IMPROVING RELIABILITY AND REDUCING RISK BY SEPARATION

Michael Todinov
Oxford Brookes University
Department of Mechanical Engineering and Mathematical sciences,
Oxford, Wheatley, OX33 1HX, UK
mtodinov@brookes.ac.uk

ABSTRACT
The paper introduces the method of separation for improving reliability and reducing technical risk and provides insight into the various mechanisms through which the method of separation attains this goal. A comprehensive classification of techniques for improving reliability and reducing risk, based on the method of separation has been proposed for the first time. From this classification, three principal categories of separation techniques have been identified: (i) assuring distinct functions/properties/behaviour for distinct components or parts (ii) assuring distinct properties/behaviour at distinct time, value of a parameter, conditions or scale and (iii) distancing risk-critical factors.
The concept 'stochastic separation' of random events and methods for providing a stochastic separation have been introduced. It is shown that separation of properties is an efficient technique for compensating the drawbacks associated with a selection based on homogeneous properties. It is also demonstrated that the method of deliberate weak links and the method of segmentation can be considered as a special case of the method of separation. Finally, the paper demonstrates that the traditional reliability measure 'safety margin' is misleading and should not be used as a measure of the relative separation between load and strength.

Keywords: Generic principles, risk reduction, separation, reliability improvement; technical risk.

1. INTRODUCTION

A systematic classification of generic methods for reducing technical risk is crucial to safe operation, engineering designs and software. However, this very important topic has not been covered with sufficient depth in the reliability and risk literature. For many decades, the focus of reliability research has been primarily on reliability prediction rather than reliability improvement and risk reduction. Acquiring the relevant knowledge and data related to the failure mechanisms and quantifying all types of uncertainty, necessary for a correct prediction of the time to failure is a formidable task. However, this task does not need to be addressed if the focus is placed on the reliability improvement and risk reduction instead of reliability prediction. The generic reliability improvement and risk reduction methods do not normally rely on reliability data or on a detailed knowledge of physical mechanisms underlying possible failure modes. These methods derive their strength from generic laws, invariants and patterns associated with increased reliability and reduced risk. As a result, these methods are particularly useful in developing new designs, with no failure history and with insufficiently researched failure mechanisms. Recent work on formulating generic principles and methods for improving reliability and reducing technical risk has been done in (Todinov, 2015). The present paper contributes to the exiting work an important generic reliability improvement and risk reduction method referred to as '*the method of separation'*.

Harmful interaction of factors critical to reliability and risk is a major source of failures. Separation is the act of disuniting risk-critical factors. Separating risk-critical factors to reduce this harmful interaction is therefore a major avenue for improving reliability and

reducing risk. Surprisingly, the method of separation has not yet been discussed as a risk-reduction tool. Despite that some techniques used in engineering are clearly instances of the method of separation, they have never been recognised as such and have never been linked with this method.

Thus, *deliberate weak links* and *stress limiters* have already been used for preventing the stresses from reaching dangerous levels. The deliberate weak links are consciously designed weak points that are easily replaced (Eder, 2008) and usually protect expensive devices. As it will be demonstrated later, this technique is essentially an instance of a separation on a parameter but it has never been recognised as such and has never been linked with the method of separation.

Another example can be given with the concept '*barrier*' (Svenson, 1991; Leveson, 2011; Hollangel, 2016). Barriers have also been used as accident prevention tools and protection measure mitigating the consequences from an accident. A classification of barriers has been proposed (Eder and Hosnedl, 2008). Despite that barriers are also instances of separation, no link has ever been made with the method of separation. Barriers distancing triggers from hazards reduce the likelihood of an accident while barriers distancing hazards from targets reduce the consequences given that accident has occurred.

It is only recently that the method of separation has been suggested in (Todinov, 2015) as a potential risk reduction tool but the discussion of the separation method was very limited and did not cover mechanisms through which separation achieves reliability improvement and risk reduction. Furthermore, no classification of separation techniques has been proposed.

One of the ways to make engineering systems more efficient is to increase the temperature, pressure and speed of the operating fluids and reduce the cross sections of components to reduce weight. However, the increased temperature, pressure and speed of the fluids accelerate the degradation of the components while reducing the cross sections leads to increased stresses and stress amplitudes which increase the risk of failure due to fast fracture and fatigue. Similarly, increasing the efficiency of manufacturing requires increasing the speed of operations which leads to low precision and unreliability.

As a result, there is a constant struggle between efficiency and reliability which is a fertile ground for technical contradictions. As a result, separation has been applied in the *TRIZ* methodology for inventive problem solving (Altshuller, 1984, 1996, 2007) for resolving physical contradictions in engineering of the type: 'the object must have attribute A during one mode of use or during one stage of a particular process and the opposite attribute (not A) during an alternative mode of use or an alternative process stage'. An example of such type of separation has been introduced (Altshuller, 1984) to resolve the contradiction between the required attributes 'rigid' versus 'not rigid' (flexible)'. An example was given with the bicycle chain which is rigid on the level of a separate link and not rigid (flexible) on the macro level 'chain of links'. An example of time separation used in *TRIZ* to ensure mutually exclusive attributes is frequently given with the attributes of a pile driven into the soil. It needs to be sharp while being driven into the soil and blunt while supporting.

Despite that the utility of the separation principle used in *TRIZ* for resolving contradictory requirements cannot be questioned, the principle of separation introduced in *TRIZ* was primarily formulated as a way of generating inventive solutions by resolving extreme contradictions of the type 'attribute A is required and attribute A is not required'. In this sense, the separation principle in *TRIZ* is understood and practiced primarily in the sense of separation of mutually exclusive (incompatible) attributes and not necessarily as an act of distancing of factors to avoid their harmful interaction. However, the act of distancing of interacting factors has a direct implication to reliability and risk because the unreliability and risk of failure are very often a direct result from a harmful interaction of reliability-critical factors. A well-documented example is the interference of load and strength which are often

associated with significant variation (Carter, 1986; Lewis, 1996). During load-strength interference, failure materialises when the random instance of load exceeds the random instance of strength.

Furthermore, the separation in *TRIZ* is not oriented towards reliability improvement and risk reduction and no specific treatment has been provided in the *TRIZ* methodology related to the mechanisms through which the separation works in increasing reliability and reducing risk. No specific discussion regarding the mechanisms through which the method of separation increases reliability has been presented in more recent literature related to *TRIZ* (Terninko et al., 1998; Savransky, 2000; Orloff, 2006, 2012; Rantanen and Domb, 2008; Gadd, 2011). In addition, the separation as a problem-solving tool in *TRIZ*, has been introduced in a rather narrow context: primarily as a separation in space, time, between the parts and the whole and separation on a condition. However, these are not the only instances when separation can be performed. Separation can essentially be performed on any selected parameter: temperature, pressure, strength, stiffness, mass, size, etc. Separation can also be performed on various functions and even on the geometry of the component. The lack of discussion related to the mechanisms through which the different separation techniques work does not allow to exploit the full potential of the separation method in the area of the reliability improvement and risk reduction.

More importantly, the *TRIZ* theory never considered 'stochastic separation', for which the separation is guaranteed only with a certain probability. *TRIZ* also never considered logical separation where no time, space, or separation on a condition is present yet the dangerous proximity of hazards and targets is prevented.

In summary, no systematic analysis and classification currently exist of the fundamental techniques and mechanisms through which the method of separation improves reliability and reduces risk. To the best of our knowledge, no comprehensive work currently exists on the method of separation applied for reliability improvement and risk reduction and the classification of the various separation techniques achieving this goal.

Consequently, this paper provides the first comprehensive introduction to the powerful method of separation for improving reliability and reducing risk. The mechanisms through which the method of separation works and the classification of separation techniques are discussed for the first time through various application examples.

2. METHODS

Separating risk-critical factors to reduce harmful interaction is a major avenue for improving reliability and reducing risk. Other major avenues for improving reliability and reducing risk is separating distinct functions/properties/behaviour for distinct components/parts and separating properties and behaviour at distinct time, scale or conditions. The method of separation presented in this paper is based on a large number of available solutions improving reliability and reducing risk in various engineering fields. They achieve reliability improvement and risk reduction through one of the three distinct avenues discussed earlier. Each of the available solutions was analysed to verify and, in some cases to quantify, its effect on the reliability improvement and risk reduction.

The available solutions were also analysed for recurring patterns and invariants. A certain level of abstraction was used to strip the observations in different engineering fields from their specific engineering context in order to uncover the underlying act of separation. This approach helped uncover hidden patterns and reach conclusions. Such was for example the conclusion that introducing deliberate weak links to reduce risk is essentially an act of separation.

From the large body of observations, patterns and invariants emerged which were captured and separated into distinct categories, classes and individual techniques. This is essentially a

process of distillation of generic principles and techniques for improving reliability and reducing risk achieved by separation. A classification summarising these categories, classes and techniques has been presented in Fig.1. From this classification, three principal categories of separation techniques have been identified: (i) techniques assuring distinct functions/properties/behaviour for distinct components/parts (ii) techniques assuring distinct properties/behaviour at distinct time, value of a parameter, conditions or scale and (iii) techniques involving distancing risk-critical factors.

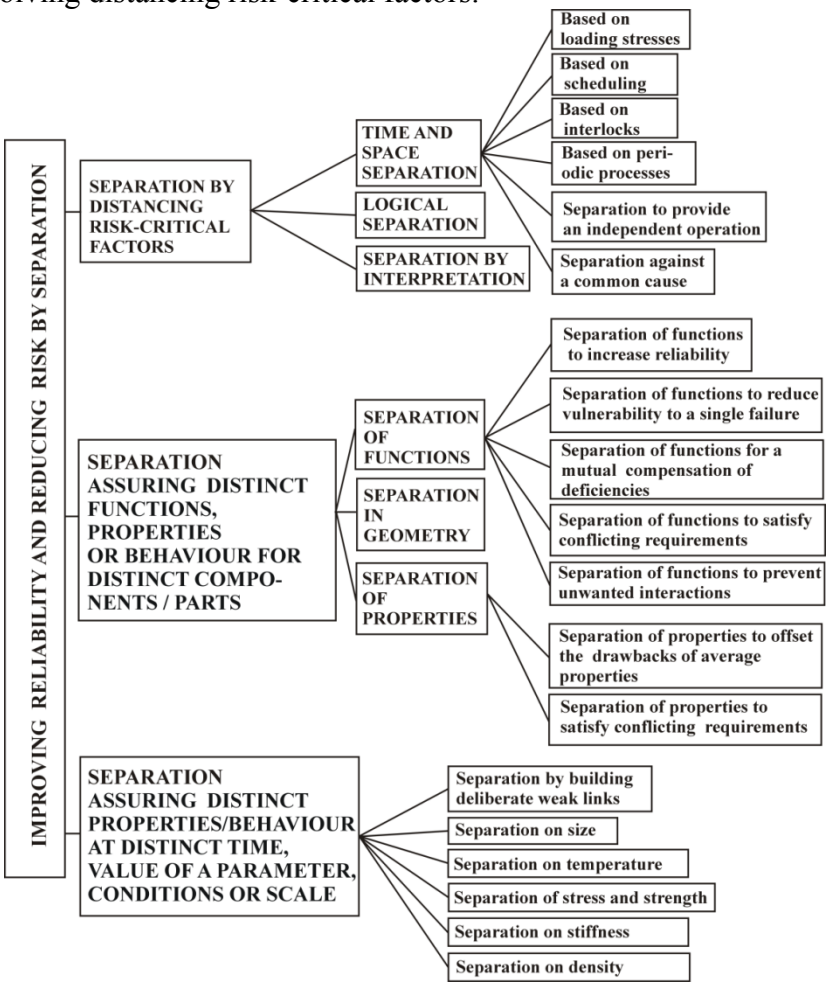

Figure 1. Classifications of various techniques for improving reliability and reducing risk by separation.

In what follows, each category of separation techniques is discussed in detail.

3. SEPARATION ASSURING DISTINCT PROPERTIES/BEHAVIOUR AT DISTINCT VALUES OF A PARAMETER, AT DISTINCT TIMES, AT DISTINCT CONDITIONS OR SCALE

Reducing risk by a separation on a risk-critical parameter is present when different characteristics of an object at different values of a risk-critical parameter are ensured to reduce the likelihood of failure or the consequences given that failure has occurred.

Separation of properties is an efficient method countering the drawbacks associated with average properties. The separation of properties essentially ‘assures’ distinct properties to different parts so that the overall risk of failure is reduced.

3.1 Separation of stress and strength (mechanical stress, current, voltage, pressure, etc.)

3.1.1 Critical weaknesses of the traditional reliability measures 'safety margin' and 'loading roughness' as measures of the degree of separation between load and strength.

The loading stress is a risk-critical parameter. In many cases, the reliability on demand is determined by the probability of a relative configuration of the load and strength, in which load is smaller than strength. Reliability on demand is controlled by the two risk-critical parameters 'load' and 'strength', characterised by distinct distributions.

Load and strength are much broader concepts than their mechanical interpretation. Any two interacting random parameters can be interpreted as 'load' and 'strength'. Load and strength, for example, could stand for 'demand' and 'supply', 'rate of damage' and 'rate of recovery', 'corrosion rate' and 'corrosion resistance', 'stress intensity' and 'fracture toughness', etc.

In a load-strength interaction, risk is strongly dependent on the degree of relative separation of the load variation and strength variation.

Consider a load distribution characterised by a mean μ_L and standard deviation σ_L and strength distribution characterised by a mean μ_S and standard deviation σ_S. A common measure quantifying the degree of relative separation of the load and strength is the reliability index, defined as

$$\beta = \frac{\mu_S - \mu_L}{\sqrt{\sigma_S^2 + \sigma_L^2}} \tag{1}$$

The reliability on demand *R* characterising the load-strength configuration is given by

$$R = \Phi(\beta) = \Phi\left(\frac{\mu_S - \mu_L}{\sqrt{\sigma_S^2 + \sigma_L^2}} \right) \tag{2}$$

where $\Phi(\bullet)$ is the cumulative distribution of the standard normal distribution (O'Connor 2002, Carter 1986, 1997).

From equation (2), it is easy to see that a larger difference $\mu_S - \mu_L$ between the means of the strength and load distribution and smaller variances σ_L^2 and σ_S^2 of the load and strength distributions, lead to a larger reliability index β and larger reliability on demand. The safety margin has been used as a measure for the relative separation of the load and strength distribution even for load and strength distributions which do not follow the Gaussian distribution (O'Connor 2002).

In what follows, it is demonstrated that for load and strength that do not follow the normal distribution, the traditional reliability measure safety margin is misleading and cannot be used to measure the degree of relative separation between load and strength.

Consider the load and strength distributions from Fig.2a. The figure shows a case where a low safety margin $\beta = (\mu_S - \mu_L)/\sqrt{\sigma_S^2 + \sigma_L^2}$ exists ($\mu_S - \mu_L$ is small and $\sigma_S^2 + \sigma_L^2$ is large) yet the reliability on demand is high. In Fig.2a, μ_S and μ_L are the mean values of the strength and load; σ_S and σ_L are the corresponding standard deviations. Now consider Fig.2b which has been obtained by reflecting symmetrically the distributions from Fig.2a with respect to axes r_1 and r_2, parallel to the probability density axis. Since the reflections do not change the variances of the distributions, the only difference is the larger difference of the means $\mu_S' - \mu_L' > \mu_S - \mu_L$ (Fig.2b). Despite the larger new safety margin

$$\beta' = \frac{\mu_S' - \mu_L'}{\sqrt{\sigma_S^2 + \sigma_L^2}} > \beta = \frac{\mu_S - \mu_L}{\sqrt{\sigma_S^2 + \sigma_L^2}}$$

the reliability on demand related to the load-strength configuration in Fig.2b is smaller than the reliability related to the configuration in Fig.2a. Clearly, the safety margin concept applied without considering the shape of the interacting distribution tails can be very misleading.

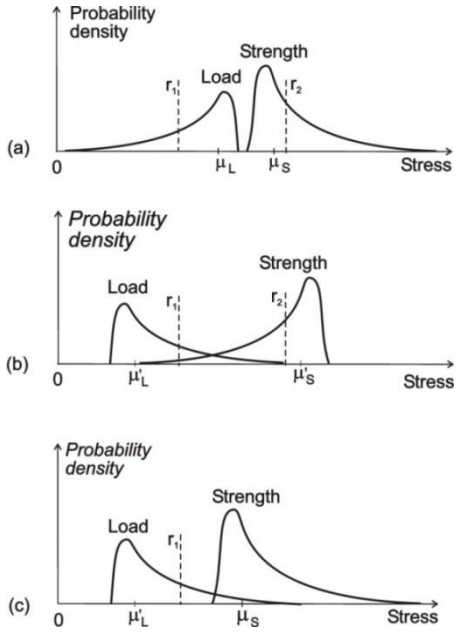

Figure 2. A counterexample showing that for skewed load and strength distribution, the traditional reliability measures 'reliability index' is very misleading.

Similar considerations are valid regarding the parameter *loading roughness* $\sigma_L / \sqrt{\sigma_L^2 + \sigma_S^2}$ introduced in (Carter, 1986, 1997). If only the load in Fig.2a is reflected symmetrically with respect to axis r_1, the loading in Fig.2c is obtained. Since the standard deviation σ_L of the load has not been affected by the reflection, the loading roughness in Fig.2c, calculated from $\sigma_L / \sqrt{\sigma_L^2 + \sigma_S^2}$, is the same as in Fig.2a, despite the much more severe type of loading.

3.1.2 Interaction between the upper tail of the load distribution and the lower tail of the strength distribution.

The problems outlined in the previous section do not exist if for load and strength which do not follow a normal distribution, a numerical integration is used to quantify the relative separation between load and strength. The most important aspect of the load-strength interaction is the interaction of the upper tail of the load distribution and the lower tail of the strength distribution (Fig.3). Consequently, only information related to the lower tail of the strength distribution and the upper tail of the load distribution is necessary.

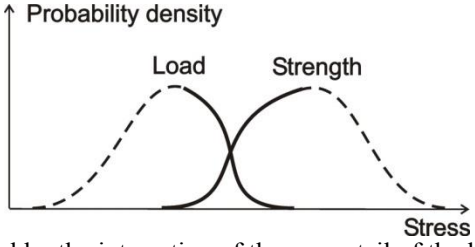

Figure 3. Reliability is determined by the interaction of the upper tail of the load distribution and the lower tail of the strength distribution.

The values from the lower tail of the strength distribution and the upper tail of the load distribution control reliability, not the values corresponding to the other parts of the distributions (Fig.3).
Consequently, an adequate model of the strength distribution should faithfully represent its lower tail and an adequate model of the load distribution should faithfully represent its upper tail.
The interaction of the upper tail of the load distribution and the lower tail of the strength distribution can be quantified. Consider the load-strength integral (Lewis, 1996) which gives the probability of failure p_f for a single load application:

$$p_f = \int_{S\min}^{S\max} [1 - F_L(x)] f_S(x)\, dx \tag{3}$$

where $F_L(x)$ is the cumulative distribution of the load and $f_S(x)$ is the probability density distribution of the strength.
Suppose that the $S_{\min}$ and $S_{\max}$ in Fig.4 correspond to stress levels for which $f_S(x) = 0$ if $x < S_{\min}$ or $x > S_{\max}$.

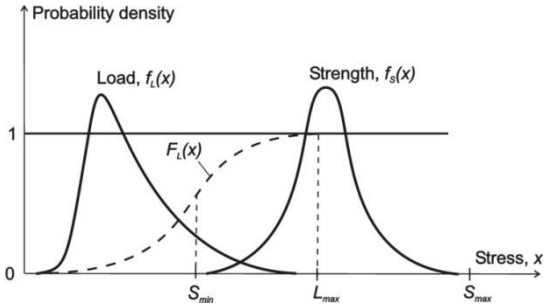

Figure 4. Deriving the reliability on demand, by integrating within the interval (*Smin, Lmax*) including only the upper tail of the load distribution and the lower tail of the strength distribution.

The integral in Equation (3) can also be presented as

$$p_f = \int_{S\min}^{L\max} [1 - F_L(x)] f_S(x)\, dx + \int_{L\max}^{S\max} [1 - F_L(x)] f_S(x)\, dx \tag{4}$$

For $x > L\max$, $F_L(x) \approx 1$ holds for the cumulative distribution of the load (Fig.4) and the second integral in Equation (4) becomes zero ($\int_{L\max}^{S\max} [1 - F_L(x)] f_S(x)\, dx \approx 0$). Consequently, the probability of failure becomes

$$p_f = \int_{S\min}^{L\max} [1 - F_L(x)] f_S(x)\, dx \tag{5}$$

Finally, for the reliability on demand, we get

$$R = 1 - \int_{S\min}^{L\max} [1 - F_L(x)] f_S(x)\, dx \tag{6}$$

The reliability integral (6) which quantifies the relative separation of the load distribution and the strength distribution has a clear advantage: To quantify the separation of the load and strength, data covering the lower tail of the load distribution and the upper tail of the strength distribution are no longer necessary.

The variance of the load distribution or the strength distribution can be reduced by shortening the lower tail of the load distribution without altering its upper tail and by reducing the upper tail of the strength distribution without altering its lower tail. However, the result from equation (6) shows that reducing the standard deviation of the load distribution and the strength distribution in the described fashion will have no impact on the risk of failure. Consequently, reducing the variances of the interacting random factors does not necessarily increase their relative separation and reduce risk.

Separation of the load distribution and the strength distribution can for example, be done by focusing on the strength distribution only (Fig.5c). The well-known *burn-in* operation (O'Connor, 2002) essentially increases the relative separation of the lower tail of the strength distribution and the load distribution thereby reducing the load-strength interference and reducing the risk of failure.

The relative separation of the load distribution and the strength distribution can also be increased by introducing deliberate weak links or stress limiters.

3.2 Separation by designing deliberate weak links

The consequences from failure and the risk of failure can be decreased if potential failures are channelled into deliberately designed weak links. Should the unfavourable conditions occur, the weak links are the ones to fail and protect the structure or component. In this way, the conditional losses are limited.

In case of M mutually exclusive failure modes, the expected conditional loss $\overline{C}_f$ (given that failure has occurred) is given by

$$\overline{C}_f = p_{1|f}\,\overline{C}_{1|f} + p_{2|f}\overline{C}_{2|f} + ... + p_{M|f}\overline{C}_{M|f} \tag{7}$$

where $\overline{C}_{k|f}$ is the expected conditional loss associated with the *k-th* failure mode ($k=1,2,...,M$) and $p_{k|f}$ is the conditional probability that given failure, it is the *k-th* failure mode that has initiated it ($\sum_{k=1}^{M} p_{k|f} = 1$). Indeed, the loss from failure $\overline{C}_f$ can take its values in M distinct, mutually exclusive ways: if the first failure mode materializes and the loss of failure is equal to the loss of failure $\overline{C}_{1|f}$ associated with the first failure mode; if the second failure mode materializes and the loss of failure is equal to the loss of failure $\overline{C}_{2|f}$ associated with the second failure mode;...;and finally, if the Mth failure mode materializes and the loss of failure is equal to the loss of failure $\overline{C}_{M|f}$ associated with the Mth failure mode. Considering the conditional probabilities $p_{k|f}$ ($k=1,2,...,M$) associated with the failure modes (given that failure has occurred) and applying the total probability theorem yields equation (7). For M failure modes characterised by constant failure rates λ_1, λ_2,..., λ_M, it can be shown that the conditional probabilities $p_{k|f}$ are given by

$$p_{k|f} = \frac{\lambda_k}{\lambda_1 + \lambda_2 + ... + \lambda_M},\ k = 1,2,...,M \tag{8}$$

Without restricting generality, suppose that a deliberate weak link has been designed and $\overline{C}_{M+1|f}$ is the conditional loss associated with the failure of the deliberate weak link. The loss $\overline{C}_{M+1|f}$ is the smallest among all conditional losses $\overline{C}_{k|f}$, ($\overline{C}_{M+1|f} << \overline{C}_{k|f}$, $k=1,2,...,M$). Suppose that the conditional probability $p_{M+1|f}$ of the deliberately built weak link has been made to be significantly larger than any other conditional probability ($p_{M+1|f} >> p_{k|f}$,

k=1,2,...,M). Given failure, it is now highly likely that the deliberate weak link has caused it (failure mode M+1, associated with the smallest conditional losses $\overline{C}_{M+1|f}$). As a result, it is highly likely that the conditional loss $\overline{C}_f$ given failure will be equal to the conditional loss of the deliberate weak link and the consequences of failure will be limited.

The deliberately introduced weak links essentially separate expensive components/systems from excessive loading stress. If the loading stress increases to a particular value p^* of the parameter (Fig.5a), a deliberate weak link fails and prevents a further increase of the loading stress (Fig.5a) on the component/system. As a result, components/systems are protected from overloading.

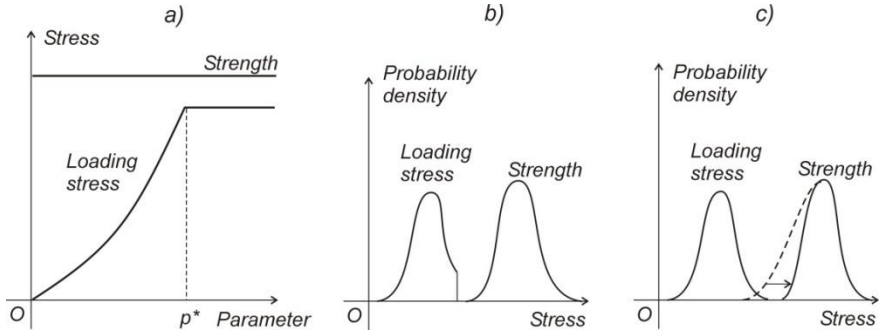

Figure 5. a) Separation of the load and strength by including a deliberate weak link; b) separation of the load and strength by a stress limiter; c) separation of the load and strength by a burn-in operation.

A shear pin in a mechanical coupling, for example, transmits torque up to a specified level p^*, beyond which the shear pin fails and disconnects the driving shaft from the mechanical device (Fig.5a). The shear pin acts as a deliberate weak link. As a result, the critical torque resistance of the mechanical device cannot be reached and the mechanical device is separated from overload. Rupture discs on pressure vessels and blow out panels built in buildings are typical examples of deliberate weak links. They are sacrificial components which separate from excessive pressure.

Electric fuses are also a common example of deliberate weak links separating electronic circuits from excessive levels of current.

Crash cones used in race cars separate against excessive deceleration during impact by deforming during an impact thereby increasing the time Δt during which the deceleration force F is present. For a given change $m\Delta v$ in the moment of the impacting car (m is the mass of the car and Δv is the change of its velocity) from the well-known relationship $F\Delta t = m\Delta v$, an increase of the impact (deceleration) time Δt results in a proportional reduction of the magnitude F of the average impact force.

Sacrificial anodes can also be considered to be deliberate weak links separating components (underground pipes, underwater installations, ship hulls) from excessive corrosion.

Deliberate weak links can also separate against excessive wear. For example, cheap rubber segments bolted on top of a conveyor act as deliberate weak links. They take all the excessive wear and their failure is followed by a replacement of a cheap rubber segment rather than by a replacement of an expensive conveyor belt.

In separating from excessive levels of the loading stress, the stress limiters can also be considered as instances of deliberate weak links (Fig.5b). A common example of a stress limiter is the anti-surge protector preventing voltage from reaching dangerous levels that could damage the electronic equipment. The safety pressure valve, activated when pressure reaches a critical level, is another common example of a stress limiter separating the loading stress from the strength of the material (Fig.5b).

While deliberate weak links are designed to fail and separate from excessive loading stress, stress limiters separate from excessive loading stress without suffering failure. The specially designed shoulder on the screw in Figure 6 is an example of a stress limiter. The shoulder prevents over-tightening of the screw and damaging the plastic component. The magnitude of the loading stress on the plastic part has been limited without any failure occurrence.

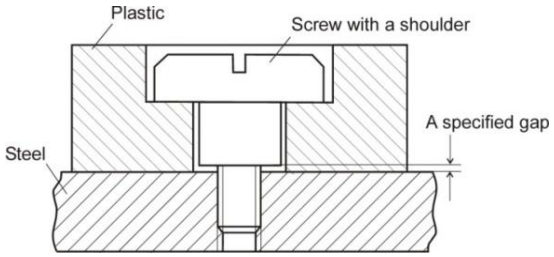

Figure 6. An example of a stress limiter: eliminating the risk of damaging the plastic part by a special design of the screw.

The friction clutches is another example of a stress limiter, that has been specifically designed to slip during a torque overload. The friction clutch acts as a deliberate weak link which does not suffer failure.

Expansion offsets are a common generic solution for separating from excessive temperature and developing excessive thermal stresses by accommodating thermal expansion/contraction. The shape and the size of the offsets depend on the amount of thermal expansion/contraction that needs to be accommodated.

In summary, the considered deliberate weak links can be classified as components/devices separating from excessive levels of: (i) torque, (ii) shear, (iii) tension, (iv) compression, (v) pressure, (vi) current, (vii) deceleration, (viii) temperature and (ix) wear.

3.3 Separation on loading stresses

Separation on loading stresses is present in cases where the design has been made in such a way that part of the structural elements are experiencing only tensile loading stresses while the rest of the components are experiencing only compressive stresses. Such loading, for example, is present in trusses, where the members experience only tension or compression. Bending and shear are completely eliminated from the truss structures. Eliminating bending and separating the loading stresses into tensile and compressive only, provides the possibility to reduce the cross sections of the tensile elements and reduce the weight of the design without compromising reliability. The cross sections of the tensile elements can be reduced because the buckling failure mode is absent for tensile elements. At the same time, for the same weight, a separation on loading stresses results in enhanced reliability.

Tensegrity systems (Wang, 2004) are composite forms of structures where a set of discontinuous compression components interacts with a set of continuous tensile components to define a stable space structure. Tensegrity systems are a good example of separation on loading stresses where most of the components are loaded in tension and only few components experience compression. The cross sections of the components experiencing tension can be reduced which results in an overall reduction of the weight of the structure or in an increased load-carrying capacity at a specified weight.

3.4 Separation on temperature

Temperature is a common risk-critical parameter. A typical example of separation on temperature is present in thermostats switching on and off heating circuits when temperature reaches a critical level.

Separation on operating temperature is also present when the aggregate state of the material is changed deliberately to reduce the hazard potential and risk. Thus, freezing a volatile and flammable substance during transportation eliminates the risk of spillage and explosion during a road accident.

Materials selection can also be made to provide a separation on temperature. During induction heating of a steel part for example, overheating beyond the Curie temperature T_{cr} of the steel, from which the part is made, is impossible. At a temperature T_{cr}, from ferromagnetic (temperatures $T < T_{cr}$), the steel becomes diamagnetic (temperatures $T > T_{cr}$). Because the magnetic properties of the steel change at $T = T_{cr}$, induction heating of the steel beyond $T = T_{cr}$ is no longer possible therefore overheating beyond $T = T_{cr}$ is not possible.

Separation on operating temperature has been used to protect control equipment in rockets from overheating. Rockets are placed in a foam shell which evaporates after the rocket launching.

Separation on operating temperature is present in thermistors which are essentially resistors whose resistance changes significantly with changing temperature. In thermistors with negative temperature coefficient (NTC thermistors) the resistance coefficient decreases significantly with increasing temperature. This has important applications with protective devices (e.g. cooling fans, protective circuits) which are activated when temperature increases to dangerous levels.

3.4.1 Separation on temperature to achieve compressive residual stress at the surface

Separation on temperature can be used to achieve a particular state characterised by increased reliability. For example, compressive residual stresses after quenching of cylindrical rods is beneficial for the fatigue resistance because of the crack closure effect of the compressive stress field. To achieve compressive residual stresses at the surface for steel cylindrical rods, separation on temperature can be used which results in a different behaviour of the quenching medium.

During quenching of a steel rod there are two counteracting principal factors controlling the formation of the residual stress at the surface (Todinov, 1999). The first factor is the thermal strain formed during the thermal contraction of the quenched rods which results in a compressive residual stress at the surface. The second factor is the thermal strain formed during the martensitic phase transformation at low temperatures which results in the formation of tensile residual stresses at the surface. If during quenching of a cylindrical steel specimen, the net plastic strain generated in the thermal contraction region is greater than that generated in the transformation region, the residual stress at surface is compressive and vice versa. This conclusion is supported by the continuum model proposed in (Todinov, 1998) and the experiments presented in (Todinov, 1999).

Consequently, the initial phase of quenching starting from high temperatures should be conducted in a quench medium characterised by a high heat transfer coefficient. This will create large thermal plastic strains which will result in large compressive residual stresses at the surface. At low temperatures, the quenching medium should be characterised by a low heat transfer coefficient. This will create small thermal gradients in the martensitic temperature interval and the transformation plastic strains which promote tensile residual stresses at the surface will be minimal. The result will be compressive residual stresses at the surface.

A separation on temperature can also be achieved by quenching sequentially in two separate quench media. The component is initially quenched in the first quench medium, characterised by a high heat transfer coefficient at high temperatures, for creating large thermal strains. Subsequently, the component is transferred in a second quench medium,

characterized by a low heat transfer coefficient in the martensitic transformation region, to reduce the thermal gradients and achieve small plastic strains. The heat-treatment process will result in compressive stresses at the surface (Todinov, 1999).

3.5 Separation on size

A common separation on size is present in strainers whose function is to stop debris from entering a system. The strainer separates on the parameter 'size' because it retains debris beyond a critical size capable of clogging the system whilst letting through debris with a smaller size. The size of the debris is a risk-critical parameter because clogging of a filter by debris for example, results in a low pressure after the filter, which normally constitutes failure.

4. SEPARATION ASSURING DISTINCT FUNCTIONS, PROPERTIES OR BEHAVIOUR FOR DISTINCT COMPONENTS/PARTS

4.1 Separation of functions

4.1.1 Separation of functions to reduce vulnerability to a single failure

Separation of functions consists of assigning different functions to different parts of a component/system.

In general, it is difficult to optimise a single component carrying many functions with regard to every single function. Separating critical functions to different parts/components is often the key to improving reliability and reducing risk.

The separation of functions will be illustrated with the design of flexible pipes carrying hydrocarbons (Fig.7) under water.

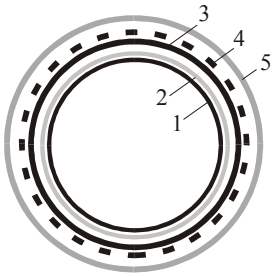

Figure 7. Separation of functions in a flexible pipe transporting hydrocarbons under water.

The pipe is composed of a stainless steel *internal carcass* (1); an *internal sheath* (2) which is extruded polymer barrier; *a pressure armour* (3) – a carbon-steel interlocked circumferential layer and *a tensile armour* (4) – helically wound carbon-steel layers for axial strength. Externally, the pipe is protected by *extruded sheath* (5). The internal carcass (1) prevents collapse of the internal sheath (2) due to the hydrostatic pressure of the water and also ensures mechanical protection. The internal sheath (2) ensures the integrity of the transported fluid while the function of the pressure armour and the tensile armour is to provide resistance against radial and tensile loads. The external sheath (5) is a mechanical barrier shielding the pipe's internal structural elements from the marine environment.

As a result, the different parts of the flexible pipe carry different functions: to protect against external corrosion, to resist tensile loads, to resist radial loads resulting from internal pressure, to make the pipe leak-proof and to prevent collapse due to external pressure. It is difficult to optimise a homogeneous pipe with respect to each of these functions. The separation of functions to different parts permits the optimisation of each part with respect to the single function it carries. The result is increased overall reliability of the pipe.

Seemingly, the separation of functions among separate parts contradicts the principle of improving reliability by reducing the number of parts. Indeed, the separation of functions

works towards increasing the complexity of the system instead of reducing it. This apparent contradiction can be resolved by considering that a loss of a part carrying a separate function does not mean a loss of all system functions. Furthermore, the separation of functions to different parts decreases the vulnerability of the component/system because the parts with the separated functions often provide *k-out-of-n redundancy* (Ramakumar, 1993) and the system is operational if k parts carrying separate functions are lost.

This particular property of the separation of functions can be illustrated particularly well by the case where the same function has been separated to multiple components to reduce risk, which is essentially a process of segmentation. Segmentation is the act of separating an entity (assembly, system, process, task, time, etc.) into a number of distinct parts. In this respect, the method of segmentation used for risk reduction introduced in (Todinov, 2015) can be interpreted as a special case of separation where the same function has been separated into different parts.

By separating a function into separate parts carrying the same function, separation replaces a single critical failure occurring at a macro-level with non-critical failures occurring at a micro-level. Suppose that a single resource is used for servicing a critical manufacturing process. Failure of the resource entails an interruption of the manufacturing process and causes severe delays and lost production. If the resource is separated into multiple smaller resources, failure of a single resource will not be catastrophic and will not entail a shutdown of the manufacturing process. Separation reduced risk by replacing a single critical failure with non-critical failures.

Analytical justification of this technique will be given with a single bolt used for fixing a critical part. Suppose that the only failure mode is 'manufacturing defect in the bolt causing a fast fatigue failure' and the reliability of the bolt, associated with one year of continuous operation, is $r=0.75$. Suppose that the function of the bolt has been separated into four smaller bolts made of the same material as the initial single bolt and suffering the same failure mode. The four bolts give the same total clamping force as the single bolt and each of the bolts is characterised by a lower reliability $s=0.70$, associated with the operational time interval of one year. Any two of the smaller bolts are sufficient to perform the function of fixing reliably the critical part.

The probability that the separated bolt assembly will survive 1 year of continuous operation without failure to support the critical part is now equal to the probability that at least two (two, three or all four) bolts will survive the operational interval of one year. If the bolts fail independently from one another, this probability is given by

$$R=s^4+\frac{4!}{3!\times(4-3)!}s^3(1-s)^1+\frac{4!}{2!(4-2)!}s^2(1-s)^2=0.916$$

As a result of the separation, the reliability of the segmented bolt assembly increased significantly compared to the single bolt, despite that the reliability of each of the smaller bolts was inferior to the reliability of the single initial bolt.

The separation of a single function greatly enhances the reliability of devices obtaining a signal from a sensor and triggering a particular action/alarm if the signal indicates a dangerous concentration of a particular chemical, dangerous magnitude of a force, torque, pressure, temperature, humidity, etc. Separation of a single sensor into multiple sensors, even with inferior reliability, makes the system less vulnerable to a single malfunction of a sensor or even to simultaneous failures of several sensors. Separation of a function into smaller-size components carrying the same function also reduces inertia forces and the response of the device to driving forces of small magnitudes. This is used in twin turbochargers where, to improve the response of the turbine to exhaust gases with small energy, two smaller turbochargers are used instead of a single large turbocharger.

Suppose that a component is responsible for three functions. Failure of the component will then cause a loss of all three functions. If functions are assigned to different components such that each component performs a single function, the loss of a component will cause a loss of a single function only. Furthermore, separating the functions performed by a single component relieves the stress on the component. Sensitivity to a single component failure is often present in cases where a particular component is overloaded with too many functions and demands. The component is 'over-stretched' and its strength can be easily exceeded if combined multiple demands are present. Additional functions required from a component also increase the number of different failure modes of the component. With increasing the number of failure modes, the overall hazard rate increases and the reliability decreases.

A common design error of this type, which has caused high-impact failures is combining the critical functions of load carrying and sealing in the design of a joint. Such was the case with the space shuttle *Challenger* booster's O-ring, which was simultaneously sealing the section of the assembly and taking the pressure of combustion (Ullman, 2003).

Separation of functions can reduce risk even in the case where the parts carrying the separate functions are logically arranged in series. Indeed, if λ_1, λ_2,..., λ_n are the hazard rates characterising the different failure modes (tensile failure mode, torsion failure mode, buckling, leaking, etc.), the reliability R_0 of a component for which the functions have not been separated is: $R_0 = \exp\left(-t\sum_{i=1}^{n}\lambda_i\right)$, where t is the specified operation time interval. The separation of functions to distinct parts provides the opportunity to optimise each corresponding part (carrying a separate function) against the failure mode it resists. As a consequence, after the separation of functions, the hazard rates $\lambda_1^{'}$, $\lambda_2^{'}$,..., $\lambda_n^{'}$ characterising the different failure modes are reduced: $\lambda_1^{'} < \lambda_1$, $\lambda_2^{'} < \lambda_2$,..., $\lambda_n^{'} < \lambda_n$. Suppose that the number of failure modes has not been increased after the separation (only the number of parts has been increased).

The reliability R' of the system where the functions have been separated into different parts is then increased: $R_1 = \exp\left(-t\sum_{i=1}^{n}\lambda_i^{'}\right) > R_0 = \exp\left(-t\sum_{i=1}^{n}\lambda_i\right)$. As a result, given that no new failure modes are introduced, by separating the functions into separate parts logically arranged in series, despite the increase in the number of parts, the risk of failure is usually decreased because of the possibility for an optimization.

4.1.2 Separation of functions for a mutual compensation of deficiencies

The separation of functions can be provided for mutual compensation of deficiencies associated with the different components building a system. A typical example is the hybrid joint, combining an adhesive joint and mechanical fixing. There is a clear separation of functions: the adhesive part reduces the stress concentration along the joint while the mechanical fixing increases the peel resistance of the adhesive joint and its stiffness.

Separation of functions is often present in the design of complex alloys where some of the microstructural constituents provide wear resistance, while other constituents provide toughness (resistance to crack propagation).

Reinforced concrete used in the construction industry is a good example of separation of functions. Concrete is a material with good compressive strength but small tensile strength. Accordingly, the concrete is reinforced with steel bars placed in areas loaded in tension where the concrete cannot resist tensile stresses.

4.1.3 Separation of functions to satisfy conflicting requirements

Mixture of several structural constituents carrying different functions can be used to create components satisfying conflicting requirements. Thus, to reduce the risk of injury to divers, the water must be dense to provide buoyancy and must not be dense to cushion the impact and prevent injury. This is achieved with water saturated with air bubbles. The water provides the necessary buoyancy while the air bubbles soften the water by decreasing the water density thereby mitigating the consequences from an impact.

4.1.4 Separation of functions to prevent unwanted interactions

Separation of functions (also known as 'separation of concerns') is a well-known principle of design of computer programs (Reade, 1989). A concern is a relatively simple, self-contained task, addressed by a programme section. Separation of functions in programming is achieved by encapsulating data and statements inside a section of code that has a well-defined interface. This results into a modular programme, consisting of procedures and functions. The encapsulation means that the variables defined into the encapsulated module (procedure or a function) remain only visible within the module and can be altered only within the module. Encapsulation avoids unwanted interactions between different pieces of code in the same programme. Avoiding unwanted interactions avoids the possibility of side effects and difficult to rectify bugs if a variable from a particular section of code is altered from another section of code. Furthermore, the encapsulated sections of code can be updated and tested independently, without having to alter code in the rest of the sections, which significantly decreases the possibility of introducing bugs. The encapsulated piece of code is essentially a black box with specified input and output, whose content can be independently developed and replaced without affecting the logic of the programme.

4.2 Separation of properties to counter the drawbacks associated with average properties

Separation of properties is necessary when the average property characterising a homogeneous state cannot provide the required reliability.

Consider a loaded steel component working in aggressive environment. Suppose that a compromise has been made by striking balance between the corrosion resistance of the steel and its cost. The result is a component which still corrodes but at a smaller rate. Instead, the surface in contact with the aggressive environment could be coated with a small quantity anti-corrosion coating with a very high corrosion resistance. To offset the overall cost, cheap steel, with a small corrosion resistance, can be selected for the rest of the steel part. As a result, the overall corrosion resistance is significantly increased, with little or no increase of the overall cost of the component.

This method works also in selecting materials to resist loading stresses. Selecting a homogeneous material with average value of the strength (resistance) cannot provide a sufficient resistance in the high-stress zones while in low-stress zones, the strength is insufficient to resist the load (Fig.8c). Consider the bracket in Fig.8a resisting a high-magnitude occasional overload with magnitude $F_{\max}$. The material of zone *A* must resist large tensile stresses while the material of zone *B* does not experience stresses of large magnitude. Commonly, the entire bracket is made of homogeneous steel, striking a compromise between tensile strength and cost. Instead, zone *A* which experiences high tensile stresses, could be made of expensive steel with high tensile strength. Zone *B* could be made of cheap steel, with low tensile strength, and the two parts could be joined to form the bracket. As a result, the reliability of the bracket will be increased significantly with a small or no increase of the cost. If the loading stress varies in the volume of the component, selecting a homogeneous material with average resistance cannot guarantee that the resistance curve will always be greater than

the curve representing the load (Fig.8c). The separation of properties guarantees that the resistance will be greater than the load and the required reliability will be guaranteed (Fig.8c). Offsetting the drawbacks of a homogeneous state by a separation of properties is very useful in cases where reliability is balanced against weight and cost. Improving reliability only locally, where it matters, saves resources and results in lightweight designs.

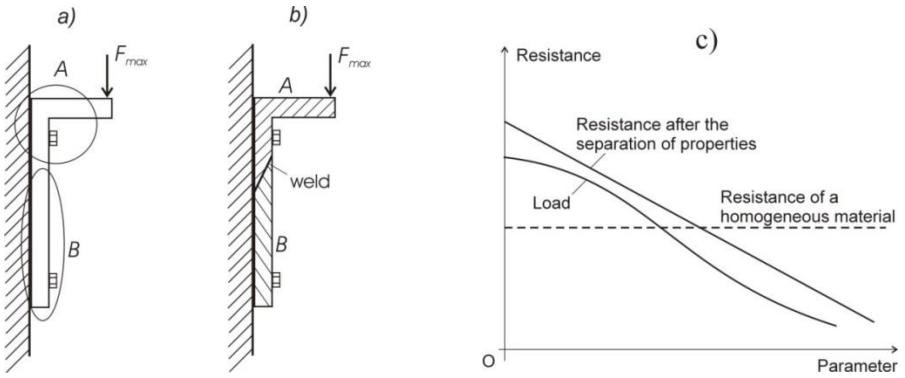

Figure 8. a,b) Separation of properties in a loaded bracket; c) Separation of the properties guarantees that the resistance will be greater than the load.

Separation of properties can be used to create a component/system satisfying conflicting requirements. Guaranteeing different properties at different places is the underlying principle behind coatings improving the wear resistance and corrosion resistance of surfaces subjected to wear and contact with aggressive environments.

This principle is often used in composite materials, combining structural constituents with different properties in different directions.

Consider a rod subjected to a combination of tension shock loading and intensive wear. The rod must be hard in order to resist the wear and at the same time it must be tough to resist the stress from the shock loading. In a compromise, with no separation of properties, a homogeneous material will be selected which has a satisfactory hardness to resist wear and a satisfactory toughness to resist shock loading. The result is an inferior solution which is neither optimised against wear nor against shock loading. These contradicting requirements can be simultaneously guaranteed if a separation of properties is implemented.

Consider a process of gas nitriding which saturates the surface layers of steel components with nitrides. The nitride coating is brittle and provides no resistance against the stresses from shock loading. However, the nitride coating is hard and can successfully resist wear. At the same time, gas nitriding provides resistance against fatigue, because it creates compressive residual stresses inducing crack closure and reducing the rate of fatigue crack propagation. The inner parts of the rod are not subjected to wear. Consequently, no special wear-resistant properties for these parts are necessary. These parts must remain tough to resist shock loading. A clear separation of properties is present, which provides the best service conditions for both - the surface and the inner parts of the rod.

In a related example, a gear must be hard, to endure large contact stresses and intensive wear and soft, to endure impacts. These conflicting requirements require conflicting material properties: the surface of the gear must be hard while the core must be soft. These conflicting properties can be guaranteed by the separation of properties achieved through *case hardening* (Kalpakjian and Schmid, 2001). This consists of a local induction heating of the surface layers followed by quenching. Case hardening improves the resistance of the surface to large contact stresses and wear, while leaving the core tough makes it resistant against impact loads.

In another example of separation of properties, imparting different resonance frequencies to connected components in a vibrating system dampens the amplitude of vibrations of the system and prolongs its life.

5. IMPROVING RELIABILITY AND REDUCING RISK BY DISTANCING RISK-CRITICAL FACTORS

5.1 Distancing risk-critical factors by a stochastic separation

Stochastic separation is present if minimal distances between risk-critical events is guaranteed with a specified probability.

Risk often depends on the non-overlapping of random events in a time interval. The competition of random demands for a particular resource/service on a finite time interval is a common example of risk and reliability controlled by the simultaneous presence of critical events. The appearance of a critical event engages the servicing resources and if a new critical event occurs during the service time of the first critical event, no servicing resources will be available for the second event.

Suppose that only a single repair resource is available for servicing failures on a power line. In the case of a power line failure, the repair resource will be engaged and if another failure occurs during the repair time associated with the first failure, no free repair resource will be available for recovering the power distribution system from the last failure. The delay in the second repair could lead to overloading of the power distribution system thereby inducing further failures.

There are cases where a very low probability of simultaneous presence of risk-critical critical events can be tolerated. Such is for example, the case of breakdowns of heating elements attached to different sections of a long subsea oil pipeline. Each failed heating element demands the intervention of a special repair vessel. If an intervention vessel is available and a breakdown of a heating element occurs, the repair vessel is engaged in the repair of the failed heating element. If during the repair, another breakdown of a heating element occurs, no free repair vessel will be available to service the new failure. As a result, the delay in recovering from the second failure will result in the formation of waxy deposits in the affected pipeline section which could block the flow through the pipeline. Blocking the flow by waxy deposits entails lengthy and expensive intervention involving cutting and replacing large sections of the pipeline.

Another category of failures controlled by the simultaneous presence of critical events is present when the simultaneous appearance of critical events increases the load on the system to a level which exceeds the system's strength. Commonly, the simultaneous appearance of random demands whose number is greater than the capacity of the system leads to overloading which often has catastrophic consequences for the system.

Yet another category of failures controlled by the simultaneous presence of risk-critical events relates to the case where *some of the critical events weaken/degrade the system's strength while some of the events increase the load on the system*. Thus, the simultaneous presence of the critical event 'repair of a failed power line' and the critical event 'sudden increase in power consumption' often cause overloading and failure of other power lines and disruption of the power supply.

A very low probability of a simultaneous presence of random demands can be tolerated for example, in a situation where critically injured people demand a particular piece of life-saving equipment.

Suppose that the times of the risk-critical events follow a homogeneous Poisson process in the interval $(0,L)$ and each event has a duration 's'. In other words, the number of events in the time interval is a random variable. According to an equation discussed in (Todinov,

2005), the probability p_0 that there will be no clustering of two or more random events within a critical distance *s* is

$$p_0 = \exp(-\lambda L)\left(1+\lambda L+\frac{\lambda^2(L-s)^2}{2!}+\ldots+\frac{\lambda^r[L-(r-1)s]^r}{r!}\right), \quad (9)$$

where *r* denotes the maximum number of time gaps of length *s*, which can be accommodated into the finite time interval with length *L* ($r=[L/s]+1$), where $[L/s]$ is the greatest integer which does not exceed the ratio *L/s*).

Consider the important practical problem of requests arriving randomly in time to a source that can service only a single request at a time. Assume for the sake of simplicity that time *s* is needed to service each random request. The random requests could be related to using unique piece of equipment, using a particular resource (e.g. water vapour, electrical power, compressed air, etc.). The list can be continued.

Equation (9) can be used for setting reliability requirements to provide a stochastic time separation (avoiding overlapping of events) of duration at least *s*, with high probability, in the common case where the random events follow a homogeneous Poisson process. For any specified time of demand *s* and a minimum probability p_0 with which the separation intervals of length at least *s* must exist, solving the equation with respect to λ yields an upper bound λ^* (an envelope) for the number density of the random events. The envelope guarantees with a minimum probability p_0 that whenever for the number density λ of events, $\lambda \le \lambda^*$ is fulfilled, the specified minimum separation of a distance at least *s* will exist between successive random demands,. In other words, with the specified probability p_0, there will be no unsatisfied demand.

In an illustrative example, the number density envelope of random demands will be determined which guarantees that the probability of unsatisfied demand will be below a specified level. A single source servicing random requests is available and each random request requires a minimum time interval of 0.5h to be serviced. The demands follow a homogeneous Poisson process in a finite time interval of 100h. If two or more demands follow within the critical service time interval of 0.5h, there will be unsatisfied demand. The maximum acceptable probability of unsatisfied demand has been specified to be $p_c = 0.1$. By solving equation (9) with respect to λ, where $p_0 = 1 - p_c = 0.9$, the upper bound $\lambda^* = 0.0467\,\mathrm{h}^{-1}$ of the number density of demands can be obtained. Whenever for the number density λ of demands $\lambda \le \lambda^* = 0.0467\ \mathrm{h}^{-1}$ is fulfilled, the probability of unsatisfied demand is smaller than 0.1. Monte Carlo simulations (one million trials) of a homogeneous Poisson process with density $\lambda^* = 0.0467$ yielded 0.1 for the probability of clustering two or more random demands within the critical interval of 0.5 h, which confirms the result from solving equation (9). Thus, for an expected number of 5 demands in 100h, the probability of unsatisfied demand is substantial (≈ 0.1). Even for the mean number density of 2 demands in 100h, the calculation from equation (9) shows that there is still approximately 2% chance of unsatisfied demand.

Figure 9 gives the dependence of the probability of unsatisfied demand for a single source and time of $s = 1\mathrm{h}$ for servicing a single random demand. The operating time interval is $a = 100\,\mathrm{h}$. The probability of unsatisfied demand has been plotted for different values of the number density of the random demands.

For a mean number of 14 demands per 100h, there is already 80% probability of unsatisfied demand. Clearly, the probability of unsatisfied demand is substantial and should always be taken into consideration in risk assessments.

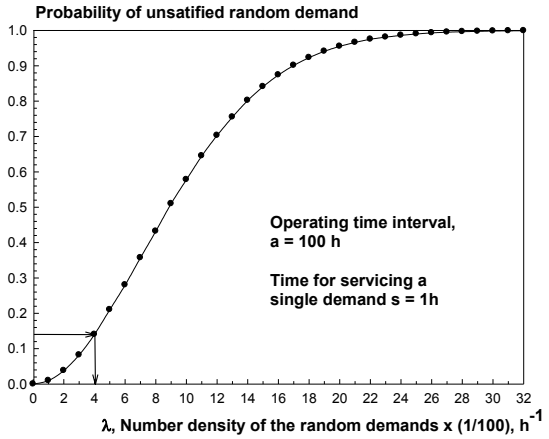

Figure 9. Probability of unsatisfied demand on a finite operational time interval of 100h. The random demands follow a homogeneous Poisson process and each random demand requires 1h service time.

Plots similar to the one in Figure 9 can be used for setting reliability requirements which provide the required degree of stochastic separation. For a specified maximum acceptable probability of unsatisfied demand, for example 14%, the number density envelope of $\lambda^* = 0.04 \ \ h^{-1}$ can be determined (see the arrows in Figure 9). This envelope guarantees that whenever the number density of the demands does not exceed $\lambda^* = 0.04$ ($\lambda \le \lambda^*$), the probability of unsatisfied demand will not exceed the critical level of 14%.

This example demonstrates the importance of setting reliability requirements not only to provide a stochastic separation and minimise the probability of unsatisfied demand below a maximum acceptable level but also to provide an optimal balance between risk and cost.

The problem of stochastic separation appears frequently in critical situations. Consider an example of emergency demands for a medical specialist arriving from patients in a critical state requiring immediate medical attention (e.g. demands from patients with toxic poisoning), In this case the obtained hazard rate envelope can be used to determine the maximum number of such patients that could be seen by a single medical specialist so that the probability of unattended emergency demands remains below a maximum acceptable level. The consequences of an unattended emergency demand can be fatal and to keep the risk low, the tolerable probability of unsatisfied demand should be very low.

If the average number density of the demands from a single patient in a critical state is λ_0, the total number density of the demands characterising all n critical patients is $\lambda = n \times \lambda_0$. Determining the maximum acceptable demand rate λ^* which guarantees with a specified probability that there will be no patient demand while the medical specialist is servicing another patient can be determined by using a method similar to the method illustrated in Figure 9. Dividing the maximum acceptable demand rate λ^* to λ_0 yields the maximum acceptable number $n*$ of patients that can be seen by a single medical specialist:

$$n^* = \lambda^* / \lambda_0 \tag{10}$$

Equation (9) is relevant to a wide class of reliability and risk problems. It can also be applied to determine the probability of clustering of two or more flaws in fibres or wires, within a critical distance *s*, given that the flaw number density is λ. It is assumed that the locations of the flaws follow a homogeneous Poisson process in the finite length *L*. Solving Equation (9) with respect to the flaw number density λ defines an upper bound for the flaw number density which guarantees with a specified minimum probability a space separation (no clustering) of flaws within a small critical distance *s*. This is important in cases where the probability of failure during loading is strongly correlated with the probability of clustering of

flaws. Solving equation (9) with respect to the flaw number density λ in fact specifies requirements regarding the maximum acceptable flaw content in the material in order to provide stochastic separation and reduce the probability of failure caused by the clustering of flaws.

5.2 Deterministic time separation by scheduling

Separation in time is required by processes, objects, entities whose simultaneous presence is associated with increased risk of failure. A common example of time separation of risk-critical factors is the traffic lights, preventing collision between intersecting flows of traffic and flows of pedestrians.

The time separation by scheduling enforces consistent time spacing between hazardous events. The time separation is used in air traffic control, where it enforces consistent time spacing between arriving aircrafts, on the basis of real-time information about the weather, headwinds, altitude and speed. The time separation guarantees a sufficient runway approach capacity which keeps the risk of accidents low.

Time separation based on scheduling can be used to create a process combining conflicting requirements. During welding operations for example, the voltage must be high in order to initiate the welding arc and low in order to produce a fault-free weld. These conflicting requirements can be attained by time separation where a very short period of high voltage is followed by a longer period of low voltage. In this way, the conflicting requirements are satisfied.

Time separation can be done by changing a process from a non-periodic to a periodic. In this way, two incompatible risk-critical factors can be introduced simultaneously by transforming their action from continuous to periodic and inserting the action of one of the factors in the pauses of the other factor.

5.3 Time and space separation by using interlocks

Preventing the simultaneous occurrence of two events can be done by using interlock devices. Consider a device working in forward and reverse mode (Fig.10a). The simultaneous pressing of the forward (F) push-button and reverse (R) push-button causes failure of the operated electro-mechanical device and must be prevented. The time separation can be easily done by two normally closed contacts *f* and *r*. The normally closed contact *f* opens when the forward push-button *F* is activated (Fig.10b) while the normally closed contact *r* opens when reverse push-button *R* is activated (Fig.10c). While the forward push button is activated, the normally closed contact *f* is open which eliminates the danger of accidentally or deliberately activating the reverse push-button R (Fig.10b). Similarly, while the reverse push button *R* is activated, the normally closed contact *r* is open which eliminates the danger of accidentally or deliberately activating the forward push-button *F* (Fig.10c). The dangerous simultaneous occurrence of the two actions has been excluded by a time interlock.

The space separation can also be based on interlocks. A common example of space separation based on interlocks is the railway signalling. It is impossible to display a signal to a train to proceed along a particular route unless the route is safe.

A common example of a space interlock is the presence-sensing safeguard interlocks which stop the operation of hazard equipment if a person is detected in a location where injury can occur. The presence-sensing system could be based on laser beams, light or infra-red beams. Beams of light forming a curtain are generated and if any of the beams is blocked by a person moving towards the hazard equipment, a control circuit switches off the power to the hazard equipment.

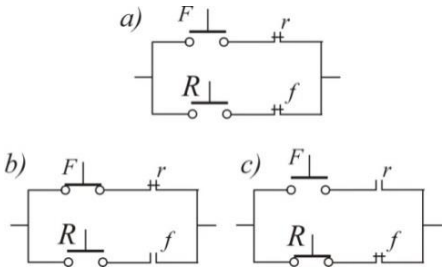

Figure 10. Time separation based on interlocks

An example of a simple mechanical space separation interlock is the interruption of the power supply from removing the protective shield of rotating machinery. As a result, it is impossible to operate the machine without positioning first the protective shield in place, which reduces the likelihood of injury.

A more sophisticated version of the mechanical space separation interlock is the trapped-key interlock. In one of the possible implementations, the access for repair to the hazard equipment is through a door operated by a key which is held trapped on the door until the door is firmly closed again by operating the key. While opening the door, a switch is operated interrupting the power supply. The hazard equipment cannot be re-energized until the door is closed and the key released. Releasing the key essentially guarantees that the hazard equipment has been made safe.

5.4 Separation based on barriers and interpretation

Deterministic space separation based on barriers has a wide application. Reliable operation often depends on critical properties, events or factors not being present in the same space region. Separating people from hazards has always been an important damage-reduction measure, if the control over hazards is lost. Separation in space is present in cases where processes, objects, entities whose interaction is associated with risk, are physically separated.

A familiar example of deterministic space separation based on barriers is the isolation of intersecting flows of traffic and flows of pedestrians at different levels to eliminate the risk of collisions and accidents. Limiting the spread of infection by quarantine measures involving isolating infected individuals is another common example of space separation creating a passive barrier. Separating hazardous sources (e.g. fuel tanks) at sufficient distances from one another avoids the domino-effect of multiple explosions and reduces significantly the amount of damage in the case of fire.

Deterministic space separation based on barriers is often used to separate sources of hazards and targets. Deterministic space separation is the essence of the safety practice of building residential areas beyond the radius of harmful influence of toxic substances from chemical plants, compost production facilities, fuel depots, etc. Separating people from hazards is an important damage-reduction measure if the control over hazards is lost. Physical barriers provide passive protection against the spread of fire, radiation, toxic substances or dangerous operating conditions. A blast wall for example, guards against the effects of a blast wave. Increasing the distance between sources of hazards and targets, minimises damage in case of an accident.

Examples of passive barriers are the safeguards protecting workers from flying fragments caused by the disintegration of parts rotating at a high speed; the protective shields around nuclear reactors or containers with radioactive waste; the fireproof partitioning; the double hulls in tankers preventing oil spillage if the integrity of the outer hull is compromised, etc.

The boundaries introduced by the separation often help reduce the damage escalation and the consequences given that failure has occurred. Thus, dividing a pipe into many separate sealed segments limits the damage from a propagating crack within a single segment only,

which reduces significantly the consequences from failure. Separation of the corridors in a building, with fireproof doors is used to delay the spread of fire.

Separation of different parts of computer networks has clear security benefits. For a separated section of a computer network, accessing a computer in one part of the network does not automatically give the attacker an easy access to other parts of the network. Such network design can significantly slow down the rate at which an attacker moves towards the valuable service and provides more opportunities for a successful detection. In addition, separating each segment by firewalls makes accessing the valuable service much more difficult because numerous security walls must be breached before an access can be gained. The result is a reduced likelihood of unauthorised access.

Deterministic separation can also be achieved through interpretation. Separation by interpretation includes acts of interpretation and compliance in order to perform its function. Typical examples are the road signs, the reflective studs on the road separating traffic, various warnings, cautions, prohibitions, etc.

The speed limit in a built region for example, separates a hazard (the car) from the target (pedestrian, structure) because when interpreted and obeyed, it provides more time for the driver to react and avoid an accident. The obeyed speed limit also provides separation from severe consequences given that an accident has occurred because the impulse of a car with speed below the speed limit is not sufficient to inflict a fatal damage to a pedestrian. The separation by interpretation is low-cost, easy to implement but unreliable.

5.5 Logical separation

Deterministic logical separation is present in cases where it is *logically impossible* for a dangerous operation to occur at a given point in time or at a given space location. Deterministic logical separation is also in place if it is logically impossible for two or more objects to be in a dangerous proximity at a given location or at a given time.

In logical separation, no barriers of any kind are set between the different parts of the system yet separation is present. The dangerous proximity of hazards and triggers and hazards and targets is made to be logically impossible.

Consider the safety problem related to preventing the hand of an operator from being in the cutting area of a guillotine. If the guillotine can be activated only by a simultaneous pressure on two separate knobs/handles which engage both hands of the operator, it is logically impossible for the operator's hand to accidentally reside in the cutting area, at any time. The operator's hands have been separated from the cutting area through the logic of the guillotine activation. This is an example of a logical separation reducing the risk of an accident.

Suppose that a dangerous action can occur during a measurement of the residual stresses on the surface of a specimen, by *X*-ray equipment (Fig.11). For example, the dangerous action can occur if a person appears in Room *A* and switches on the *X*-ray control panel while the operator is still positioning the specimen under the *X*-ray head in room *B* (Fig.11).

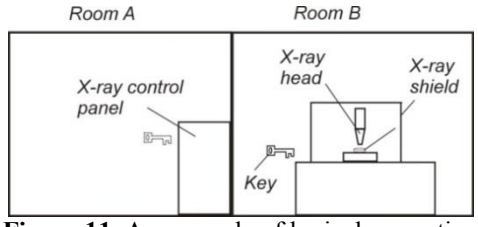

Figure 11. An example of logical separation

The dangerous action can be prevented from occurring by a logical separation. The design of the *X*-ray equipment can be made in such a way that the only way to lift the *X*-ray shield in order to position the specimen and the only way to operate the *X*-ray control panel in room *A*

is by using the same unique key. Switching on the *X*-ray control panel and adjusting the specimen under the X-ray shield would then require that the same object (the key) to be in two different places, at the same time, which is impossible. The safety risk has been eliminated by a logical separation.
Logical separation leads to low-cost yet very efficient designs eliminating safety risks. It is a simple yet underutilized generic tool for improving safety and reducing risk.

5.6 Separation in geometry
Separation in geometry is present when different parts of an object or assembly have different geometry to provide optimal conditions maximizing reliability and minimizing the risk of failure. A common example of separation in geometry is shaping concave one of the contact objects in order to reduce the contact stress and the likelihood of failure.
Separation in geometry can be used to improve the load resistance of designs, for a specified volume of material.

Consider the cantilever beam with length L and uniform rectangular cross section with thickness t and width b (Fig.12a) which has been overloaded by a force with magnitude P. Suppose that the tensile strength of the material is σ_s. According to the theory of elasticity, the maximum tensile stress σ_t acting on the beam, is at the cantilever support and is given by

$$\sigma_t = \frac{6PL}{bt^2} \tag{11}$$

Consequently, the maximum magnitude of the overloading force which the beam 'a' (in Fig.12a) can sustain is

$$P_{\max,a} < \frac{\sigma_s bt^2}{6L} \tag{12}$$

Suppose that the beam has been tapered in the way shown in Fig.12b, such that the right end of the beam has thickness $0.5t$ and the left end has thickness $1.5t$. If the width b and the length L of the initial beam remain unchanged, the volume of material $V_b = \frac{(0.5t+1.5t)}{2} L \times b = tLb$ used for beam 'b' will be equal to the volume of material $V_a = tLb$ used for beam 'a'. The maximum overload stress which beam 'b' can sustain is now 2.25 times bigger than the maximum overload stress characterising beam 'a':

$$P_{\max,a} < \frac{\sigma_s b(1.5t)^2}{6L} = 2.25\frac{\sigma_s bt^2}{6L} \tag{13}$$

The load-carrying capacity of the beam has been increased and the risk of failure has been reduced by a separation in geometry.

Separation in geometry can, for example, increase the resistance to torsion and bending with no increase in the amount of material used. The maximal torque which a circular component with radius r and cross-sectional area $A = \pi r^2$ can resist (Fig.12c) is given by (Gere and Timoshenko, 1999)

$$T_0 = \frac{\tau_s I_0}{r} \tag{14}$$

where $I_0 = \frac{\pi r^4}{2}$ is the central moment of area of the circular cross section and τ_s is the shear strength of the material.

Now suppose that by separation in geometry, a hollow circular cross section has been created (Fig.12d), with outer radius $r_1 = 1.5r$ and inner radius $r_0 = (\sqrt{5}/2)r$. The cross

sectional area of the new section (Fig.12d) is exactly equal to the cross-sectional area of the initial solid circular section (Fig.12c) ($\pi r_1^2 - \pi r_0^2 = \pi(1.5r)^2 - \pi(\sqrt{5}/2r)^2 = \pi r^2$).

The maximal torque which the hollow circular component can resist is given by

$$T_1 = \frac{\tau_s I_1}{r_1} \tag{15}$$

where $I_1 = \frac{\pi r_1^4 - \pi r_0^4}{2} = \frac{\pi(r_1^2 - r_0^2)(r_1^2 + r_0^2)}{2} = \frac{\pi r^2 \times (7/2) r^2}{2} = \frac{7}{2} \times \frac{\pi r^4}{2} = (7/2) I_0$

As a result,

$$\frac{T_1}{T_0} = \frac{(7/2)r}{r_1} = \frac{(7/2)r}{1.5r} = 2.33 \tag{16}$$

As a result of the separation in geometry, the maximum torque the component can resist has been increased 2.33 times.

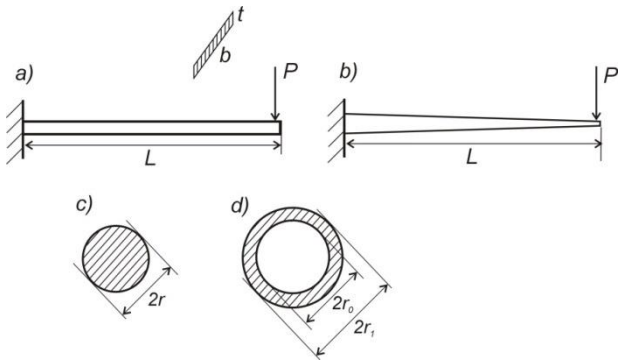

Figure 12. Separation in geometry to improve the load-carrying capacity of beams.

5.7 Separation to provide an independent operation and protection against a common cause

Separation can be used for improving reliability and reducing risk by providing an independent operation of components. If failure of one component makes the failure of another component more likely, the reliability of the system can be improved by making the two components operate independently from one another. Consider the operation of two devices where the second device is powered from the first device. Failure of the first device will cause a loss of power for the second device. The reliability of the system can be improved if the power supply to the devices is separated.

Next, consider a dual control system based on two control modules CM1 and CM2 controlling an electro-mechanical device *M*. To cut the cost, the two control modules share the same cable (Fig.13).

The two control channels in Fig.13a are not independent because failure of the cable connecting the electro-mechanical device will cause both control channels to fail which entails a loss of control over the electro-mechanical device M. Separating (decoupling) the control channels (Fig.13b) ensures the independent operation of the control channels and improves the reliability of the system. Failure of any of the connecting cables will not result in a loss of control over the electro-mechanical device M.

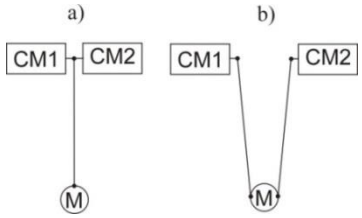

Figure 13. Improving the reliability of a control system by decoupling and ensuring independent operation

Decoupling of circuits is usually done by a decoupling capacitor and is often used in electronics when a portion of a circuit is prevented from being affected by fluctuations of the power supply due to switching occurring in another portion of the circuit.

Separation to block a common cause is present when a component or a group of components are distanced (insulated) from the action of a common cause, simultaneously affecting the performance of the components.

A common-cause failure is usually due to a single cause with multiple failure effects which are not consequences from one another (Billinton and Allan, 1992). A common cause reduces the reliability of a number of components simultaneously. The affected components are then more likely to fail, which reduces the overall system reliability.

Typical conditions promoting common cause failures are: common design faults, common manufacturing faults, common installation and assembly faults, common maintenance faults, shared environmental stresses by several components: for example high temperature, pressure, humidity, erosion, corrosion, vibration, radiation, dust, electromagnetic radiation, impacts and shocks. Common cause may also be due to: a common power supply, common communication channels, a common piece of software, etc. Thus, two programmable devices produced by different manufacturers, assembled and installed by different people can still suffer a common cause if the same faulty piece of software code has been installed in the devices.

Failure to account for common causes usually leads to optimistic reliability predictions - the actual reliability is smaller than the predicted. In many cases, the separation method helps in blocking out common causes thereby reducing the risk of failure. Separating the components at distances greater than the radius of influence of a common cause is an efficient way of reducing the risks of common-cause failures. Thus, separating large fuel containers at safe distances from one another prevents cascading explosions initiated by the explosion of one of the containers. Separating two or more communication centres at distances greater than the radius of destruction of a missile, increases the probability of survival of at least one of the centres. Multiple back-ups of the same vital piece of information kept in different places protects against the loss of information in case of fire, theft or sabotage.

Another implementation of this principle is the separation of vital components from a component whose failure could inflict damage. A typical example is separating the control lines at safe distances from the aeroplane jet engines. In case of engine explosion, the flight controls will still be operational which permits safe landing of the plane. Separating redundant components by insulating them from contact with an environment characterised by excessive dust, humidity, heath or vibrations, is also an efficient way of protecting against a common cause failure.

Providing maintenance of redundant components by separate operators reduces the likelihood of common cause failure due to faulty maintenance. Separating the physical principles on which redundant devices operate, provides diversity in design and is a very efficient way of blocking out a common cause and reducing common cause failures. The idea is to prevent several components from being affected by the same common cause. If two pumps (a main pump and an emergency pump) participate in cooling of a chemical reactor, failure of both pumps will create an emergency situation. If the two cooling devices are from separate manufactures or operate on separate physical principles, the common cause faults will be blocked out. For redundant cooling devices, if one of them is powered by electricity and the other uses natural gravitation to operate, the common cause “absence of power supply” will be blocked out. If, in addition, the two cooling devices are serviced/maintained by separate operators, the common cause ‘faulty maintenance’ will also be blocked out.

Similarly, a common cause due to an incorrect calibration of measuring instruments can be avoided if the calibration is done by separate operators. If finally, the cooling devices are separated in different rooms, the common cause failure due to fire will also be blocked out.

A common cause failure due to a software bug for example, can be avoided if a separate algorithm and implementation are provided for the same task or if a separate team is involved in developing the same piece of software, independently.

Separating investment in unrelated sectors protects against a common cause reducing simultaneously the return from all sectors (e.g. agricultural sectors simultaneously affected by bad weather or disease, consumer sectors simultaneously affected by a health scare, investments in different sectors in a country affected by a political crisis, economic crisis, social unrest, etc.

It needs to be pointed out that judgement needs to be exerted in applying the method of separation. The act of separation cannot automatically guarantee that a risk reduction will always be achieved. In some cases, the antipode of separation - the *unification* achieves the risk reduction. In cases where the separate components performing separate functions are lightly loaded, unifying the components into a single component performing all functions often achieves the risk reduction, because of reducing the complexity of the system. Unification can often eliminate a failure mode thereby reducing the risk of failure. Thus, testing the corrosion resistance of specimens made of different alloys could be done by shaping the test specimens into reservoirs which hold the corrosive agent. This is an act of unification because the specimen now performs two functions: (i) test specimen and (ii) a reservoir for the corrosive agent. By the act of unification, the need for a reservoir to hold all test specimens is avoided and with it, the failure mode 'corrosion of the reservoir holding the specimens' is also avoided. Making contacting parts of materials with similar chemical composition (unification of properties) often eliminates harmful electrochemical corrosion. Unifying the thermal expansion properties of different contacting parts by selecting materials with similar thermal expansion coefficients reduces the thermal stresses and thermal fatigue and enhances reliability.

CONCLUSIONS

Analysis of various separation techniques and the mechanisms through which they improve reliability and reduce risk was presented for the first time. A comprehensive classification of techniques for improving reliability and reducing risk, based on the method of separation was also proposed for the first time.

From the presented classification, three principal categories of separation techniques reducing risk have been identified: (i) assuring distinct functions/properties/behaviour for distinct components/parts (ii) assuring distinct properties/behaviour at distinct time, value of a parameter, conditions or scale and (iii) distancing risk-critical factors.

The concept 'stochastic separation' of random events has been introduced. Stochastic separation is present if the separation of the risk-critical events is guaranteed with a specified probability. A method for providing stochastic separation with a specified probability, between random events, was also introduced.

The method of segmentation for reducing risk can be considered as a special case of the method of separation where the same function is separated into distinct parts carrying the function.

The traditional reliability measure ‘safety margin’ is very misleading and must not be used to determine the relative separation of the load and strength for non-Gaussian distributions of the load and strength.

Separation on a parameter is an efficient technique for reducing risk. The deliberate weak links technique and the stress limiters technique are essentially instances of separation on the parameter ‘loading stress’.

Separation on properties is an efficient technique for compensating the drawbacks associated with a selection based on homogeneous properties.

Introducing deliberate weak links to reduce risk is essentially an application of the method of separation.

The logical separation, making it logically impossible for a dangerous operation to occur at a given point in time or space, can be an efficient and low-cost risk reduction technique.

REFERENCES

Altshuller G.S. (1984). *Creativity as an exact science*: *The theory of the solution of inventive problems*. New York: Gordon and Breach Science Publishing.

Altshuller G.S. (1996). *And suddenly the inventor appeared, TRIZ, the theory of inventive problem solving*, *Translation from Russian by Lev Shulyak*, Worcester, MA: Technical Innovation Center.

Altshuller G.S. (2007). *The innovation algorithm, TRIZ, systematic innovation and technical creativity*, Worcester: Technical Innovation Center, Inc.

Billinton, R. & Allan, R. N. (1992). *Reliability evaluation of engineering systems* (2nd ed.). Plenum press.

Carter A.D.S. (1986). *Mechanical Reliability*, Macmillan Education Ltd.

Carter A.D.S. (1997). *Mechanical reliability and design*, Macmillan Press Ltd.

Eder W.E. & Hosnedl S.. (2008). *Design Engineering*, CRC Press.

Gadd K. (2011). *TRIZ for engineers*: *Enabling inventive problem solving*, Wiley.

Gere J.M. & Timoshenko S.P. (1999). *Mechanics of materials, 4th SI ed.*, Stanley Thornes (Publishers) Ltd.

Hollangel E. (2016). *Barriers and accident prevention*, Routlege.

Kalpakjian S. & Schmid S. (2001). *Manufacturing engineering and technology 4th ed*., Prentice Hall.

Leveson N. (2011). *Engineering a safer world: systems thinking applied to safety*, Cambridge, Massachusetts: The MIT Press.

Lewis E.E. (1996). *Introduction to reliability engineering*, New York: John Wiley & Sons, Inc.

O'Connor P.D.T. (2002). *Practical reliability engineering*, 4th ed., New York: John Wiley & Sons, Ltd.

Orloff M.A. (2006). *Inventive thinking through TRIZ*, *A practical guide*, 2nd ed., Springer.

Orloff M.A. (2012). *Modern TRIZ A Practical Course with EASyTRIZ Technology, A practical guide*, 2nd ed., Springer.

Ramakumar R. (1993). *Engineering reliability, fundamentals and applications*, Englewood Cliffs: Prentice Hall.

Rantanen K. & Domb E. (2008). *Simplified TRIZ*, 2nd edition, Auerbach Publications.

Reade C. (1989). *Elements of Functional Programming*. Boston, MA: Addison-Wesley Longman.

Ross S. (1997). *Simulation*, (2nd ed.). Harcourt academic press.

Savransky S.D. (2000). *Introduction to TRIZ methodology of inventive problem solving*, CRC press LLC.

Svenson O. (1991). The accident evolution and barrier function (AEB) model applied to incident analysis in the processing industries. *Risk Analysis*, 11(3), 499-507.

Terninko J., Zusman A. & Zlotin B. (1998). *Systematic Innovation: An introduction to TRIZ*, CRC Press LLC.

Todinov M.T. (1998). Mechanism for the formation of the residual stresses from quenching, *Modelling and Simulation in Materials Science and Engineering*, 6, 273-291.

Todinov M.T. (1999). Influence of some parameters on the residual stresses from quenching, *Modelling and Simulation in Materials Science and Engineering*, 7(1), 25-41.

Todinov M.T. (2005). Limiting the probability of failure for components containing flaws, *Computational Materials Science*, 32, 156-166.

Todinov M.T. (2015). Reducing risk through segmentation, permutations, time and space exposure, inverse states, and separation, *International Journal of Risk and Contingency Management* 2015, 4(3),1-21.

Ullman D.G. (2003). *The Mechanical Design Process*, 3rd ed., McGraw Hill.
Wang B.B. (2004). Free-standing tension structures, Spon press, London.