Optimal Discrimination between Transient and Permanent Faults

Mathematical Details

CSR Tecnical report, v 2.1, September 1998

M. Pizza[a], L. Strigini[a] , A. Bondavalli[b] and F. Di Giandomenico[c]

a) Centre For Software Reliability, City University, Northampton Square, London EC1V OHB, UK. E-mail {mpizza, strigini}@csr.city.ac.uk

b) CNUCE/CNR, via S.Maria 36, 56126 Pisa, Italy.
E-mail a.bondavalli@cnuce.cnr.it

c) IEI/CNR, via S.Maria 46, 56126 Pisa, Italy.
E-mail digiandomenico@iei.pi.cnr.it

This report contains additional material concerning sections 3 and section 4 of the paper "Optimal Discrimination between Transient and Permanent Faults' (Proc. 3rd IEEE High-Assurance System Engineering Symposium, November 13-14, 1998, Washington, DC; IEEE Computer Society Press).

This report together with the above paper supersedes the CSR Technical Report "Bayesian Diagnosis of Transient vs Permanent Faults", January 1998.

Note: In this report, there are references to equations (1) and (2), figures 1-8, tables 1 and 2 and reference [9] which are in the above paper.

Contents

Addenda to section 3. Behaviour of the proposed method 1
3.1 Effect of sequences of successes 2
3.2 Effect of sequences of failures 5
3.3 Useful descriptive measures 7
3.3.1 Reaction to observing a first failure 7
3.3.2 How frequently tests must fail to produce a high posterior probability of permanent faults 9
Addenda to section 4.Application to different scenarios 11
Appendix 12

Addenda to section 3. Behaviour of the proposed method

A designer would not need to forecast the detailed responses of the diagnosis algorithm, as they are 'normatively correct'. Sill we may want to understand the general pattern of response and how it is affected by the parameters of the failure model. Here, we study how the posterior probability of the conjecture "the component is permanently faulty" changes over time when the test results are a sequence of either s consecutive successes or f consecutive failures. Assume that this sequence starts after the i-th test round, at which time $P_{post}(Perm(i))=p0$. We define these two functions:

$P_{post}(s, p0) = P_{post}(Perm(i+s) \mid P_{post}(Perm(i) = p0 \ , success(j) \text{ for } j=i+1,...i+s)$

$P_{post}(f, p0) = P_{post}(Perm(i+f) \mid P_{post}(Perm(i) = p0 \ , failure(j) \text{ for } j=i+1,...i+f)$

The expressions of the two probabilities above

$P_{post}(f, p0)$ and $P_{post}(s, p0)$ are obtained by repetedly applying equation (2), starting with a $P_{post}(Perm(i))=p0$. The following equations can be proved true by induction:

$P_{post}(f, p0)=$

$$\Big[p0\cdot\big[P(failure(i)\mid Pperm(i-1))-P(failure(i)\wedge\neg Perm(i)\mid\neg Perm(i-1))\big]+$$
$$(1-p0)\cdot\theta p\cdot P(failure(i)\mid Perm(i))\cdot(1-Q^{f})\Big]\Big/$$
$$\Big[(1-p0)\cdot\big[P(failure(i)\mid Pperm(i-1))-P(failure(i)\wedge\neg Perm(i)\mid\neg Perm(i-1))\big]\cdot(Q)^{f}+$$
$$p0\cdot\big[P(failure(i)\mid Pperm(i-1))-P(failure(i)\wedge\neg Perm(i)\mid\neg Perm(i-1))\big]+$$
$$(1-p0)\cdot\theta p\cdot P(failure(i)\mid Perm(i))\cdot(1-Q^{f})\Big]$$

where $Q=\left(\dfrac{P(failure(i)\wedge\neg Perm(i)\mid\neg Perm(i-1))}{P(failure(i)\mid Perm(i-1))}\right)$.

$P_{post}(s, p0) =$

$$\Big[p0\cdot\big[P(success(i)\wedge\neg Perm(i)\mid\neg Perm(i-1))-P(success(i)\mid Perm(i-1))\big]+$$
$$(1-p0)\cdot\theta p\cdot(P(success(i)\mid Perm(i-1)))\cdot\left(Q^{s}-1\right)\Big]\Big/$$
$$\Big[(1-p0)\cdot\big[P(success(i)\wedge\neg Perm(i)\mid\neg Perm(i-1))-P(success(i)\mid Perm(i-1))\big]\cdot Q^{s}+$$
$$p0\cdot\big[P(success(i)\wedge\neg Perm(i)\mid\neg Perm(i-1))-P(success(i)\mid Perm(i-1))\big]+$$
$$+(1-p0)\cdot\theta p\cdot(P(success(i)\mid Perm(i-1)))\cdot\left(Q^{s}-1\right)\Big]$$

where $Q=\left(\dfrac{P(success(i)\wedge\neg Perm(i)\mid\neg Perm(i-1))}{P(success(i)\mid Perm(i-1))}\right)$

NOTE: In the following subsections, all the results indicated as "approximate" have been proven to have negligible error under the (sufficient but not necessary) condition that the

probability of observing failures becomes at least ten times higher after a component becomes permanently faulty:

$$P(\,failure(i) \wedge \neg\, Perm(i) | \neg Perm(i-1)) < 0.1 \cdot P(\,failure(i) | Perm(i))$$

or, in other words, that transient faults do make a sizeable difference in the probability that the component fails a test. Details are in [9].

3.1 Effect of sequences of successes

As tests accumulate in which only successes are observed, the probability of permanent fault asymptotically tends to a limiting value *L*:

$$L = \lim_{s \to +\infty} P_{post}\left(Perm(i+s) \mid P_{post}(Perm(i)) = p0,\ success(j) \quad for\ all\ j \in \{i+1,..i+s\}\right)$$
$$= \lim_{s \to +\infty} Ppost(s,\ p0)$$

Usually, $L > 0$: there is a non-zero probability of the component being permanently faulty, even though it has never failed a test (yet). The expression for $P_{post}(s, p0)$ is given in the box at the beginning of this section. L takes different values depending on whether:

$$P(success(i) \wedge \neg Perm(i) | \neg Perm(i-1)) > P(success(i) | Perm(i-1)) \qquad (3)$$

$$L = \begin{cases} 1 & if \quad \dfrac{P(success(i) \wedge \neg Perm(i) \mid \neg Perm(i-1))}{P(success(i) \mid Perm(i-1))} < 1 \\ p0 & if \quad \dfrac{P(success(i) \wedge \neg Perm(i) \mid \neg Perm(i-1))}{P(success(i) \mid Perm(i-1))} = 1 \\ \dfrac{\theta p}{\dfrac{P(success(i) \mid \neg Perm(i-1))}{P(success(i) \mid Perm(i-1))} - 1} & if \quad \dfrac{P(success(i) \wedge \neg Perm(i) \mid \neg Perm(i-1))}{P(success(i) \mid Perm(i-1))} > 1 \end{cases}$$

If, as is usually the case, (3) is verified, one obtains:

$$L = \frac{\theta p}{\dfrac{P(success(i) \mid \neg Perm(i-1))}{P(success(i) \mid Perm(i-1))} - 1} \qquad (4)$$

As can be seen from (4), if $P(success(i)|\neg Perm(i-1)) > 2*P(success(i)\ |Perm(i-1))$, L is small ($L < \theta p$).

If $p0 < L$, $P_{post}(s, p0)$ quickly tends to L as s increases. If $p0 > L$, $P_{post}(s, p0)$ decreases approaching L (Fig. 9). This decrease can be described in terms of a parameter c, defined as:

$$c = \log\left(\frac{P(success(i) | Perm(i-1))}{P(success(i) \wedge \neg Perm(i) | \neg Perm(i-1))}\right)$$

The number *s* of successes needed to reach a certain value of $P_{post}(s, p0)$ is approximately proportional to *1/c*, so long as $P_{post}(s, p0) >> L$. This behaviour can be observed in Fig. 9.

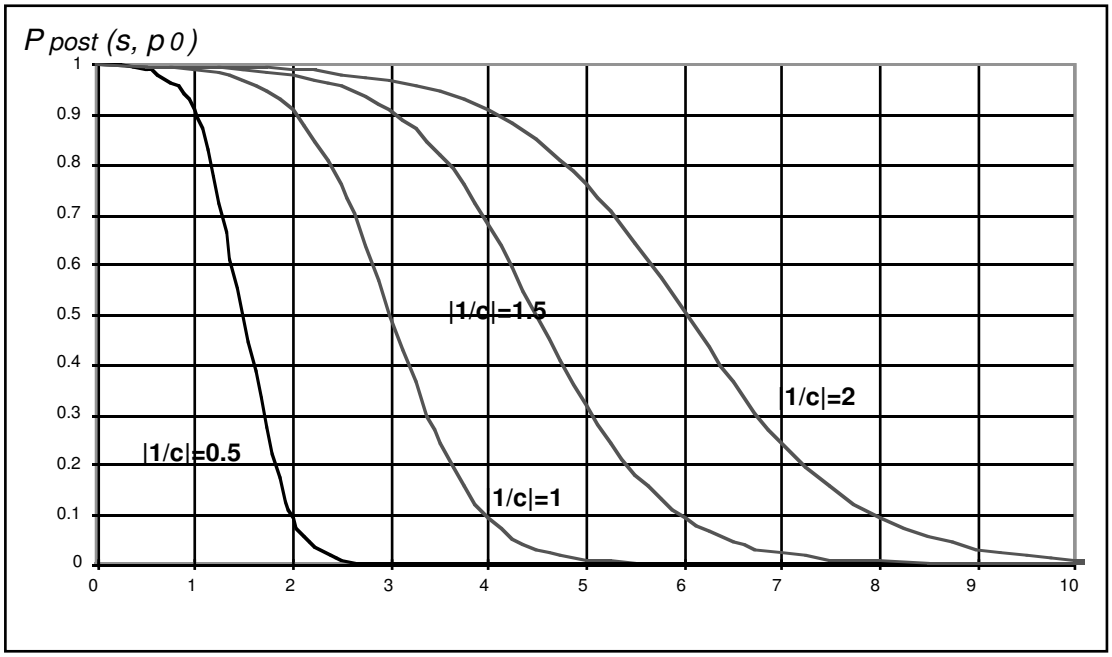

Figure 9. $P_{post}(s, p0)$ as a function of the number of consecutive successes, *s*, starting from $p0=0.999$. For $s \to \infty$, each curve tends to its own asymptote *L*, very close to *0*.

Useful equations (used for most of the following mathematical derivations)

$P(failure(i) \wedge \neg Perm(i) | \neg Perm(i-1)) = P(failure(i) | \neg Perm(i)) - \theta p * P(failure(i) | Perm(i))$

$P(success(i) \wedge \neg Perm(i) | \neg Perm(i-1)) = P(success(i) | \neg Perm(i)) - \theta p * P(success(i) | Perm(i))$

$P(failure(i) | Perm(i-1)) = P(failure(i) | Perm(i))$

$P(success(i) | Perm(i-1)) = P(success(i) | Perm(i))$

$P(success(i) | Perm(i-1)) = 1 - P(failure(i) | Perm(i-1))$

$log\ (x*y) = log(x) + log(y)$

$log(1/x) = -log(x)$

Why Ppost(s,p0) is approximately proportional to 1/c so long as $P_{post}(s, p0) >> L$

The function $S_{p0}(p)$ gives the number of consecutive successes that one would need to observe for the probability $P_{post}(Perm)$ to decrease from *p0* to *p*. Here we are only considering the case in which *p* and *p0* are greater than *L* ($p0 > p > L$). The opposite case ($p0 < p < L$) has been studied as well but we will not report it as it is not of much interest. An expression for $S_{p0}(p)$ can be obtained by solving for *s* the equation $P_{post}(s, p0) = p$.

$$S_{p0}(p)=\frac{1}{\log\left(\left(\frac{P(success(i)|\,Perm(i-1))}{P(success(i)\wedge\neg Perm(i)|\,\neg Perm(i-1))}\right)\right)}\cdot$$

$$\left(\begin{array}{l}\log\frac{1-p0}{1-p}+\\ +\log\frac{\left(p\cdot\left(1-\frac{P(success(i)\wedge\neg Perm(i)|\,\neg Perm(i-1))}{P(success(i)|\,Perm(i-1))}-\theta p\right)+\theta p\right)}{\left(p0\cdot\left(1-\frac{P(success(i)\wedge\neg Perm(i)|\,\neg Perm(i-1))}{P(success(i)|\,Perm(i-1))}-\theta p\right)+\theta p\right)}\end{array}\right)$$

Using the definition of c we can write

$$S_{p0}(p)=\frac{1}{c}\cdot\left(\begin{array}{l}\log\frac{1-p0}{1-p}+\\ +\log\frac{\left(p\cdot\left(\frac{P(success(i)|\,Perm(i-1))-P(success(i)|\,\neg Perm(i-1))}{P(success(i)|\,Perm(i-1))}\right)+\theta p\right)}{\left(p0\cdot\left(\frac{P(success(i)|\,Perm(i-1))-P(success(i)|\,\neg Perm(i-1))}{P(success(i)|\,Perm(i-1))}\right)+\theta p\right)}\end{array}\right).$$

Since $\left(\frac{P(success(i)|\,Perm(i-1))-P(success(i)|\,\neg Perm(i-1))}{P(success(i)|\,Perm(i-1))}\right)=\theta p\cdot\frac{1}{L}$, this last expression can be rewritten as:

$$S_{p0}(p)=\frac{1}{c}\cdot\left(\log\frac{1-p0}{1-p}+\log\frac{\left(p\cdot\frac{1}{L}+1\right)}{\left(p0\cdot\frac{1}{L}+1\right)}\right),$$

that is:

$$S_{p0}(p)=\frac{1}{c}\cdot\left(\log\frac{1}{1-p}+\log\left(p\cdot\frac{1}{L}+1\right)\right)+\frac{1}{c}\cdot\left(\log\frac{1-p0}{\left(p0\cdot\frac{1}{L}+1\right)}\right),$$

One can see that, when $p>>L$, $\log\left(p\cdot\frac{1}{L}+1\right)\approx\log\left(p\cdot\frac{1}{L}\right)$, thus $S_{p0}(p)$ is well approximated by:

$$S_{p0}(p)\approx\frac{1}{c}\cdot\left(\log\frac{1}{1-p}+\log\left(p\cdot\frac{1}{L}\right)\right)+\frac{1}{c}\cdot\left(\log\frac{1-p0}{\left(p0\cdot\frac{1}{L}+1\right)}\right),$$

Now *log(x*y)=log (x) + log (y)* and therefore*:*

$$S_{p0}(p)\approx\frac{1}{c}\cdot\left(\log\frac{p}{1-p}+\log\left(\frac{1}{L}\right)\right)+\frac{1}{c}\cdot\left(\log\frac{1-p0}{\left(p0\cdot\frac{1}{L}+1\right)}\right)$$

$$= \frac{1}{c} \cdot \left(\log \frac{p}{1-p} \right) + \frac{1}{c} \cdot \left(\log \frac{1-p0}{\left(p0 \cdot \frac{1}{L} + 1 \right)} + \log\left(\cdot \frac{1}{L} \right) \right)$$

$$= \frac{1}{c} \cdot \left(\log \frac{p}{1-p} \right) + C(p0)$$

C(p0) is a constant value (i.e. it does not depend on *p*), such that $S_{p0}(p0)=0$.

The absolute error produced in using this approximate expression for $S_{p0}(p)$ is $\frac{1}{c} \cdot \left(-\log\left(1 - \frac{L}{p} \right) \right)$, which is negligible when $p >> L$.

Conclusion: if $p >> L$ ($p0 > p >> L$), the number of successes s required to reduce *Ppost(po,s)* from *p0* to *p* is approximately proportional to c

3.2 Effect of sequences of failures

Ppost(f, p0) obviously increases with *f* and tends to 1 for $f \to \infty$. We will use the notation $F_{p0}(p)$ to denote the number of consecutive failures that need to be observed for *Ppost(f, p0)* to reach a value $p > p0$. Figure 10 shows two curves of *Ppost(f, p0)* as a function of *f*, for $p0=0$. The curves can be described in terms of two parameters:

$$a = \log\left(P(Perm(i) \mid \neg Perm(i-1)) \right),$$

(related to the rate of occurrence of permanent faults) and

$$b = \log\left(\frac{P(failure(i) \wedge \neg Perm(i) \mid \neg Perm(i-1))}{P(failure(i) \mid Perm(i-1))} \right)$$

(describing how much a permanent fault increases the probability of failure))

To give an idea of the speed of increase of *Ppost(f, p0)*, we can say that $F_0(0.5) \approx a/b$ (with negligible error bounded by $\left| \frac{\log(0.9 + \theta p)}{\log \theta p} \right|$); for any pair $\hat{p}, p0$ with $1 > \hat{p} > p0 > 0$, the number of failures, *f*, needed to make *Ppost(f, p0)* = $\hat{p}$ is approximately proportional to $\frac{1}{|b|}$:

$$F_{p0}(\hat{p}) \propto \frac{1}{|b|}$$ (this is an approximation with negligible error if $p0 >> \theta p$)

The shape of each curve is determined by the system parameters, not by *p0*. If $p0 \neq 0$, it is sufficient to shift the origin of the *x* axis so that the chosen curve intersects the point *(0,p0)*.

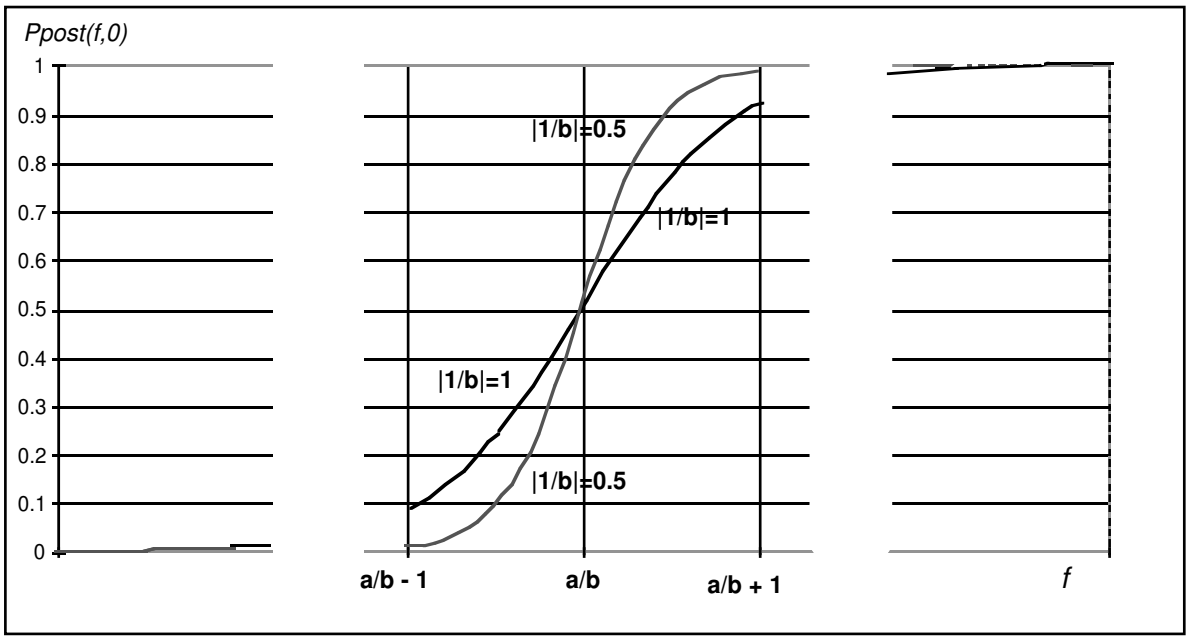

Figure 10. Increase of $P_{post}(f, p0)$ as a function of the number of consecutive failures, f, starting from the value $p0=0$. For $f \to \infty$, all curves tend to 1.

How we derived the results concerning a and b:

The function $F_{p0}(p)$ gives the number of consecutive failures to be observed for the probability $P_{post}(Perm)$ to increase from $p0$ to p. We will only consider $F_0(p)$, considering that for any $p0 \neq 0$ one can write $F_{p0}(p)=F_0(p)-F_0(p0)$. An expression for $F_0(p)$ can be obtained by solving for f the equation $P_{post}(f, 0) = p$.

$$F_0(p) = \frac{1}{\log\left(\frac{P(failure(i) \wedge \neg Perm(i) | \neg Perm(i-1))}{P(failure(i) | Perm(i-1))}\right)} \cdot \left(-\log\frac{1}{1-p} - \log\frac{\left(p \cdot \left(1 - \frac{P(failure(i) \wedge \neg Perm(i) | \neg Perm(i-1))}{P(failure(i) | Perm(i-1))} - \theta p\right) + \theta p\right)}{\theta p} \right)$$

We can approximate the previous equation, with negligible error when $p >> \theta p$ (the error has been proved to be negligible under the sufficient condition $\frac{P(failure(i) \wedge \neg Perm(i) | \neg Perm(i-1))}{P(failure(i) | Perm(i-1))} < 0.1$), by

$$F_0(p) \approx \frac{1}{\log\left(\frac{P(failure(i) \wedge \neg Perm(i) | \neg Perm(i-1))}{P(failure(i) | Perm(i-1))}\right)} \cdot \left(-\log\frac{1}{1-p} - \log\frac{p}{\theta p} \right)$$

Denoting $\log\left(\frac{P(failure(i) \wedge \neg Perm(i) | \neg Perm(i-1))}{P(failure(i) | Perm(i-1))}\right)$ by b and $log(\theta p)$ by a we will rewrite the previous expression as ($log(a/b)=log(a)-log(b)$):

$$F_0(p) \approx \frac{1}{\log\left(\frac{P(failure(i) \wedge \neg Perm(i) \mid \neg Perm(i-1))}{P(failure(i) \mid Perm(i-1))}\right)} \cdot \left(-\log\frac{1}{1-p} - \log\frac{p}{\theta p}\right)$$

$$= \frac{1}{b} \cdot \left(-\log\frac{1}{1-p} - \log\frac{p}{\theta p}\right) =$$

$$= \frac{a}{b} - \frac{1}{b} \cdot \left(-\log\frac{p}{1-p}\right)$$

One can see that the approximate expression of $F_0(p)$ is proportional to b and that for p=0.5 we have $F_0(0.5) \approx a/b$.

If $p > p0 >> \theta p$:

$$F_{p0}(p) = F_0(p) - F_0(p0) \approx$$

$$\approx \frac{a}{b} - \frac{1}{b} \cdot \left(-\log\frac{p}{1-p}\right) - \frac{a}{b} + \frac{1}{b} \cdot \left(-\log\frac{p0}{1-p0}\right) =$$

$$= \frac{1}{b} \cdot \left(-\log\frac{p \cdot (1-p0)}{(1-p) \cdot p0}\right)$$

As one can see, $F_{p0}(\hat{p}) \propto \frac{1}{|b|}$

3.3 Useful descriptive measures

For given values of the parameters, i.e., characteristics of the component and of the testing procedure, the diagnosis algorithm needs simply to apply equation (2) after each test round. This produces the correct posterior probability of permanent fault, as a function of the whole past history of observed successes and failures. Every possible history will produce a different evolution of $P_{post}(Perm)$. To understand the influence of the parameters, some less detailed description is desirable: we now define two measures that describe important macroscopic aspects of how $P_{post}(Perm)$ changes with test results, and will be useful in the rest of the paper.

3.3.1 Reaction to observing a first failure

In the operation of a system, long series of successes will usually separate any failures or clusters of failures. So, a common case is that a failure is observed at a test round i when $P_{prior}(Perm(i)) \approx L$. A useful descriptive measure is thus $F_L(0.5)$, the number of consecutive failures which have to be observed for $P_{post}(Perm)$ to increase from L to 0.5. This value is approximated by:

$$F_L(0.5) \approx \frac{a}{b} - \frac{\log\left(P(failure(i) \mid Perm(i))\right)}{b} \approx \frac{1}{b} \cdot \log(\theta p + L)$$

Why $F_L(0.5) \approx \frac{a}{b} - \frac{\log\left(P(failure(i) \mid Perm(i))\right)}{b} \approx \frac{1}{b} \cdot \log(\theta p + L)$

If $\frac{P(success(i) \wedge \neg Perm(i) \mid \neg Perm(i-1))}{P(success(i) \mid Perm(i-1))} > 1$, which is usually the case, we have that:

$$L=\frac{\theta p}{\frac{P(success(i)\mid\neg Perm(i-1))}{P(success(i)\mid Perm(i-1))}-1}=\frac{\theta p\cdot P(success(i)\mid Perm(i-1))}{P(success(i)\mid\neg Perm(i-1))-P(success(i)\mid Perm(i-1))}=$$

$$F_0(L)=\frac{1}{\log\left(\frac{P(failure(i)\wedge\neg Perm(i)\mid\neg Perm(i-1))}{P(failure(i)\mid Perm(i-1))}\right)}\cdot$$

$$\left(-\log\frac{1}{1-L}-\log\frac{\left(L\cdot\left(\frac{P(failure(i)\mid Perm(i-1))-P(failure(i)\wedge\neg Perm(i)\mid\neg Perm(i-1))-\theta p\cdot P(failure(i)\mid Perm(i-1))}{P(failure(i)\mid Perm(i-1))}\right)+\theta p\right)}{\theta p}\right)$$

Seeing that $P(failure(i)\wedge\neg Perm(i)\mid\neg Perm(i-1))=P(failure(i)\mid\neg Perm(i))-\theta p*P(failure(i)\mid Perm(i-1))$:

$$\frac{P(failure(i)\mid Perm(i-1))-P(failure(i)\mid\neg Perm(i-1))}{P(failure(i)\mid Perm(i-1))}=\frac{P(success(i)\mid\neg Perm(i-1))-P(success(i)\mid Perm(i-1))}{P(failure(i)\mid Perm(i-1))}=$$

$$=\frac{P(success(i)\mid\neg Perm(i-1))-P(success(i)\mid Perm(i-1))}{P(success(i)\mid Perm(i-1))\cdot\theta p}\cdot\frac{P(success(i)\mid Perm(i-1))\cdot\theta p}{P(failure(i)\mid Perm(i-1))}=$$

$$=\frac{1}{L}\cdot\frac{P(success(i)\mid Perm(i-1))\cdot\theta p}{P(failure(i)\mid Perm(i-1))}$$

We can rewrite:

$$F_0(L)=\frac{1}{b}\cdot\left(-\log\frac{1}{1-L}-\log\frac{\left(L\cdot\left(\frac{\theta p\cdot P(success(i)\mid Perm(i-1))}{L\cdot P(failure(i)\mid Perm(i-1))}\right)+\theta p\right)}{\theta p}\right)$$

$$=\frac{1}{b}\cdot\left(-\log\frac{1}{1-L}-\log\frac{1}{P(failure(i)\mid Perm(i-1))}\right)$$

$$=\frac{1}{b}\cdot\left(-\log\left[(1-L)\cdot P(failure(i)\mid Perm(i-1))\right]\right)$$

$$\approx\frac{1}{b}\cdot\left(-\log\left[P(failure(i)\mid Perm(i-1))\right]\right)$$

Seeing that $F_L(0.5)=F_0(0.5)-F_0(L)$ and that $F_0(0.5)\approx a/b$ we have the result:

$$F_L(0.5)\approx\frac{a}{b}-\frac{\log\left(P(failure(i)\mid Perm(i))\right)}{b}$$

We can also write

$F_L(0.5)=F_0(0.5)-F_0(L)=$

$$= \frac{1}{b} \cdot \left[\log 0.5 - \log \frac{\left(0.5 \cdot \left(1 - \frac{P(failure(i) \wedge \neg Perm(i) | \neg Perm(i-1))}{P(failure(i) | Perm(i-1))} - \theta p \right) + \theta p \right)}{\theta p} \right.$$

$$\left. - \log \frac{1}{1-L} + \log \frac{\left(L \cdot \left(1 - \frac{P(failure(i) \wedge \neg Perm(i) | \neg Perm(i-1))}{P(failure(i) | Perm(i-1))} - \theta p \right) + \theta p \right)}{\theta p} - \right]$$

$$= \frac{1}{b} \cdot \left[\log \frac{0.5}{\left(0.5 \cdot \left(1 - \frac{P(failure(i) \wedge \neg Perm(i) | \neg Perm(i-1))}{P(failure(i) | Perm(i-1))} - \theta p \right) + \theta p \right)} \right.$$

$$\left. - \log \frac{1}{1-L} + \log \left[L \cdot \left(1 - \frac{P(failure(i) \wedge \neg Perm(i) | \neg Perm(i-1))}{P(failure(i) | Perm(i-1))} - \theta p \right) + \theta p \right] \right]$$

Under the assumption that $P(failure(i) \wedge \neg Perm(i) | \neg Perm(i)) < 0.1 \cdot P(failure(i) | Perm(i))$, the first logarithm from the left is approximately zero. Since $\log \frac{1}{1-L}$ is also close to zero, we can write

$$F_L(0.5) = F_0(0.5) - F_0(L) \approx \frac{1}{b} \cdot \log \left[L \cdot \left(1 - \frac{P(failure(i) \wedge \neg Perm(i) | \neg Perm(i-1))}{P(failure(i) | Perm(i-1))} - \theta p \right) + \theta p \right] \right]$$

$$\frac{1}{b} \cdot \log(L + \theta p).$$

$F_L(0.5)$ is a good indicator of how rapidely the probability of permanent fault grows when one first starts observing failures. Of course, it is usually a non-integer. Since test rounds happen in integer numbers, the diagnosis can only be updated after an integer number of observations, and any decision based on it will be somewhat insensitive to small errors in parameter values.

3.3.2 How frequently tests must fail to produce a high posterior probability of permanent faults

It is also interesting to consider the case in which successes and failures alternate (not necessarily in a regular pattern), causing $P_{post}(Perm)$ to oscillate. We would then like to know by how much failures must outnumber successes for our diagnosis to veer decidedly towards "permanently faulty".

It turns out that, as long as $P_{post}(Perm) >> L$, the effects of one failure will be completely cancelled by a series of b/c successes. These need not be in an uninterrupted sequence, provided every *additional* failure is followed by b/c additional successes. With a higher ratio of successes to failures, P_{post}(Perm) decreases towards L; with a lower ratio, it increases towards 1. We shall call the value b/c the *success/failure ratio threshold* , or T_{sfr}.

Why $T_{sfr}=b/c$

$F_{p1}(p2)$ gives the number of failures that will increase $P_{post}(perm)$ from $p1$ to $p2$ while $S_{p2}(p1)$ gives the number of successes that will return to the probability $p1$ from $p2$. The ratio $\frac{S_{p2}(p1)}{F_{p1}(p2)}$ gives the average number of successes that will cancel the effects of one failure. From the previous studies it has been said that if $p1,p2>>max\ (L,\theta p)$, $F_{p1}(p2)$ is proportional to $1/b$, while $S_{p1}(p2)$ is proportional to $1/c$. Thus, if $p1,p2>>max\ (L,\theta p)$, we can expect $\frac{S_{p2}(p1)}{F_{p1}(p2)}$ not to depend on $p1$ and $p2$ but to be approximately constant:

$$\frac{S_{p2}(p1)}{F_{p1}(p2)} \approx \frac{b}{c}.$$

We will now show it analytically.

$F_{p1}(p2)$ and $S_{p1}(p2)$ are well approximated (when $p1,p2>>max(L,\theta p)$) by :

$$S_{p2}(p1) \approx \frac{1}{c}\cdot\left(\log\frac{1-p2}{1-p1}+\log\frac{\left(p1\cdot\left(\frac{P(success(i)|\,Perm(i-1))-P(success(i)|\,\neg Perm(i-1))}{P(success(i)|\,Perm(i-1))}\right)\right)}{\left(p2\cdot\left(\frac{P(success(i)|\,Perm(i-1))-P(success(i)|\,\neg Perm(i-1))}{P(success(i)|\,Perm(i-1))}\right)\right)}\right)$$

$$=\frac{1}{c}\cdot\left(\log\frac{1-p2}{1-p1}+\log\frac{p1}{p2}\right)$$

and

$$F_{p1}(p2) \approx \frac{1}{b}\cdot\left(-\log\frac{1-p1}{1-p2}-\log\frac{p1}{p2}\right)$$

Under the previous conditions on $p1$ and $p2$, the ratio $\frac{S_{p2}(p1)}{F_{p1}(p2)}$ is approximately constant value and independent of values of $p1$ and $p2$:

$$\frac{S_{p2}(p1)}{F_{p1}(p2)} \approx \frac{\frac{1}{c}\cdot\left(\log\frac{1-p2}{1-p1}+\log\frac{p1}{p2}\right)}{\frac{1}{b}\cdot\left(-\log\frac{1-p1}{1-p2}-\log\frac{p1}{p2}\right)}=\frac{b}{c}$$

Table 3 summarises the parameters and descriptive measures defined in this section.

Intermediate parameters	Descriptive measures
$a=\log\left(P(Perm(i)\,|\,\neg Perm(i-1))\right)$ $b=\log\left(\frac{P(failure(i)\wedge\neg Perm(i)\,|\,\neg Perm(i-1))}{P(failure(i)\,|\,Perm(i-1))}\right)$ $c=\log\left(\frac{P(success(i)\,|\,Perm(i-1))}{P(success(i)\wedge\neg Perm(i)\,|\,\neg Perm(i-1))}\right)$	$L=\frac{\theta p}{\frac{P(success(i)\,|\,\neg Perm(i-1))}{P(success(i)\,|\,Perm(i-1))}-1}$ $F_L(0.5)\approx\frac{a}{b}-\frac{\log\left(P(failure(i)\,|\,Perm(i))\right)}{b}$ $T_{sfr}=\frac{b}{c}$

Table 3. Expressions for the intermediate parameters and the descriptive measures defined

Addenda to section 4.Application to different scenarios

Table 4 lists the hypotheses characterising the three scenarios. Taking account of the additional model parameters we define here, the expressions defined in Section 3 become rather complex; they can be found in the Appendix. Here, we specify the assumptions for the three scenarios and describe the general behaviour of the diagnosis algorithm (details are in [9]):

Scenario A: All faults in a component cause it to fail tests. So, a test success implies certainty that the component is non-faulty (as can be confirmed by applying equation (2) using success as the value for *evidence(i)*). Clearly, $L = 0$. The only non-obvious problem in diagnosis is discriminating between transient and permanent faults, after test failures are observed.

Scenario B: Even when faulty, a component may still produce correct test data, with probabilities α_p if there are permanent faults and α_t in case of transient faults only. The parameters α_p and α_t describe both the fault manifestation characteristics of the component (e.g., rate of intermittent manifestation of permanent faults) and the coverage of the tests. The adjudication is 'perfect' (error-free).

The implications of sequences of successes are less obvious than in Scenario A. Observing a success does not warrant certainty that the component is non-faulty: in particular, the limit L is greater than 0. The expressions for *Ppost(f,p0)* and *Ppost(s,p0)* in this scenario are given in the appendix in equations (A2) and (A3), respectively. If $\alpha_t = \alpha_p$, *Ppost(f,p0)* is the same as for Scenario A. Indeed, a test failure means that there is a fault (since we assume perfect adjudication): all that the diagnosis has to do is to assign probabilities of this being permanent vs transient, based on the ratio of their respective probabilities of causing a failure. With $\alpha_t=\alpha_p$, this ratio is the same as in Scenario A. For a given pair $(s,p0)$, and if $P(success(i) \wedge \neg Perm(i) | \neg Perm(i-1))$ is substantially greater than $P(success(i) | Perm(i-1))$, the right-hand term in (A3) depends almost only on α_p, irrespective of the values taken by the other parameters: intuitively, the main reason for doubting that a test success implies absence of permanent faults is the possibility that the latter may not become manifest in testing.

Scenario C: The same as scenario B, but the adjudication process may misinterpret the data collected. This behaviour is modelled by 2 parameters: with probability $\beta 1$ erroneous data are interpreted as a success, and with probability $\beta 2$ correct data are interpreted as a failure. These parameters model both the possibility of physical faults affecting the adjudication, and of the adjudication being imperfect by design (e.g., software-implemented “reasonableness tests” on program results).

Equations (A4) and (A5) describe *Ppost(f,p0)* and *Ppost(s,p0)* for this scenario. If both $\beta 1$ and $\beta 2$ are smaller than θt and θp, (which is plausible if adjudication errors are due to physical faults only), these functions will take very similar values to those in Scenario B.

	i-th Observation phase		i-th Adjudication phase	
	state of component during observation	**correctness of the data collected**	**correctness of the data collected**	**test results**
Scen.A	*Ok(i)*	*correct(i)*	*correct(i)*	*success(i)*
	Perm(i)	*erroneous(i)*	*erroneous(i)*	*failure(i)*
	Temp(i)	*erroneous(i)*		
Scen.B	*Ok(i)*	*correct(i)*		
	Perm(i)	*erroneous(i)* with prob *1-*α_p *correct(i)* with prob α_p	*correct(i)*	*success(i)*
	Temp(i)	erroneous(i) with prob *1-*α_t *correct(i)* with prob α_t	*erroneous(i)*	*failure(i)*
Scen.C	*Ok(i)*	*correct(i)*	*correct(i)*	*success(i)* with prob *1-*$\beta 2$ *failure(i)* with prob $\beta 2$
	Perm(i)	*erroneous(i)* with prob *1-*α_p *correct(i)* with prob α_p	*erroneous(i)*	*failure(i)* with prob *1-*$\beta 1$ *success(i)* with prob $\beta 1$
	Temp(i)	*erroneous(i)* with prob *1-*α_t *correct(i)* with prob α_t		

Table 4. Hypotheses for the three scenarios

Appendix

Here we list all the formulas needed for applying this diagnosis method, and the expressions of the intermediate parameters and descriptive measures used in the paper, for the three scenarios considered. More details are in [9].

Scenario A ($\alpha_t=\alpha_p=\beta 1=\beta 2=0$)**:**

$$P_{post}(f,0)=\frac{\theta p\cdot(1-\theta t^{f})}{\theta p\cdot(1-\theta t^{f})+\theta t^{f}\cdot(1-\theta t)} \tag{A1}$$

Scenario B ($\beta 1=\beta 2=0$)**:**

$$P_{post}(f,p0)=\frac{p0\cdot(1-\theta t\cdot(1-\alpha_t)-\alpha_p)+(1-p0)\cdot\theta p\cdot(1-\alpha_p)\cdot\left(1-\left(\frac{\theta t\cdot(1-\alpha_t)}{1-\alpha_p}\right)^{f}\right)}{(1-p0)(1-\theta t\cdot(1-\alpha_t)-\alpha_p)\cdot\left(\frac{\theta t\cdot(1-\alpha_t)}{1-\alpha_p}\right)^{f}+} \tag{A2}$$

$$+p0\cdot(1-\theta t\cdot(1-\alpha_t)-\alpha_p)+(1-p0)\cdot\theta p\cdot(1-\alpha_p)\cdot\left(1-\left(\frac{\theta t\cdot(1-\alpha_t)}{1-\alpha_p}\right)^{f}\right)$$

$$P_{post}(s,p0)=\frac{p0\cdot(1-\theta p-\theta t\cdot(1-\alpha_t)-\alpha_p)+(1-p0)\cdot\theta p\cdot\alpha_p\cdot\left(\left(\frac{1-\theta p-\theta t\cdot(1-\alpha_t)}{\alpha_p}\right)^s-1\right)}{(1-p0)(1-\theta p-\theta t\cdot(1-\alpha_t)-\alpha_p)\cdot\left(\frac{1-\theta p-\theta t\cdot(1-\alpha_t)}{\alpha_p}\right)^s+}$$

$$\frac{}{p0\cdot(1-\theta p-\theta t\cdot(1-\alpha_t)-\alpha_p)+(1-p0)\cdot\theta p\cdot\alpha_p\cdot\left(\left(\frac{1-\theta p-\theta t\cdot(1-\alpha_t)}{\alpha_p}\right)^s-1\right)} \quad \text{(A3)}$$

Scenario C:

$$P_{post}(f,p0)=\frac{p0\cdot\left((1-\alpha_p)\cdot(1-\beta1)+\alpha_p\cdot\beta2-\theta t\cdot(1-\alpha_t)\cdot(1-\beta1)-(1-\theta p-\theta t\cdot(1-\alpha_t))\cdot\beta2\right)+}{(1-p0)\left((1-\alpha_p)\cdot(1-\beta1)+\alpha_p\cdot\beta2-\theta t\cdot(1-\alpha_t)\cdot(1-\beta1)-(1-\theta p-\theta t\cdot(1-\alpha_t))\cdot\beta2\right)\cdot Q^f+}$$

$$\frac{(1-p0)\cdot\theta p\cdot\left((1-\alpha_p)\cdot(1-\beta1)+\alpha_p\cdot\beta2\right)\cdot\left(1-Q^f\right)}{p0\cdot\left((1-\alpha_p)\cdot(1-\beta1)+\alpha_p\cdot\beta2-\theta t\cdot(1-\alpha_t)\cdot(1-\beta1)-(1-\theta p-\theta t\cdot(1-\alpha_t))\cdot\beta2\right)}$$

$$\frac{}{(1-p0)\cdot\theta p\cdot\left((1-\alpha_p)\cdot(1-\beta1)+\alpha_p\cdot\beta2\right)\cdot\left(1-Q^f\right)} \quad \text{(A4)}$$

where $Q=\frac{\left(\theta t\cdot(1-\alpha_t)\cdot(1-\beta1)+(1-\theta p-\theta t\cdot(1-\alpha_t))\cdot\beta2\right)}{(1-\alpha_p)\cdot(1-\beta1)+\alpha_p\cdot\beta2}$

$$P_{post}(s,p0)=\frac{p0\cdot\left((1-\theta p-\theta t\cdot(1-\alpha_t))\cdot(1-\beta2)+\theta t\cdot(1-\alpha_t)\cdot\beta1-\beta1\cdot(1-\alpha_p)-(1-\beta2)\cdot\alpha_p\right)+}{(1-p0)\cdot\left((1-\theta p-\theta t\cdot(1-\alpha_t))\cdot(1-\beta2)+\theta t\cdot(1-\alpha_t)\cdot\beta1-\beta1\cdot(1-\alpha_p)-(1-\beta2)\cdot\alpha_p\right)\cdot Q^s+}$$

$$\frac{(1-p0)\cdot\theta p\cdot\left(\beta1\cdot(1-\alpha_p)+(1-\beta2)\cdot\alpha_p\right)\cdot\left(Q^s-1\right)}{p0\cdot\left((1-\theta p-\theta t\cdot(1-\alpha_t))\cdot(1-\beta2)+\theta t\cdot(1-\alpha_t)\cdot\beta1-\beta1\cdot(1-\alpha_p)-(1-\beta2)\cdot\alpha_p\right)+}$$

$$\frac{}{(1-p0)\cdot\theta p\cdot\left(\beta1\cdot(1-\alpha_p)+(1-\beta2)\cdot\alpha_p\right)\cdot\left(Q^s-1\right)} \quad \text{(A5)}$$

where $Q=\frac{(1-\theta p-\theta t\cdot(1-\alpha_t))\cdot(1-\beta2)+\theta t\cdot(1-\alpha_t)\cdot\beta1}{\beta1\cdot(1-\alpha_p)+(1-\beta2)\cdot\alpha_p}$.

Scenario A	$a = \log \theta p\,, \quad b = \log \theta t\,, \quad L = 0, \quad F_L(0.5) \approx a / b$
Scenario B	$a = \log \theta p\,, \quad b = \log \frac{\theta t \cdot (1-\alpha_t)}{1-\alpha_p}, \quad c = \log\left(\frac{\alpha_p}{1-\theta t \cdot (1-\alpha_t) - \theta p}\right) \approx \log \alpha_p$ $L = \frac{\frac{\theta p}{1-\theta p \cdot (1-\alpha_p) - \theta t \cdot (1-\alpha_t)}}{\alpha_p} - 1$ $F_L(0.5) \approx \frac{a}{b} - \frac{\log\left(1-\alpha_p\right)}{b} \approx \frac{1}{b} \cdot \log(\theta p + L), \qquad T_{sfr} = \frac{b}{c}$
Scenario C	$a = \log \theta p\ , \quad b = \frac{\theta t \cdot (1-\alpha_t) \cdot (1-\beta 1) + (1-\theta p - \theta t \cdot (1-\alpha_t)) \cdot \beta 2}{(1-\alpha_p) \cdot (1-\beta 1) + \alpha_p \cdot \beta 2}$ $c = \frac{\beta 1 \cdot (1-\alpha_p) + (1-\beta 2) \cdot \alpha_p}{(1-\theta p - \theta t \cdot (1-\alpha_t)) \cdot (1-\beta 2) + \theta t \cdot (1-\alpha_t) \cdot \beta 1}$ $L = \frac{\theta p}{\frac{(1-\theta p - \theta t) \cdot (1-\beta 2) + \theta t \cdot \left((1-\alpha_t) \cdot \beta 1 + \alpha_t \cdot (1-\beta 2)\right) + \theta p \cdot \left(\beta 1 \cdot (1-\alpha_p) + (1-\beta 2) \cdot \alpha_p\right)}{\beta 1 \cdot (1-\alpha_p) + (1-\beta 2) \cdot \alpha_p} - 1}$ $F_L(0.5) \approx \frac{a}{b} - \frac{\log\left((1-\alpha_p) \cdot (1-\beta 1) + \beta 2 \cdot \alpha_p\right)}{b} \approx \frac{1}{b} \cdot \log(\theta p + L), \qquad T_{sfr} = \frac{b}{c}$

Table 5. Expressions for the intermediate parameters and descriptive measures