-step false alarm probability $\hat{P}_F$ for a range of thresholds H. The qualitative shape of the single-frame and multiple-frame curves is similar but the underlying dependence on the threshold (not shown) is significantly different. Specifically, the optimal threshold for $f/\sigma = 3.5$ is given by $H^* \approx 5.0$, yielding an optimal system failure probability of $P^*_{S,N} \approx 4.1 \times 10^{-4}$. This optimal threshold corresponds to the following single and N-step performance: $P_F = 4.9 \times 10^{-7}$, $P_D = 0.06$, $\hat{P}_F = 0.16$, and $\hat{P}_D = 0.73$. These performance metrics are shown by the squared locations in Fig. 5 on the respective single and N-step ROC curves. As discussed in the previous subsection, false alarms must be very small in order to obtain a good system reliability. These results emphasize this point as the optimal performance in this example is obtained with a very small single-step false alarm probability.

6. Conclusions and future work

This paper analyzed the reliability of a dual-redundant actuator system with an analytical fault detection scheme. The system failure probability per hour can be exactly computed provided that certain probabilistic information is known for actuator failures and fault detection performance. A numerical example with a

simple model-based fault detection logic was given to demonstrate the approach. The proposed technique can be combined with high-fidelity Monte Carlo simulations to assess system reliability. Future work will consider a more realistic example, e.g., the duplex actuation system on the Airbus A380. This will require the analysis framework to incorporate intermittent faults as well as advanced fault detection filters and decision functions. The future work will also evaluate the reliability estimates calculated with the proposed approach using empirical data (flight tests and bench tests) obtained from a small UAV. This will provide additional confidence in the practical utility of the proposed method.

Acknowledgment

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

ADDSAFE (2012). *ADDSAFE: Advanced Fault Diagnosis for Sustainable Flight Guidance and Control*, European 7th Framework Program, `http://addsafe.deimos-space.com/`.

Aldous, D. (1989). *Probability Approximations via the Poisson Clumping Heuristic*, Springer-Verlag, New York, NY.

Asmussen, S.R. and Glynn, P.W. (2007). *Stochastic Simulation: Algorithms and Analysis*, Springer, New York, NY.

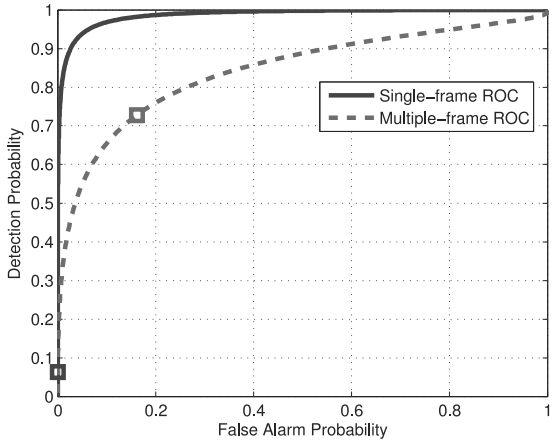


Fig. 5. ROC curves P_D vs. P_F and $\hat{P}_D$ vs. $\hat{P}_F$ for $f/\sigma = 3.5$.

Belcastro, C. and Belcastro, C. (2003). On the validation of safety critical aircraft systems, Part I: An overview of analytical and simulation method, *Proceedings of the AIAA Conference of Guidance, Navigation and Control, GNC 2003, Austin, TX, USA*, paper no. AIAA 2003–5559.

Bleeg, R. (1988). Commercial jet transport fly-by-wire architecture considerations, *AIAA/IEEE Digital Avionics Systems Conference, San Jose, CA, USA*, pp. 399–406.

Brook, D. and Evans, D.A. (1972). An approach to the probability distribution of CUSUM run length, *Biometrika* **59**(3): 539–549.

Chen, J. and Patton, R. (1999). *Robust Model-Based Fault Diagnosis for Dynamic Systems*, Kluwer, Boston, MA.

Collinson, R. (2011). *Introduction to Avionic Systems, 3rd Edition*, Springer, New York, NY.

Ding, S. (2008). *Model-Based Fault Diagnosis Techniques: Design Schemes, Algorithms, and Tools*, Springer-Verlag, Berlin.

Efimov, D., Cieslak, J., Zolghadri, A. and Henry, D. (2013). Actuator fault detection in aircraft systems: Oscillatory failure case study, *Annual Reviews in Control* **37**(1): 180–190.

Egan, J. (1975). *Signal Detection Theory and ROC Analysis*, Academic Press, New York, NY.

Embrechts, P., Kluppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*, Springer, New York, NY.

Fawcett, T. (2006). An introduction to ROC analysis, *Pattern Recognition Letters* **27**(8): 861–874.

Freeman, P., Pandita, R., Srivastava, N. and Balas, G. (2013). Model-based and data-driven fault detection performance for a small UAV, *IEEE Transactions on Mechatronics* **18**(4): 1300–1309.

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

Goupil, P. (2011). AIRBUS state of the art and practices on FDI and FTC in flight control system, *Control Engineering Practice* **19**(6): 524–539.

Gustafsson, F., Åslund, J., Frisk, E., Krysander, M. and Nielsen, L. (2008). On threshold optimization in fault-tolerant systems, *Proceedings of the IFAC World Congress, Seoul, Korea,* pp. 7883–7888.

Heller, M., Niewoehner, R. and Lawson, P.K. (2001). F/A-18E/F super hornet high-angle-of-attack control law development and testing, *Journal of Aircraft* **38**(5): 841–847.

Hu, B. and Seiler, P. (2013). A probabilistic method for certification of analytically redundant systems, *Proceedings of the 2nd International Conference of Control and Fault-Tolerant Systems, SysTol 2013, Nice, France*, pp. 13–18.

Isermann, R. (2006). *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*, Springer-Verlag, Berlin.

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

Krasich, M. (2000). Use of fault tree analysis for evaluation of system-reliability improvements in design phase, *Proceedings of the IEEE Annual Reliability and Maintainability Symposium, RAMS 2000, Los Angeles, CA, USA*, pp. 1–7.

Lee, W., Grosh, D., Tillman, A. and Lie, C. (1985). Fault tree analysis, methods, and applications: A review, *IEEE Transactions on Reliability* **34**(3): 194–203.

Lucas, J.M. and Saccucci, M.S. (1990). Exponentially weighted moving average control schemes: Properties and enhancements, *Technometrics* **32**(1): pp. 1–12.

Murthy, D., Xie, M. and Jiang, R. (2004). *Weibull Models*, John Wiley & Sons, Hoboken, NJ.

Nakagawa, T. and Osaki, S. (1975). The discrete Weibull distribution, *IEEE Transactions on Reliability* **24**(5): 300–301.

Patton, R.J. and Chen, J. (1991). Robust fault detection using eigenstructure assignment: A tutorial consideration and some new results, *Proceedings of the IEEE Conference on Decision and Control, CDC 1991, Brighton, UK*, pp. 2242–2247.

Åslund, J., Biteus, J., Frisk, E., Krysander, M. and Nielsen, L. (2007). Safety analysis of autonomous systems by extended fault tree analysis, *International Journal of Adaptive Control and Signal Processing* **21**(2–3): 287–298.

Rausand, M. and Hoyland, A. (2004). *System Reliability Theory: Models, Statistical Methods, and Applications*, Wiley-Interscience, Hoboken, NJ.

Renfrow, J., Liebler, S. and Denham, J. (1994). F-14 flight control law design, verification, and validation using computer aided engineering tools, *Proceedings of the IEEE Conference on Control Applications, CCA 1994, Glasgow, UK,* pp. 359–364.

Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods*, Springer, New York, NY.

Rubino, G. and Tuffin, B. (2009). *Rare Event Simulation Using Monte Carlo Methods*, Wiley, New York, NY.

Singpurwalla, N.D. (2006). *Reliability and Risk: A Bayesian Perspective*, John Wiley & Sons, Chichester.

Stein, W. and Dattero, R. (1984). A new discrete Weibull distribution, *IEEE Transactions on Reliability* **33**(2): 196–197.

United States Congress (2012). House resolution 658: FAA modernization and reform act of 2012, Section 332: Integration of civil unmanned aircraft systems into national airspace system.

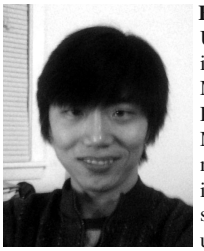
Vanek, B., Bauer, P., Gozse, I., Lukatsi, M., Reti, I. and Bokor, J. (2014). Safety critical platform for mini UAS insertion into the common airspace, *Proceedings of the AIAA Guidance, Navigation and Control Conference, GNC 2014, National Harbor, MD, USA*, AIAA–2014–0977.

Wheeler, T.J., Seiler, P., Packard, A.K. and Balas, G.J. (2011). Performance analysis of fault detection systems based on analytically redundant linear time-invariant dynamics, *Proceedings of the American Control Conference, ACC 2011, San Francisco, CA, USA*, pp. 214–219.

Willsky, A.S. and Jones, H.L. (1976). A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems, *IEEE Transactions on Automatic Control* **21**(1): 108–112.

Yeh, Y. (1996). Triple-triple redundant 777 primary flight computer, *Proceedings of the 1996 IEEE Aerospace Applications Conference, Aspen, CO, USA*, pp. 293–307.

Yeh, Y. (2001). Safety critical avionics for the 777 primary flight controls system, *Proceedings of the 20th Digital Avionics Systems Conference, DASC 2001, Daytona Beach, FL, USA*, pp. 1.C.2.1–1.C.2.11.



Bin Hu received his Bachelor's degree from the University of Science and Technology of China in 2008 and a Master's degree from Carnegie Mellon University in 2010. Currently he is a Ph.D candidate at the Aerospace Engineering and Mechanics Department of the University of Minnesota, twin cities.His main research interests include reliability theory, fault tolerant control, stochastic hitting time problems, rare event simulations, and their applications in the design and certification processes of safety-critical systems.

Peter Seiler received his Ph.D. from the University of California, Berkeley, in 2001. His graduate research focused on coordinated control of unmanned aerial vehicles and control over wireless networks. In 2004–2008, Dr. Seiler worked at the Honeywell Research Labs on various aerospace and automotive applications, including the redundancy management system for the Boeing 787, sensor fusion algorithms for automotive active safety systems and re-entry flight control laws for NASA's Orion vehicle. Since joining the University of Minnesota in 2008, Dr. Seiler has been working on fault-detection methods for safety-critical systems and advanced control techniques for wind turbines.

Appendix

Correlated actuator analysis

This appendix summarizes the key results when the failure times for the primary and backup actuators (T_1 and T_2) are correlated. The duplex system failure rate $P_{S,N}$ can be computed for the correlated case given the joint probability mass function $P[T_1 = j, T_2 = k]$ for all $1 \leq j, k \leq N + 1$. The FDI logic only monitors the primary actuator and hence the FDI performance does not depend on T_2. Therefore, a result similar to Eqn. (14) can

be obtained following the framework in Section 3.1:

$$\begin{aligned} P_{S,N} &= \sum_{k=1}^{N} P[T_S \geq k + N_0 \mid T_1 = k] P[T_1 = k] \\ &\quad + P[T_S \leq N \mid T_1 = N+1] \\ &\quad \times P[T_1 = N+1, T_2 \leq N] \\ &\quad + \sum_{k=1}^{N} P[T_S < k + N_0 \mid T_1 = k] P[T_1 = k, T_2 \leq N]. \end{aligned} \tag{A1}$$

Equation (16) can also be extended as

$$\begin{aligned} P_{S,N} &= P[T_1 \leq N, T_2 \leq N] \\ &\quad + P[T_S \leq N \mid T_1 = N+1] P[T_1 = N+1, T_2 \leq N] \\ &\quad + \sum_{k=1}^{N} P[T_S \geq k + N_0 \mid T_1 = k] \\ &\quad \times P[T_1 = k, T_2 = N+1]. \end{aligned} \tag{A2}$$

The discussion in Section 3.1 provides similar insights regarding the basic failure modes for this correlated case. The main difficulty in applying these results for correlated failures is that it would be difficult to determine the joint probability mass function for the actuator failure times.

Received: 30 January 2014
Revised: 17 April 2014