TOLERABLE HAZARD RATE FOR FUNCTION WITH INDEPENDENT SAFETY BARRIER ACTING AS FAILURE DETECTION AND NEGATION MECHANISM

Résumé
On s’intéresse ici aux dispositifs de sécurité comportant une barrière qui détecte des défaillances contraires à la sécurité et les neutralise en amenant le système dans un état de repli sécuritaire lorsqu’elles se produisent. Ce type de dispositif se rencontre par exemple, dans les centrales ou réseaux électriques, dans les industries de process ou dans les systèmes ferroviaires. La question posée est l’allocation d’objectifs quantitatifs de sécurité à la fonction principale ainsi qu’à la barrière, en vue de limiter le risque d’accident en respectant la fréquence acceptable de danger (THR) qui a été prescrite. On présente une méthode qui, contrairement par exemple à celle qui est sous-jacente à l’annexe A4 de la norme CENELEC EN 50129 de la signalisation ferroviaire, ne fait pas l’hypothèse d’un retour immédiat du système à l’état nominal à partir de l’état de repli sécuritaire. Dans l’approche présentée ici, on obtient la probabilité, en fonction du temps, que le système réside dans un état sûr, ainsi que le taux de transition, fonction du temps et asymptotique, vers un état dangereux. La méthodologie repose sur la résolution des équations de Chapman-Kolmogorov en régime transitoire pour la chaîne de Markov décrivant le dispositif. La comparaison avec la méthode de la norme EN 50129 confirme que celle-ci peut conduire à des prévisions optimistes, donc potentiellement à sous-évaluation des risques. Il est important avant d’appliquer toute formule de calcul du THR du système, de préciser clairement toutes les hypothèses sous-jacentes relatives à la maintenance et au régime de fonctionnement du système étudié.

Summary
The present work addresses safety systems where a safety barrier detects wrong-side failures and mitigates them by leading the system to a safe state upon their occurrence. This type of safety philosophy is found for instance in electric power plants or networks, the process industries or in railway systems. The key question is that of allocating safety targets both to the main function and to the safety barrier in order to limit the risk of accident, by complying with a prescribed tolerable hazard rate (THR). A method is presented which, in contrast to the one implicit in Annex A4 of the CENELEC EN 50129 railway signaling standard, does not assume an instant restoration of the system to its nominal state once it has reached the safe state following the triggering of the barrier. In the approach presented here, the probability of the system residing in a safe state is derived, as a function of time, and so is the transition rate (time-varying or asymptotic) to the unsafe state, i.e. the system hazard rate. The method is based on the explicit resolution of the time-varying Chapman Kolmogorov equations for the continuous time Markov chain that models the system. Comparison with the method of the above-mentioned CENELEC standard shows that the latter may lead to optimistic predictions for the system hazard rate, and therefore may potentially underestimate risks. It is very important, before applying any hazard rate formula, to make explicit all underlying assumptions on maintenance and operating regime of the system under study.

Introduction and Problem Statement
Safety risks in complex systems such as are found in aerospace, power plants or rail transport, are often mitigated by so-called safety barriers, i.e. mechanisms which, once a functional failure has occurred, detect this occurrence and lead the system to a safe state. In safety studies, the types of failures of interest are the hazards, i.e. the failures which cause some harm (jauntily defined according to the context). A safe state is a state where no harm may occur. For instance, in a rail transport system, an emergency braking system can bring the train to stop and the passengers are then considered safe. In a nuclear power plant, a partial or total controlled plant shut-down will result in a safe state. The safety barrier in those examples is the mechanism that triggers an emergency braking or a plant shut-down upon detection of a hazard occurrence. For instance, in communication based train control, if a train does not receive its movement authority in a prescribed time frame due to a technical failure in the data communication system, the automated train protection (ATP) function will trigger an emergency braking to prevent possible collisions, i.e. it will act as a barrier to prevent the system from evolving to an unsafe state (where the train would collide with the preceding train). A barrier must therefore first detect the hazards that may be caused by functional failures, and then negate them, i.e. counter their effects.

Any sound safety analysis must take into account, not only the functional failures that lead to the hazards in the first place, but also the possibility of a failure of the safety barrier itself. If such a failure occurs before the functional failure happens, then the barrier will not be able to fulfil its role once such a functional failure takes place. If a tolerable hazard rate (THR) is defined for the hazard, i.e. the maximal frequency of occurrence that is acceptable for the hazard in a given accident scenario, one must allocate constraints to the functional failure rate and the barrier failure rate, as well as the detection and negation times, such that, if those constraints are complied with, the hazard rate will not exceed the THR.
Doing so requires the use of a model. In the current CENELEC EN 50129:2003 standard, the model used is a continuous time Markov chain (CTMC). It is assumed that the functional failures and the failures of the safety barrier are independent events (if not, common cause failures must be modeled, with their own failure rate). It is further assumed in this model that, once the safe state has been reached, the system is immediately restored to its initial, nominal operating condition.

For instance, in the above example, that would mean that, once the ATP had detected a hazard occurrence due to an ATO functional failure, and then had triggered the emergency brake, the train could immediately return to its nominal operation. Under this assumption, a simple formula is obtained for the THR as a function of the functional hazard rate, the barrier failure rate, the detection and negation times. This formula is symmetric between the functional failure rate and the barrier failure rate. However, this assumption is not always justified in practice. In most cases, restoration is not immediate. Furthermore, restoration does not even need to be assumed in the analysis. This is why, in the new provisional version of EN 50126, prEN 50126-2, a model has been proposed where no restoration is assumed. Based on the corresponding Markov chain models, the transient and steady-state probabilities for the system being in a safe or unsafe state are calculated. Subsequently, the transition rate (transient and asymptotic) to the unsafe state is also expressed in terms of the input parameters. Those models then allow for the allocation of safety targets to the main function and the safety barrier to guarantee system compliance with the THR. Obviously the allocations will be different depending on the assumptions.

In a first part, the symmetric model depicted in Annex A.4 of the EN 50129 standard (where immediate restoration is assumed) is recalled. Subsequently, the new proposed model (in the prEN 50126-2 standard) is presented and solved. Both models are then illustrated on an example and concluding statements follow.

It is very important in using those models to be fully aware of the assumptions they call upon rather than keeping those implicit.

### General model

The model which will be used is defined by the continuous time Markov chain of Figure 1.

![Markov Model for functional failure and safety barrier](image)

The meaning of the notation is as follows:

- The tolerable hazard rate of the function is denoted \( \text{THR}_F \) or \( \lambda_1 \)
- The tolerable hazard rate of the safety barrier is denoted \( \text{THR}_G \) or \( \lambda_2 \)
- The safe down time, \( \text{SDT}_F \), refers to the sum of the average times needed to detect and negate the technical hazard.
- The safe down rate, \( \text{SDR}_F \), is the reciprocal of the safe down time, and is denoted \( \mu_1 \):
  \[
  \text{SDR}_F = \frac{1}{\text{SDT}_F} = \mu_1 \quad \{1\}
  \]
- The mean time to detect and negate the failure of the safety barrier is denoted \( \text{SDT}_G \), and its reciprocal, \( \text{SDR}_G \), or \( \mu_2 \):
  \[
  \text{SDR}_G = \frac{1}{\text{SDT}_G} = \mu_2 \quad \{2\}
  \]

Finally, from the safe state \( Z_0 \), the system can be restored to its initial nominal state \( Z_0 \) with a restoration rate \( \mu_3 \).

It is to be remarked that \( \text{SDR}_F \) and \( \text{SDR}_G \) very much depend on the operational context. The “restoration” step (as proposed in prEN 50126-2) may in some cases include the time to complete a mission, or the time until the end of the operational day (if logistic constraints prevent an earlier action). The initial (nominal state) is denoted \( Z_0 \). The system leaves this state either when a functional failure occurs (with rate \( \lambda_1 \)) or when a failure of the barrier occurs (with rate \( \lambda_2 \)), with these events leading to states \( Z_1 \) and \( Z_2 \) respectively. Once in \( Z_1 \) or \( Z_2 \), a second failure brings the system to the unsafe state \( Z_3 \), i.e. \( Z_3 \) is reached if either a functional failure is followed by a failure of the barrier, which takes place before the latter has been able to act, or if a functional failure takes place when the barrier is already in a failed state and therefore is incapable of bringing the system to safety.
The model is similar to that of a 1-out-of-2 repairable redundant structure. Using classical techniques for solving the steady-state Chapman-Kolmogorov equations, e.g. (Birolini 2010), it is straightforward to derive an expression for MTTF₀, i.e. the mean time to reach the unsafe state Z₃ from the initial state Z₀:

\[
MTTF₀ = \frac{1 + \frac{\lambda_1}{\lambda_2 + \mu_1} + \frac{\lambda_2}{\lambda_1 + \mu_2} + \left(\frac{\lambda_1 \mu_1}{\lambda_1 + \mu_1} + \frac{\lambda_2 \mu_2}{\lambda_2 + \mu_2}\right)}{\lambda_1 + \lambda_2 - \left(\frac{\lambda_1 \mu_1}{\lambda_1 + \mu_1} + \frac{\lambda_2 \mu_2}{\lambda_2 + \mu_2}\right)}
\]  

\[\text{(3)}\]

**Instant Restoration Assumption**

It is readily seen from (3) that the expression simplifies when instantaneous restoration to nominal state is assumed, i.e. \(\mu_3\) goes to infinity. If furthermore it is assumed that:

\[\lambda_1 \ll \mu_1 \text{ and } \lambda_2 \ll \mu_2, \text{ then}\]

\[MTTF₀ \to \frac{1}{\lambda_1 \lambda_2 \left(\frac{1}{\mu_1} + \frac{1}{\mu_2}\right)} \]  

\[\text{(4)}\]

Or the equivalent failure rate is given by

\[\Lambda = \lambda_1 \lambda_2 \left(\frac{1}{\mu_1} + \frac{1}{\mu_2}\right) \]  

\[\text{(5)}\]

Also noted:

\[\Lambda = THR_F THR_G (SDT_F + SDT_G) \]  

\[\text{(6)}\]

This is the formula provided in EN 50129 but without any comment on the assumptions on which it rests. From the above discussion, it is important to keep in mind that this expression is only valid if two conditions hold:

(A) There is ‘instant’ restoration of the system from state ZD to nominal state Z₀ : \(\mu_3 \to \infty\)

(B) \(SDF \ll THR\) and \(SDR \ll THR\).

Assumption (A) in particular is very strong.

**The model without restoration to nominal state**

**1. Problem Formulation**

In order to represent the fact that no restoration takes place from the safe state Z₀, one would have to let the restoration rate \(\mu_3\) vanish in the above model. Then the Markov chain is no longer ergodic: it has two absorbing states, the unsafe state Z₃ and the safe state Z₀. The method used to calculate MTTF₀ through Eq. (3) is no longer applicable. Instead, the transient (time-varying) Chapman-Kolmogorov equations will be solved and then the asymptotic probabilities and transition rates will be determined.

The Chapman–Kolmogorov equations for the system of Figure 1 are the following (the time variable t is omitted for ease of writing and the prime sign ‘ denotes the time derivative):

\[P'_{0} = (\lambda_1 + \lambda_2) P_{0} \]  

\[\text{(7)}\]

\[P'_{1} = \lambda_1 P_{0} - (\lambda_2 + \mu_1) P_{1} \]  

\[\text{(8)}\]

\[P'_{2} = \lambda_2 P_{0} - (\lambda_1 + \mu_2) P_{2} \]  

\[\text{(9)}\]

\[P'_{3} = \lambda_2 P_{1} + \lambda_1 P_{2} \]  

\[\text{(10)}\]

\[P'_{4} = \mu_1 P_{1} + \mu_2 P_{2} \]  

\[\text{(11)}\]

They must be completed by the balance condition, stating that the system is at all times in one and only one of states 0 to 4:
\[ 1 = P_0 + P_1 + P_2 + P_3 + P_4 \]  

[12]

2. Solution of Chapman-Kolmogorov Equations

Let us calculate \( P_i(t) \) for \( (i=0,1,2,3,4) \).

These values constitute the solution of Equations (7) to (12) with initial condition:

\[
\begin{align*}
P_0(0) & = 1 \\ P_i(0) & = 0 \quad (i=1,2,3,4)
\end{align*}
\]
as the system is assumed to be initially operational.

Equations (7) to (11) constitute a system of homogeneous linear differential equations and (12) is a normalizing condition. The system can be solved by the usual methods described in any elementary calculus textbook (for instance, [Ref.4]). In fact, \( P_0 \) is obtained easily from (7); then, upon substituting the solution in (8) and (9), \( P_1 \) and \( P_2 \) are obtained, and so on. The solution (13) is as follows:

\[
\begin{align*}
P_0(t) & = e^{-(\lambda_1+\lambda_2)t} \\ P_1(t) & = \frac{\lambda_1}{\mu_1 - \lambda_1} \left( e^{-(\lambda_1+\lambda_2)t} - e^{-(\lambda_2+\mu_1)t} \right) \\ P_2(t) & = \frac{\lambda_2}{\mu_2 - \lambda_2} \left( e^{-(\lambda_1+\lambda_2)t} - e^{-(\lambda_1+\mu_2)t} \right) \\ P_3(t) & = \frac{\lambda_1 \lambda_2}{\mu_1 - \lambda_1} \left( \frac{e^{-(\lambda_2+\mu_1)t}}{\lambda_2 + \mu_1} - \frac{e^{-(\lambda_1+\lambda_2)t}}{\lambda_1 + \lambda_2} \right) + \frac{\lambda_1 \lambda_2}{\mu_1 - \lambda_1} \left( \frac{1}{\lambda_1 + \lambda_2} - \frac{1}{\lambda_2 + \mu_1} \right) + \frac{\lambda_1 \lambda_2}{\mu_2 - \lambda_2} \left( \frac{1}{\lambda_1 + \lambda_2} - \frac{1}{\lambda_1 + \mu_2} \right) \\ P_4(t) & = \frac{\lambda_1 \mu_1}{\mu_1 - \lambda_1} \left( \frac{e^{-(\lambda_2+\mu_1)t}}{\lambda_2 + \mu_1} - \frac{e^{-(\lambda_1+\lambda_2)t}}{\lambda_1 + \lambda_2} \right) - \frac{\lambda_2 \mu_2}{\mu_2 - \lambda_2} \left( \frac{e^{-(\lambda_1+\mu_2)t}}{\lambda_1 + \mu_2} - \frac{e^{-(\lambda_1+\lambda_2)t}}{\lambda_1 + \lambda_2} \right) + \frac{\lambda_2 \mu_2}{\mu_2 - \lambda_2} \left( \frac{1}{\lambda_1 + \lambda_2} - \frac{1}{\lambda_1 + \mu_2} \right)
\end{align*}
\]

When \( t \to \infty \) then \( \lim_{t \to \infty} P_0(t) = \lim_{t \to \infty} P_1(t) = \lim_{t \to \infty} P_2(t) = 0 \)

Probability of hazardous condition:

\[
\lim_{t \to \infty} P_3(t) = P_3(\infty) = \frac{\lambda_1 \lambda_2}{\mu_1 - \lambda_1} \left( \frac{1}{\lambda_1 + \lambda_2} - \frac{1}{\lambda_2 + \mu_1} \right) + \frac{\lambda_1 \lambda_2}{\mu_2 - \lambda_2} \left( \frac{1}{\lambda_1 + \lambda_2} - \frac{1}{\lambda_1 + \mu_2} \right) \quad [14]
\]

Probability of safe condition:

\[
\lim_{t \to \infty} P_4(t) = P_4(\infty) = \frac{\lambda_1 \mu_1}{\mu_1 - \lambda_1} \left( \frac{1}{\lambda_1 + \lambda_2} - \frac{1}{\lambda_2 + \mu_1} \right) + \frac{\lambda_2 \mu_2}{\mu_2 - \lambda_2} \left( \frac{1}{\lambda_1 + \lambda_2} - \frac{1}{\lambda_1 + \mu_2} \right) \quad [15]
\]

It is observed that, if the failure rates of the main function and the barrier were equal (\( \text{TFR}_1 = \text{TFR}_2 = \text{TFR} \)) and if also the safe down rate of the function were equal to that of the barrier (\( \text{SDR}_1 = \text{SDR}_2 = \text{SDR} \)), then the steady-state probability of the safe state \( Z_4 \) and of the unsafe state \( Z_3 \) would be obtained respectively as the steady state availability and unavailability of a 1-out-of-2 system with failure rate \( \text{TFR} \) and restoration rate \( \text{SDR} \). In other words, in that special case:

\[
\begin{align*}
P_3 & = \frac{\text{TFR}}{\text{TFR} + \text{SDR}} \\ P_4 & = \frac{\text{SDR}}{\text{TFR} + \text{SDR}}
\end{align*}
\]
3. Transition rate to unsafe state

The equivalent transition rate to the unsafe state Z3 at time t, denoted \( \dot{\theta}(t) \), is, by definition, the limiting conditional probability per unit of time (when the time increment \( dt \) goes to zero) of the probability for the system to be in state Z3 at time \( (t + dt) \) given that it was in a safe state at time t.

As the two states from which the system can transition to state Z3 are states Z1 and Z2, there follows:

\[
\dot{\theta}(t) = \frac{\lambda_2 P_1 + \lambda_1 P_2}{P_1 + P_2} \quad \text{[18]}
\]

From the explicit solutions for \( P_i(t) \) derived above, \( \dot{\theta}(t) \) is therefore derived.

The following expression is obtained (assuming \( \lambda_1 - \mu_1 \) and \( \lambda_2 - \mu_2 \) different from zero):

\[
\dot{\theta}(t) = \frac{1}{\lambda_1} \left( 1 - e^{(\lambda_1 - \mu_1) t} \right) + \frac{1}{\mu_2 - \lambda_2} \left( 1 - e^{(\lambda_2 - \mu_2) t} \right)
\]

\[
= \frac{\lambda_1 \lambda_2}{(\mu_1 - \lambda_2) + \lambda_2 (\mu_1 - \lambda_1)} \quad \text{[19]}
\]

One also obtains the asymptotic equivalent transition rate to the unsafe state, \( \dot{\theta}(\infty) \), by letting t go to infinity.

If \( \lambda_1 < \mu_1 \) and \( \lambda_2 < \mu_2 \), then

\[
\dot{\theta}(\infty) = \frac{\lambda_1 \lambda_2 ((\mu_1 + \mu_2) - (\lambda_1 + \lambda_2))}{\lambda_1 (\mu_2 - \lambda_2) + \lambda_2 (\mu_1 - \lambda_1)} \quad \text{[20]}
\]

If \( \lambda_1 > \mu_1 \) and \( \lambda_2 < \mu_2 \), it is seen from [19] that \( \dot{\theta}(\infty) = \lambda_2 \)

Symmetrically, if \( \lambda_2 > \mu_2 \) and \( \lambda_1 < \mu_1 \), \( \dot{\theta}(\infty) = \lambda_1 \)

4. Special Cases

Assuming \( \lambda_i < \mu_i \) (i=1, 2), then [20] reduces to:

\[
\dot{\theta}(\infty) = \frac{\mu_1 + \mu_2}{\mu_1 + \lambda_2} \quad \text{[21]}
\]

In the special case where \( \mu_1 = \mu_2 \),

Eq. [21] becomes:

\[
\dot{\theta}(\infty) = 2 \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2} \quad \text{[22]}
\]

If furthermore one assumes that \( \lambda_2 << \lambda_1 \), then

\[
\dot{\theta}(\infty) \approx 2 \lambda_2 \quad \text{[23]}
\]

If on the other hand the failure rate of the barrier were equal to that of the function, i.e. if \( \lambda_1 = \lambda_2 = \lambda \), then the following would hold:

\[
\dot{\theta}(\infty) = \lambda \quad \text{[24]}
\]

Some asymmetric limiting cases are worth noting.

Let us assume \( \mu_i \) goes to infinity, as compared to \( \lambda_i \) (i.e. \( \mu_i >> \lambda_i \)).

This means that the speed with which the occurrence of a hazard is detected and protected by the safety barrier is much, much greater than the failure rate of the safety barrier (or equivalently the average time to detect and protect is much smaller than the MTBF of the safety barrier). Then, the only way for the system to reach the unsafe state is through Z2, i.e. if after a failure of the safety barrier, the hazard occurs (because, once in Z1 on the other hand, the system would move instantly to the safe state Z4). The corresponding hazard rate will be \( \lambda_1 \). This is confirmed by dividing both terms of the fraction in the right-hand side of [20] by \( \lambda_2^2 \), which yields:
\[ \theta(\infty) = \frac{\lambda_1 ((\frac{\mu_1}{\lambda_2} + \frac{\mu_2}{\lambda_2}) - (\frac{\lambda_1}{\lambda_2} + 1))}{\lambda_2 (\frac{\mu_2}{\lambda_2} - 1) + (\frac{\mu_1}{\lambda_2} - \frac{\lambda_1}{\lambda_2})} \]

Eq. (25) shows that, for \( \frac{\mu_1}{\lambda_2} \to \infty \), \( \lim \theta(\infty) = \lambda_1 \)

Likewise, for \( \frac{\mu_2}{\lambda_1} \to \infty \), \( \lim \theta(\infty) = \lambda_2 \).

The meaning of this second limiting case is that, if the detection and negation of a failure in the safety barrier occurs much faster than the occurrence of a technical hazard, the only way for the system to reach the unsafe state is through the occurrence of a failure of the safety barrier after a technical hazard has occurred and before it has been detected and protected by the safety barrier, i.e. through state Z1 (because, through the other possible route, once in Z2 the system would move instantly to the safe state Z4).

Discussion and numerical example

The limiting cases examined in the previous section show that, in some cases, the entire safety burden may rest on the safety barrier (when \( \theta(\infty) = \lambda_2 \)) or on the technical hazard (when \( \theta(\infty) = \lambda_1 \)).

Those situations are very different from those described by Eq. (6), where essentially the hazard rates of the main function and the barrier are "multiplied". The safety allocation in the cases here considered will therefore have to be a lot more stringent to reach the same system-level THR.

Let us consider the following examples.

In a modern automatic urban train system, there is an ATO (automated train operation) function responsible for train movement control (adapting speed to the operations context) and a safety–related ATP (automated train protection) function which acts as a safety barrier (allocated at SIL 4), triggering emergency brakes in case of ATO failure, for instance.

Let us assume that those two functions are characterized by the following orders of magnitude: \( \lambda = 2 \times 10^{-5} \text{ h}^{-1} \) for the hazard rate attached to the ATO and \( \lambda = 10^{-8} \text{ h}^{-1} \) for the hazard rate attached to the ATP.

The analysis presented in this paper and illustrated on an example does not require that assumption; it investigates directly the probability for the system to reach the unsafe state, and hence the transition rate to that state, i.e. the system hazard rate.

If now the non-repairable model is used by application of Eq. (20), the equivalent asymptotic THR is found to be equal to \( 2 \times 10^{-9} \text{ h} \). This represents the asymptotic rate of occurrence of the unsafe event. (the convergence with time is extremely fast).

Obviously if it is desired to reach a value of \( 10^{-9} \text{ h} \) for instance, a lower hazard rate for the ATP would be required: in fact it should be of the order of magnitude \( 10^{-10} \text{ h} \) instead of \( 10^{-9} \text{ h} \).

It can be seen that in this example the safety barrier (in this case the ATP) is the dimensioning parameter.

Conclusion

In the modeling of safety barriers that lead systems to a safe state upon occurrence of a hazard, several assumptions are often relied upon without always being stated explicitly. One such assumption, found in CENELEC EN 50129 standard for railway systems, is that, once the system has been brought to a safe state thanks to the safety barrier, it immediately is restored to nominal operation. Such a simplifying assumption leads to a symmetric formula for the system hazard rate (symmetric between the hazard rate of the main function and that of the barrier) which is exactly that of a one-out-of-two redundant structure.

The analysis presented in this paper and illustrated on an example does not require that assumption; it investigates directly the probability for the system to reach the unsafe state, and hence the transition rate to that state, i.e. the system hazard rate.

The approach followed so far in the standard is found to be optimistic and to rest on assumptions which are generally not realistic. This is the rationale for proposing, in the prEN 50126-2 draft standard, an alternative expression, based on the present work.

References