Updating failure rates and test intervals in the operational phase: A practical implementation of IEC 61511 and IEC 61508

S. Hauge
SINTEF Technology and Society

M. A. Lundteigen & M. Rausand
NTNU Production and Quality Engineering

Demonstrating compliance with reliability targets in the operational phase is a key requirement in the follow-up of safety instrumented systems. This paper describes a new approach for demonstrating such compliance, taking into account the information from reported failures. This includes to update the failure rate and to make decisions regarding the functional test interval. A brief discussion of some practical experience with data collection and analysis in the operational phase is also included. The main focus is on safety instrumented systems in the oil and gas industry, but the approach may be applicable for other industry sectors as well.

1 INTRODUCTION

Safety instrumented systems (SIS) are frequently used in the oil and gas industry to detect the onsets of hazardous events (e.g., gas leakages and high pressures) and to mitigate their consequences to humans, the environment, and material assets. Failure to do so can lead to major accidents and it is therefore of vital importance to monitor the performance of the SIS in the operational phase.

Each SIS performs one or more safety instrumented functions (SIF), by using some electrical, electronic, or programmable electronic technology. Design and operation must follow the requirements in IEC 61508 (1998) and IEC 61511 (2003), two standards that have been widely adopted by the national authorities for the oil and gas industry. IEC 61508 is a generic standard on SIS design and construction, while IEC 61511 addresses SIS applications in the process industries.

The reliability performance of the SIS is referred to as *safety integrity* in the IEC-standards. This is the probability that the SIS satisfactorily performs the required SIFs, under stated conditions and for a stated period of time (IEC 61511 2003). In an early design phase, a hazards and risk analysis is performed to establish the required level of safety integrity of each SIF that is performed by the SIS, and in design and operation, it is necessary to demonstrate that this level of safety integrity is met.

The IEC-standards distinguish between four *safety integrity levels* (SIL), where SIL 1 is the least reliable level and SIL 4 is the most reliable level. Some requirements apply to all SILs, while others are SIL-specific. Once the required SIL has been established for each SIF, it is necessary for the plant owner to ensure that each SIF is designed, installed, and operated according to the requirements. This includes to demonstrate that a set of qualitative and quantitative requirements are met in each of these phases.

The objective of this paper is to present a new approach for demonstrating compliance with the quantitative requirements in the operational phase based on operational experience.

The new approach is applicable for *low demand* SIFs, which are the most common SIFs in the oil and gas industry. A low demand SIF is passive during normal operation and is intended to respond if a specific hazardous event occurs. In passive mode, limited information is available about the system state. Regular functional testing is therefore important to reveal failures that otherwise would be hidden until the next demand. The approach presented in this paper suggests criteria for how to adjust the functional test intervals, based on operational experience.

The structure of the paper is as follows: In Section 2, some of the factors that may influence the predicted and the actual safety integrity are discussed. Section 3 gives a brief overview of some of the challenges related to failure rate estimation, including an overview of relevant literature. The new approach, comprising

four steps, is presented in Section 4. Some case examples are also included here. In Section 5, the authors share some experience from some industry projects involving analysis of SIS related failures. Section 6 gives some conclusions and final remarks.

2 PREDICTED AND ACTUAL SAFETY INTEGRITY

The quantitative requirements for a low demand SIF are stated by means of the average probability of failure on demand, PFD. For each SIL, the PFD must be within a specified interval. For a SIL 3 function, for example, the PFD must be $\leq 10^{-3}$, which means that no more than one failure per 1 000 demands is accepted. The required SILs are usually specified by the operator, and the vendor or the system integrator has to demonstrate the the SIFs meet these SIL requirements.

Predicted PFD

This initial PFD estimate for a SIF that is made by the vendor or system integrator is here referred to as the *predicted* PFD, and denoted by PFD_0. To be accepted, for example, as SIL 3, the vendor or system integrator must demonstrate that $\text{PFD}_0 \leq 10^{-3}$.

When the SIS is supplied to the operator, it is understood that the PFD_0 is in accordance with the required SIL.

The PFD_0 is calculated based on (i) a system reliability model that is compatible with the system topology and the assumptions made, and (ii) a set of reliability parameters.

The PFD_0 is influenced by a number of assumptions about the future operating and environmental conditions, the quality of maintenance, test coverage and efficiency of testing, and deterioration mechanisms of the SIS components (Lundteigen and Rausand 2007; Lundteigen 2009).

The available reliability models (e.g., reliability block diagrams, fault trees, and Markov models) are able to capture many different system features, but no model can capture all features.

Reliability data may be obtained from manufacturers, the operator's own experience, generic data sources (e.g., OREDA (2002)), from expert judgements, and so on. In most cases, the data only reflects *random hardware failures*, that are caused by normal degradation, while *systematic failures* are not taken into account. Systematic failures are due to inadequate design, operation, maintenance, or exposure outside the design envelope, and do not have the same statistical properties as random hardware failures.

The technology is developing fast, and experience from earlier equipment may not adequately reflect the reliability of the equipment to be installed. Another issue is that several manufacturers tend to be over-optimistic when supplying reliability data for their equipment (SINTEF 2006; Summers 2008). The predicted PFD_0 is therefore subject to a significant uncertainty.

Actual PFD

When the SIS is put into operation, the operator must try to ascertain that the *actual* PFD meets the SIL requirements. How to do this, is not obvious, since we are not able to *measure* the actual PFD. The actual PFD can fail to meet the SIL requirement, and be greater than PFD_0, because of the following reliability influencing factors:

- The system has been modified (either the system or the SIS)
- The actual operational and environmental conditions are not compatible with those that were assumed in the design phase
- The testing and maintenance are not performed as assumed in the design phase
- The common cause failures are not as assumed
- The deterioration mechanisms are not as assumed
- The actual reliability of the equipment is not as indicated by the generic data sources
- The expert judgements were inaccurate
- Some failures have not been revealed

The actual PFD may change with time due to system changes and changes of the operational and environmental conditions. The actual PFD should therefore be assessed at several stages in the operational phase. The actual PFD in stage i is denoted PFD_i, for $i = 1, 2, \ldots$.

The approach that is outlined in this paper has two objectives: (i) to assess whether or not the actual PFD meets the SIL requirement, i.e., if $\text{PFD}_i \leq \text{PFD}_0$, and (ii) if PFD_i is significantly different from PFD_0, what impact should this have on the length of the functional test interval. The test interval is a decision variable, while the $\lambda_{\text{DU},i}$ is an unknown parameter that must be estimated based on experience data.

The PFD is a function of the rate λ_{DU} of dangerous undetected (DU) failures of the components of the SIS, the length, τ of the functional test interval, and of several other parameters. A dangerous (D) failure is a failure that prevents the execution of a SIF, and a failure is undetected (U) if it is hidden until there is a real demand or a functional test. Some dangerous failures may also be detected by online diagnostics and are referred to as dangerous detected (DD). If the DD failures are repaired within a short time, we may consider the effect from DD failures on the PFD to be negligible.

As shown by Rausand and Høyland (2004), the PFD is mainly an increasing function of the product

$\lambda_{\text{DU}} \cdot \tau$. This means that, if we can keep all other conditions constant, then

$$\text{PFD}_i \leq \text{PFD}_0 \Leftrightarrow \lambda_{\text{DU,i}} \cdot \tau_i \leq \lambda_{\text{DU,0}} \cdot \tau_0 \qquad (1)$$

where τ_i is the actual functional test interval, $\lambda_{\text{DU},i}$ is the actual DU failure rate, τ_0 is the functional test interval that was suggested to verify that the PFD_0 meets the SIL requirement with $\lambda_{\text{DU,0}}$, the initially assumed failure rate.

3 CHALLENGES RELATED TO FAILURE RATES

To estimate the rate of DU-failures is a challenging task. First, DU-failures are rare events since the SIS is designed for high reliability. Second, there may be only a few components of the same type for observation. Whereas an installation may have more than 100 fire detectors, the number of certain valves, for example high integrity pressure protection system (HIPPS) valves, may be limited to one or two.

The stated challenges are well addressed in literature. A general overview is given by Rausand and Høyland (2004). Several authors address parameter estimation with few observations, see e.g., Hokstad et al. (1998), Kunttu and Kortelainen (2004), and Chhibber and Apostolakis (1993). It is also frequently discussed how generic data may be adjusted to application specific environments and operational conditions, see for example an overview in Moss (2005). Røed and Aven (2009) and Røed et al. (2009) discuss an approach for taking other influencing factors into account, for example that an operator makes a wrong installation. An unsolved research challenge is to find practical ways to implement the available theory, and specifically within the scope of IEC 61508 and IEC 61511. Some initial initiatives have been taken by Vatn (2006) and Hauge and Lundteigen (2008). The presented approach builds on these initiatives.

4 NEW APPROACH

The new approach is split into four steps, as illustrated in Fig. 1. The first two steps focus on safety integrity monitoring, where the number of recorded DU-failures is used to indicate how the actual PFD_i is compared to the predicted PFD_0. In step 3, the number of recorded DU-failures is used to estimate a new (actual) failure rate. This failure rate may be used to determine PFD_i, with or without considering the other reliability influencing factors, such as the common cause failures. In this approach, the new failure rate is used in step 4 to evaluate if the current functional test interval is appropriate.

We now consider the first stage ($i = 1$). This is the first scrutiny of the actual performance of the SIF in the operational phase. We assume that changes in the

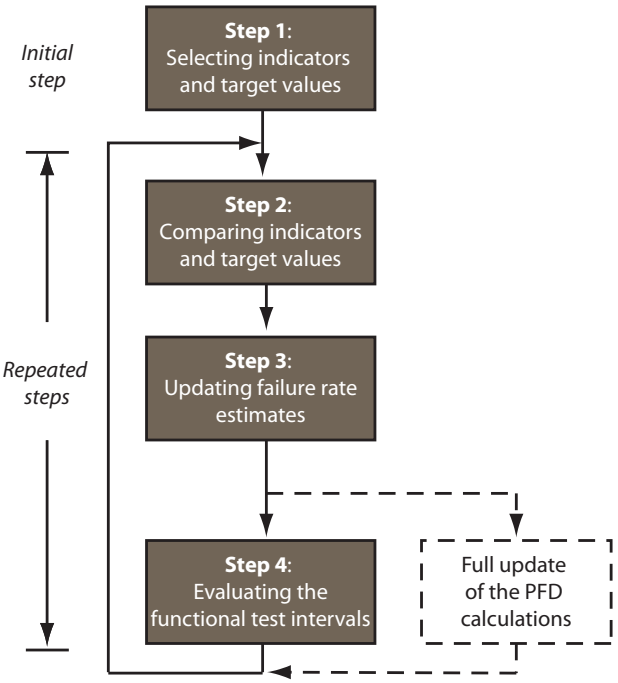

Figure 1: Main steps of the new approach

environmental and operational conditions mainly affect the failure rate and that the fraction of failures that are common cause are as assumed for PFD_0. In addition, it is assumed that all DU-failures are revealed by regular functional tests, and that the SIS has not been modified. We also assume that the test interval is as assumed in the design phase (The last assumption can, however, easily be relaxed).

Step 1: Selecting indicators and target values

In this step, we introduce two quantities that are used in step 2 to compare the actual PFD_1 with the predicted PFD_0:

Integrity performance indicator: An observable quantity for a defined population (e.g., an equipment group) that can be used to estimate the actual PFD_1. In this approach, we suggest using the recorded number of DU-failures as the integrity performance indicator.

Integrity target value: The maximum value that the integrity performance indicator may take before the actual PFD_1 (most likely) exceeds the predicted PFD_0. Here, we suggest using the expected number of DU-failures with the initially assumed failure rate ($\lambda_{\text{DU,0}}$) as the integrity target value.

Let t_1 denote the aggregated time in service from start-up until the first assessment of the actual PFD. We assume that the number of DU-failures, $X(t_1)$ during the service time t_1, can be modeled by a homogenous Poisson process with rate $\lambda_{\text{DU,1}}$. Then, the expected number of failures are (Rausand and Høyland 2004):

$$\text{E}(X(t_1)) = \lambda_{\text{DU,1}} \cdot t_1 \qquad (2)$$

We tacitly assume that all observation periods start just after a functional test and that they are terminated

just after another test. Any other observation periods will include time in service where we are prohibited from detecting DU-failures.

The integrity target value is the expected number of DU-failures that would occur if the rate of DU-failures were as assumed in the previous phase (here, in the design phase), i.e.,

$$\mathrm{E}_0(X(t_1)) = \lambda_{\mathrm{DU},0} \cdot t_1 \tag{3}$$

and we denote this number by x_0.

Step 2: Comparing integrity performance with integrity target values

Let x_1 denote the observed number of DU-failures during the aggregated time in service t_1. Once the values are established, the following rules are suggested for the comparison:

1. If $x_1 \approx x_0$, the situation is considered acceptable, but the possibility of removing the failure cause should anyhow be considered (ALARP principle).
2. If $x_1 < x_0$, the situation is acceptable – and less frequent proof testing may in some cases be considered (see step 4). The ALARP principle still applies.
3. If $x_1 > x_0$, a failure cause analysis should be performed and compensating measures to reduce the number of future failures should be considered, including the need for more frequent proof testing (see step 4).

The most important case is case 3, where more failures than expected have occurred. This indicates that the rate of DU-failures that was assumed in design may be too optimistic, and that the SIL requirements may not be fulfilled.

Case 2 indicates that the rate of DU-failures that was assumed in design was too pessimistic, and that we may extend the test interval and still fulfill the SIL requirements.

Step 3: Updating failure rate estimates

The initial rate of DU-failures can be updated based on operational experience in two different ways: (i) by estimating $\lambda_{\mathrm{DU},1}$ solely based on observed failures, or (ii) by combining operational experience with the initial estimate by a Bayesian updating procedure.

When the actual failure rate is based on observations alone, it is denoted $\hat{\lambda}_{\mathrm{DU},1}$.

If limited operational experience is available, either due to short observation period or a small number of components under observation, it is better to perform a Bayesian update of the failure rate. With Bayesian updating, we combine prior knowledge about the initial failure rate, i.e., $\lambda_{\mathrm{DU},0}$, with operational experience, i.e., the number of DU-failures. In this case, we denote the new (updated) failure rate $\tilde{\lambda}_{\mathrm{DU},1}$.

Our prior knowledge about $\lambda_{\mathrm{DU},0}$ is usually based on data provided by the manufacturer or from generic databases like OREDA (2002).

Alternative I: Using operational experience alone:

When the criteria for having sufficient operational experience are met, it is straightforward to estimate a new failure rate. Having assumed that failures occur according to a homogeneous Poisson process, the natural estimator of $\lambda_{\mathrm{DU},1}$ is (Rausand and Høyland 2004):

$$\hat{\lambda}_{\mathrm{DU},1} = \frac{\text{Number of failures}}{\text{Aggregated time in service}} = \frac{x_1}{t_1} \tag{4}$$

where x_1 is the observed number of DU-failures and t_1 is the aggregated time in service. The estimate is unbiased, and a 90% confidence interval $(\lambda_{\mathrm{DU,1,L}}, \lambda_{\mathrm{DU,1,U}})$ is given by

$$\lambda_{\mathrm{DU,1,L}} = \frac{1}{2t_1} z_{0.05,2x_1} \tag{5}$$

$$\lambda_{\mathrm{DU,1,U}} = \frac{1}{2t_1} z_{0.95,2(x_1+1)} \tag{6}$$

where $z_{\varepsilon,\nu}$ denotes the upper 100ε% percentile of the χ^2 distribution with ν degrees of freedom.

How much operational experience we need to rely on operational experience alone, is not easy to know. Where a large number of similar components are installed at the same installation, like fire and gas detectors, there may be a significant amount of experience data available already after a year of operation. With few components of the same type, it may be necessary to wait several years before sufficient operational experience has been obtained.

A potential "cut off" criterion for using only operational data may be when the confidence in the $\hat{\lambda}_{\mathrm{DU},1}$ is similar to the confidence in the $\lambda_{\mathrm{DU},0}$ from design.

For typical SIS components (detectors, sensors, and valves) in OREDA (2002), we note that the upper 95% confidence limit for the rate of DU-failures is often 2–3 times the mean value of the failure rate. In the design phase, the SIL verification is often based on OREDA data, this means that if $\lambda_{\mathrm{DU,1,U}} \leq k \cdot \hat{\lambda}_{\mathrm{DU},1}$, for a $k \approx 2.5$, our confidence in the estimate $\hat{\lambda}_{\mathrm{DU},1}$ should be comparable to our confidence in the initial estimate. We therefore have the criterion:

$$\frac{1}{2t_1} z_{0.95,2(x_1+1)} \leq k \cdot \frac{x_1}{t_1} \tag{7}$$

By solving this inequality, we see that it is fulfilled when $x_1 \geq 2$, and that this criterion is independent of the aggregated time in service t_1. This means that if we have observed 2 or more DU-failures, our confidence in the empirical estimate $\hat{\lambda}_{\text{DU},1}$ should be comparable to our confidence in the initial estimate.

In the situation where no DU-failure has been observed ($x_1 = 0$), it may be useful to reflect on the following two situations:

1. A very low initial DU-failure rate: The operational experience (no DU-failures) has (so far) confirmed this and there is no evidence to update the initial failure rate.
2. A very high initial DU-failure rate: The fact that no failure has occurred during the observation period indicates that the initial failure rate estimate may be too high.

Alternative II: Bayesian update:

When the criteria for sufficient operational experience are not met, we suggest a Bayesian update of the failure rate estimate.

In this case, we regard the initial DU-failure rate as a random variable $\Lambda_{\text{DU},0}$ with a prior distribution that expresses our uncertainty. Our prior knowledge about the rate of DU-failures may come from OREDA (2002), own previous experience, expert judgement, and so on. A common approach is to assume that our uncertainty can be modeled by a Gamma distribution with parameters α and γ (Rausand and Høyland 2004), with mean and standard deviation:

$$E(\Lambda_{\text{DU},0}) = \frac{\alpha}{\gamma} \tag{8}$$

$$\text{SD}(\Lambda_{\text{DU},0}) = \frac{\sqrt{\alpha}}{\gamma} \tag{9}$$

To determine these two parameters, we need to express our belief about $\Lambda_{\text{DU},0}$. This may be done by assuming that the mean rate $E(\Lambda_{\text{DU},0}) = \alpha/\gamma$ has a specific value $\lambda_{\text{DU},0}$ and that we can specify a conservative estimate, $\lambda^*_{\text{DU},0}$ which is approximately one standard deviation bigger than the mean value, such that $\lambda^*_{\text{DU},0} - \lambda_{\text{DU},0} = \text{SD}(\Lambda_{\text{DU},0}) = \sqrt{\alpha}/\gamma$. We therefore need to express our belief by giving numerical values for $\lambda_{\text{DU},0}$, the mean rate of DU-failures, and $\lambda^*_{\text{DU},0}$, a conservative estimate of the DU-failure rate. When these two parameters are specified, we may solve for α and γ:

$$\gamma = \frac{\lambda_{\text{DU},0}}{\left(\lambda^*_{\text{DU},0} - \lambda_{\text{DU},0}\right)^2} \tag{10}$$

and

$$\alpha = \gamma \cdot \lambda_{\text{DU},0} \tag{11}$$

Selecting the conservative estimate $\lambda^*_{\text{DU},0}$ is not always an easy task. We suggest three alternatives approaches:

- The analyst specifies a conservative estimate $\lambda^*_{\text{DU},0}$, based on her own belief and/or other sources of data.
- If the analyst has no suggestion, we use OREDA as basis for specifying a conservative estimate. From OREDA, we often find that the standard deviation for the rate of dangerous failures is comparable with the mean value itself. A conservative estimate $\lambda^*_{\text{DU},0}$ may therefore be set equal to $2 \cdot \lambda_{\text{DU},0}$.
- In either cases, the conservative estimate $\lambda^*_{\text{DU},0}$ should not take larger values than $5 \cdot 10^{-7}$ failures/hour. This limit is set for practical reasons, to avoid that the prior estimate totally overweighs the operational experience.

This means that $\lambda^*_{\text{DU},0}$ is chosen as:

$$\lambda^*_{\text{DU},0} = \\ = \max\left\{\text{user specified value}, 2 \cdot \lambda_{\text{DU}}, 5 \cdot 10^{-7}\right\} \\ \text{failures/hour} \tag{12}$$

The recorded experience data, i.e., the number, x_1 of DU-failures during the aggregated time in service t_1, can now be used to calculate the new (posterior) failure rate estimate:

$$\tilde{\lambda}_{\text{DU},1} = \frac{\alpha + x_1}{\gamma + t_1} \tag{13}$$

If desired, we may now express our uncertainty about this estimate by a 90% credibility interval. For details, e.g., see Rausand and Høyland (2004).

Depending on the amount of experience data, we may use either $\hat{\lambda}_{\text{DU},1}$ or $\tilde{\lambda}_{\text{DU},1}$ as estimate for the rate of DU-failures. Note that $\tilde{\lambda}_{\text{DU},1}$ can always be used, while $\hat{\lambda}_{\text{DU},1}$ requires at least 2 observed DU-failures.

The new estimate, $\hat{\lambda}_{\text{DU},1}$ or $\tilde{\lambda}_{\text{DU},1}$, can be used to determine the actual PFD_1.

As stated above, it may be required to update the actual PFD at several stages during the operational stage. We have shown how to do this in the first stage, and the same approach may be used also for later stages. The only main difference is that we in stage i, use the posterior distribution from stage $i-1$ as prior distribution.

We will next discuss how the updated failure rate estimate is used as basis for evaluating the functional test interval.

4.1 *Step 4: Evaluating the functional test intervals*

A simplified approach for evaluating the functional test interval was introduced by Hauge and Lundteigen (2008). In the following, a flexible and pragmatic approach for changing the functional test interval is described, starting with the following restrictions and assumptions:

- The test interval can never be *more* than doubled or halved. This assumption is made to ensure some conservatism in how much the functional test interval may be altered in a single step.
- The maximum allowed length of the functional test interval is 36 months. A 36 months interval corresponds to the longest interval between revision stops for oil and gas installations on the Norwegian continental shelf, and we assume that the functional status of all safety critical equipment should be verified with no longer intervals than this.
- The functional test intervals follow a discrete scale, corresponding to the frequently used intervals for testing and overhauls; 1 month, 3 months, 6 months, 9 months, 12 months, 18 months, 24 months, and 36 months.
- The initial test interval τ_0, was selected in the design phase such that the SIL requirement was fulfilled.

We consider the first update ($i = 1$). By the approach in step 3, we estimate the rate of DU-failures by either $\hat{\lambda}_{\text{DU},1}$ or $\tilde{\lambda}_{\text{DU},1}$. In the following, assume that we have chosen to use the Bayesian estimate $\tilde{\lambda}_{\text{DU},1}$. (If we use the empirical estimate, we get the same formulas). Next, determine the 90% confidence (or credibility) interval for $\lambda_{\text{DU},1}$. We then calculate the ratio $\lambda_{\text{DU},0}/\tilde{\lambda}_{\text{DU},1}$. This ratio indicates the fractional change in failure rate and thus the "allowed" change of the test interval. By using eq. (1), an updated test interval τ_1 can now be calculated as:

$$\tau_1^* = \frac{\lambda_{\text{DU},0} \cdot \tau_0}{\tilde{\lambda}_{\text{DU},1}} \tag{14}$$

where τ_0 is the initially selected functional test interval and τ_1^* is the new (preliminary) functional test interval. The final decision on a new functional test interval depends on how τ_1^* differs from τ_0.

Case A: τ_1^* is larger than τ_0

The main rule is that τ_1 is set equal to the first allowed test interval below τ_1^*. Assume that τ_0 is 9 months. If $\tau_1^* \approx 13$ months, then τ_1 is set equal to 12 months.

If τ_1^* is two times or more the initial test interval, we add some conservatism to the decision, using the same rule as was introduced for doubling the functional test interval in Hauge and Lundteigen (2008):

> If $\tilde{\lambda}_{\text{DU},1}$ is less than half the $\lambda_{\text{DU},0}$ and the entire 90% credibility (or confidence) interval for the $\lambda_{\text{DU},1}$, is below $\lambda_{\text{DU},0}$, then the functional test interval can be considered doubled (e.g., from one year to two years).

If doubling is not recommended, then τ_1 is set equal to next allowed test interval below the doubled interval.

Example: Consider a situation where τ_0 is 6 months and τ_1^* is calculated to 13 months. Due to the restriction in how much the interval can be changed in one step, the maximum length of the new interval would be 12 months. If we assume that a doubling of the functional test interval is not recommended by the rule above, then τ_1 is set equal to 9 months, which is the first allowed interval below 12 months.

Case B: τ_1^* is less than τ_0

The main rule is that τ_1 is set equal to the first allowed test interval above τ_1^*. If $\tau_1^* \approx 7$ months, then τ_1 is set equal to 9 months.

If τ_1^* is half or less the initial test interval, we add some conservatism to the decision, using the same rule as was introduced for halving the functional test interval in Hauge and Lundteigen (2008):

> If $\tilde{\lambda}_{\text{DU},1}$ is more than twice the $\lambda_{\text{DU},0}$ and the entire 90% credibility (or confidence) interval for the $\lambda_{\text{DU},1}$, is above $\lambda_{\text{DU},0}$, then the functional test interval should be halved (e.g., from one year to 6 months).

If halving is not recommended by the rule, then τ_1 is rounded *up* to the next allowed test interval above the halved interval.

Example: Consider a situation where τ_0 is 12 months and τ_1^* is calculated to 5 months. Due to the restriction in how much the interval can be changed in one step, the new interval should not be reduced below 6 months. If we assume that a halving of the functional test interval is not recommended by the rule above, then τ_1 is set equal to 9 months, which is the first allowed interval above 6 months.

A more detailed discussion of the rationale for doubling and halving is given in Hauge and Lundteigen (2008).

Subsequent updates

For the next update of the functional test interval, we replace τ_0 by τ_1 and denote the new failure rate estimate $\tilde{\lambda}_{\text{DU},2}$. In subsequent updates, the process is repeated for τ_i, $i = 2...$

4.2 *Case example*

Assume that one critical DU-failure has occurred during three years of operation of 35 blowdown valves.

Further assume that the initially selected $\lambda_{\text{DU},0}$ was $2.9 \cdot 10^{-6}$ failures per hour, and that a functional test interval of 12 months was sufficient to meet the reliability target. Is a change in the functional test interval recommended?

Since only one DU-failure has been observed, it is not recommended to rely on operational experience alone. Instead, we do a Bayesian update of the failure rate.

To determine the parameters α and γ in eqs. (10) and (11), we need to select a conservative estimate $\lambda^*_{\text{DU},0}$. We assume that the user finds it difficult to specify such a value, and that the $\lambda^*_{\text{DU},0}$ is set equal to $2 \cdot \lambda_{\text{DU},0} = 5.8 \cdot 10^{-6}$ failures/hour based on eq. (12). By using eqs. (10) and (11), we obtain $\gamma = 3.5 \cdot 10^6$ and $\alpha = 1$, and the new failure rate then becomes:

$$\begin{aligned} \tilde{\lambda}_{\text{DU}} &= \frac{\alpha + x}{\gamma + t_n} = \frac{1+1}{3.5 \cdot 10^5 + 9.2 \cdot 10^5} \\ &= 1.6 \cdot 10^{-6} \text{ hours}^{-1} \end{aligned} \quad (15)$$

The new (preliminary) test interval becomes:

$$\begin{aligned} \tau_1^* &= \frac{\lambda_{\text{DU}}}{\hat{\lambda}_{\text{DU}}} \cdot \tau = \frac{2.9 \cdot 10^{-6}}{1.6 \cdot 10^{-6}} \cdot 12 \text{ months} \\ &\approx 22 \text{ months} \end{aligned} \quad (16)$$

and we find that $\tau_1^* > \tau_0$, but $< 2 \cdot \tau_0$. This means that the new functional test interval τ_1 is set equal to the first allowed test interval below 22 months, i.e., 18 months.

A decision on whether or not to change the functional test interval should not be determined from quantitative calculations alone. Hauge and Lundteigen (2008) propose a checklist that addresses other issues of importance in the decision-making. This list should also be used with step 4 here before the final decision is made.

5 EXPERIENCE FROM FAILURE ANALYSIS

Failure reviews often indicate that too optimistic failure rates are used in design to calculate the PFD. In some cases, valves have been given failure rates that are 10–100 times better than the historical performance of similar valves in e.g., OREDA. One consequence is unrealistic target indicator values, i.e., that no DU failure is to be recorded during the entire lifetime of the installation. As a result, many plant owners often have to perform much more frequent testing for this equipment than what was initially assumed in design.

Certain components are not given unique tag numbers in the maintenance system, like for example programmable electronic controllers (PLCs) and fire centrals. Failures are instead recorded against other connected components, for example transmitters and fire and gas detectors. In addition, software related failures, which are highly relevant for PLCs and fire centrals, are often not recorded in the maintenance system at all, but in separate systems. This means that the performance of the PLCs is not fully known.

An important premise for having a negligible contribution from dangerous detected (DD) failures on the PFD is that these failures are corrected within *short* time, typically a few hours. However, there are several examples where the follow-up of DD-failures is not given sufficient attention and prioritization. Perhaps are the failures considered 'safe' since they have been notified. Or perhaps have the burden on the engineers in reducing failures been transferred to the operators to follow-up (too frequently occurring) dangerous failures.

DD-failures that occur close in time are sometimes reported as *one* failure, rather than as separate failure events. This may also be the case for spurious alarms. Counting the number of such failures may therefore not produce very accurate basis for estimating the corresponding failure rates.

For some components, nearly all reported failures have the same or similar failure causes. Examples comprise too long closing time for blowdown valves, incorrectly calibrated gas detectors, and incorrectly calibrated level transmitters. In the latter case, the reason for incorrect calibration may be confusion about the different ranges and readings for level transmitters located at the same vessel. As a result, it should be considered to increase the awareness to common cause failures, both during failure analysis and through training and procedures, for example as suggested by Lundteigen and Rausand (2007).

6 CONCLUSIONS AND FINAL REMARKS

The presented approach is useful in the day to day follow-up of safety instrumented systems, as the number of recorded failures may be deduced from the maintenance system at any time. However, the approach gives only an indication of the actual reliability performance, and more testing should be performed to compare decisions made, for example regarding the functional test intervals, with the results that would have been obtained in a more thorough analysis of all reliability influencing factors.

The approach refers mainly to information that is obtained from manually recorded failures in the maintenance system. In the future, the manual registration may be partly replaced by information from other systems, such as automatic shutdown re-cording systems, condition monitoring systems, and diagnostic systems. A challenge is to find ways to link registration of software failures into the maintenance systems.

This paper also outlines practical experience from

failure analysis that may be important to share with SIS designers and system integrators. One example is that the use of too optimistic failure rates leads to unrealistic reliability targets in the operational phase. Another example is that many failures share failure causes, thus indicating that more awareness and control with common cause failures are required.

7 ACKNOWLEDGEMENTS

The paper is based on experience from the Norwegian oil and gas industry, collected through the ongoing "PDS-BIP" research project funded by the Norwegian Research Council and the Norwegian PDS forum participants. PDS is a Norwegian abbreviation for reliability of computer-based systems, and the PDS forum is an initiative to gather industry and research institutes that work with reliability of SIS. The authors would like to thank the PDS forum participants for their support to work with this topic.

REFERENCES

Chhibber, S. and G. Apostolakis (1993). Some approximations useful to the use of dependent information sources. *Reliability Engineering & System Safety 42*(1), 67–86.

Hauge, S. and M. Lundteigen (2008). A new approach for follow-up of safety instrumented systems in the oil and gas industry. In *Safety, Reliability and Risk Analysis: Theory, Methods and Applications*, pp. 2921–2928. Leiden, The Netherlands: CRC Press/Balkema.

Hokstad, P., K. Øien, and R. Reinertsen (1998). Recommendations on the use of expert judgment in safety and reliability engineering studies. Two offshore case studies. *Reliability Engineering and System Safety 61*(1-2), 65–76.

IEC 61508 (1998). *Functional Safety of Electrical/Electronic/Programmable Electronic Safety-Related Systems*. Geneva: International Electrotechnical Commission.

IEC 61511 (2003). *Functional Safety - Safety Instrumented Systems for the Process Industry*. Geneva: International Electrotechnical Commission.

Kunttu, S. and H. Kortelainen (2004). Supporting maintenance decisions with expert and event data. In *Annual Reliability and Maintainability Symposium. 2004 Proceedings*, Piscataway, NJ, pp. 593–9. IEEE.

Lundteigen, M. (2009). *Safety instrumented systems in the oil and gas industry: Concepts and methods for safety and reliability assessments indesign and operation [978-82-471-1386-8 (electronic ver.), 978-82-471-1385-1 (printed ver.]*. Ph. D. thesis, Norwegian University of Science and Technology (NTNU), Trondheim, Norway.

Lundteigen, M. A. and M. Rausand (2007). Common cause failures in safety instrumented systems on oil and gas installations: Implementing defense measures through function testing. *Journal of Loss Prevention in the Process Industries 20*(3), 218–229.

Moss, T. (2005). *The reliability data handbook*. New York: ASME Press.

OREDA (2002). *OREDA Reliability Data* (4rd ed.). Available from: Det Norske Veritas, NO 1322 Høvik, Norway: OREDA Participants.

Rausand, M. and A. Høyland (2004). *System Reliability Theory: Models, Statistical Methods, and Applications* (2nd ed.). Hoboken, NJ: Wiley.

Røed, W. and T. Aven (2009). Bayesian approaches for detecting significant deterioration. *Reliability Engineering and System Safety 94*(2), 604–610.

Røed, W., A. Mosleh, J.-E. Vinnem, and T. Aven (2009, February). On the use of the hybrid causal logic method in offshore risk analysis. *Reliability Engineering and System Safety 94*(2), 445–55.

SINTEF (2006). *Reliability prediction methods for safety instrumented systems, PDS method handbook, 2006 edition*. Trondheim, Norway: SINTEF.

Summers, A. (2008). IEC 61508 Product Approvals – Veering Off Course. *Published on-line: http://www.controlglobal.com/articles/2008/187.html*.

Vatn, J. (2006). Procedures for updating test intervals based on experience data. In *Proceedings of the 30th ESReDA seminar*, pp. 185–198. Ispra, Italy: European Commission, Joint Research Centre.