Reducing risk through segmentation, separation, time and space exposure, inverse states, and separation

Michael Todinov
Oxford Brookes University
Department of Mechanical engineering
and Mathematical sciences, UK

ABSTRACT

The paper features a number of new generic principles for reducing technical risk with a very wide application area. Permutations of interchangeable components/operations in a system can reduce significantly the risk of system failure at no extra cost. Reducing the time of exposure and the space of exposure can also reduce risk significantly.

Technical risk can be reduced effectively by introducing inverse states countering negative effects during service. The application of this principle in logistic supply networks leads to a significant reduction of the risk of congestion and delays. The associated reduction of transportation costs and environmental pollution has the potential to save billions of dollars to the world economy.

Separation is a risk-reduction principle which is very efficient in the cases of separating functions to be carried out by different components and for blocking out a common cause. Segmentation is a generic principle for risk reduction which is particularly efficient in reducing the load distribution, vulnerability to a single failure, the hazard potential and damage escalation.

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

INTRODUCTION

Despite the critical importance of distilling generic principles related to reducing technical risk, very little has been published on this topic. For a long time, the principles of technical risk reduction have been exclusively focused within specific industries, technologies or operations, for example: oil and gas industry, nuclear industry, aviation, construction, medicine, banking, welding, heat treatment, machining, casting, forging, transportation, handling poisonous substances, handling heavy loads, etc.

These risk reduction principles tend to be oriented towards avoiding or mitigating particular failure modes in the specific application area and usually have no general validity. A similar feature characterises even principles of technical risk reduction related to commonly occurring failure modes across various engineering disciplines, for example, principles related to *fatigue* and *fast fracture* of engineering components (Ewalds & Wanhill, 1984; Hertzberg, 1996; Zahavi & Torbilo 1996; Anderson, 2005).

Reliability engineering usually focusses on predicting the reliability of components and systems and does not normally discuss principles for reliability improvement. Improving reliability by active and standby 'redundancy', by strengthening the weakest link, by developing physics-of-failure models, by eliminating a common cause and by reducing variability for example, are generic risk reduction principles that have been covered well in the reliability literature (Barlow & Proschan 1975; Ebeling, 1997; O'Connor, 2003; Lewis, 1996; Todinov, 2007). There exists a rather simplistic view among some reliability practitioners that improving the reliability of a system involves either improving the reliability of the components or providing redundancy. Equally simplistic is the view that only developing physics of failure models can deliver reliability improvement. This view has been fuelled by the failure of some reliability models to predict correctly the life of engineering components. A possible contributing reason is the widespread erroneous view that the quality and utility of reliability models depends strongly on the availability of failure data. Comparative statistical models however, based on assumed input data, can deliver real reliability improvement in the absence of any failure data. For example, in comparing the performance of competing network topologies and selecting the topology with the best performance, a comparative method for assessing the performance of

competing network topologies could proceed by: (i) assuming common flow capacities, failure frequencies, and repair times for the corresponding components/edges of the compared networks; (ii) determining the performance of the competing networks by using an appropriate software tool and finally (iii) selecting the best-performing topology.

These extreme views demonstrate unnecessary self-imposed constraints. Increasing reliability can be achieved by using principles which range widely from pure statistical modelling to pure physics-of-failure modelling underpinning the reliable operation and failure.

The risk literature (Vose 2002; Aven 2003; Bedford & Cooke) is oriented towards risk modelling, risk assessment, risk management and decision making and there is very little discussion related to generic principles for reducing technical risk.

The Taguchi's experimental method for robust design through testing (Phadke 1989) achieves designs where the performance characteristics are insensitive to variations of control (design) variables. This method can be considered to be an important step towards formulating the generic risk reduction principle of robust design whose performance characteristics are insensitive to variations of design parameters.

French (1999) formulated a number of generic principles to be followed in conceptual design, but they were not necessarily oriented towards reducing technical risk. Generic principles to be followed in engineering design have been discussed in Pahl 2007. Most of the discussed principles however are either not related to reducing the risk of failure or are too specific (e.g. the principle of thermal design), with no general validity. Collins (2003) discussed engineering design with a failure prevention perspective. The formulated generic guidelines to be followed by engineer-designers however were given in a specific context of mechanical design and no generic principles for reducing technical risk were formulated. An interesting classification of human errors and methods for reducing human errors has been made in Dhillon and Singh (1981) which again do not have general validity.

The struggle between the need of increasing efficiency and reducing the weight of components and systems and reliability is a constant source of technical and physical contradictions. Hence, it is no surprise that several principles for resolving technical contradictions formulated by Altshuller in the development of TRIZ methodology for inventive problem solving (Altshuller, 1984,1996, 1999) can also be used for reducing technical risk. Eliminating harmful factors and influences is the purpose of many inventions and Altshuller's TRIZ system captured a number of useful generic design principles closely related to eliminating harm.

Most of the generic principles for technical risk reduction however, are routed in the reliability and risk theory and cannot be deduced from the general inventive principles formulated in TRIZ, which serve as a general guide in developing inventive solutions, as an alternative to the trial-and-error approach. Some principles for technical risk reduction rely on concepts like 'robust fault-tolerant design' with reduced sensitivity to the variation of reliability-critical design parameters; other principles are based on guaranteeing with large probability minimum separation intervals between random events (Todinov, 2004).

Some principles are rooted in the logic of operation of devices and the logic of execution of operations ('Failure prevention interlocks'); other principles rely on specific systematic methods for discovery and elimination of failure modes. Contrary to what some authors stated, optimisation is not necessarily about finding a compromise between several parameters to maximise a particular system output. Thus, producing a robust engineering assembly, whose characteristics are insensitive to the variation of the parameters characterising the system's components can be achieved not only by finding appropriate mean values for the component parameters within a selected design. A robust (optimised) engineering assembly can also be created by departing radically from the selected design and selecting a new assembly design exhibiting less sensitivity of its characteristics to variations of the parameters characterising the components.

The systematic distilling, formulating and classifying of generic principles for reducing technical risk was started in (Todinov 2007) where the principles for risk reduction have been divided into: 'preventive' - reducing mainly the likelihood of failure; 'protective' - reducing mainly the consequences from failure and 'dual' - oriented towards reducing both the likelihood of failure and the consequences from failure. The formulated set of principles however is by no means comprehensive.

In this paper, a number of new generic principles for risk reduction with very wide application area are proposed which, to the best of our knowledge, have never been reported before. The unifying feature of the presented principles is that they are truly universal and can be applied in diverse areas of the human activity, for example in environmental sciences, project management, logistics supply, financial engineering, economics, medicine, etc.

The presented set of principles prompts risk managers not to limit themselves within few familiar ways of improving reliability and reducing risk which often leads to solutions which are far from optimal. Using appropriate combinations of diverse principles often brings a considerably larger effect.

PERMUTATIONS OF INTERCHANGEABLE COMPONENTS AND PROCESSES

Consider the system in Figure 1 which transports cooling liquid from three sources s1,s2 and s3 to the chemical reactor *t*.

The cooling system consists of identical pipeline sections (the arrows in Figure 1). Each pipeline section is coupled with a pump for transporting the cooling fluid through the section. Suppose that the pipeline sections and the pumps are old and prone to failure due to corrosion, fatigue, wear, deteriorated seals, etc. The cooling system fulfils its mission if at least one cooling line delivers cooling fluid to the chemical reactor. Suppose for the sake of simplicity that all pipeline sections are in the same state of deterioration and each section is characterized with the same reliability 0.4, associated with one year of operation. Because of the deteriorated sections, the cooling system will benefit from risk-reduction consisting of purchasing and replacing deteriorated pipeline sections with new sections. Consequently, the replacement of any of the 9 pipeline sections is a possible risk-reduction option. Now suppose that the available budget is sufficient for purchasing and replacing exactly 3 pipeline sections. Each new pipeline section is characterised by a reliability 0.9 for one year of operation.

Because of the symmetry of the system in Figure 1a, the replacement of any pipeline section is associated with the removal of the same amount of system risk. The pipeline sections work independently from one another and because all of them are identical, it seems that any three pipeline sections can be replaced with new ones (Figure 1b), with the same effect.

This impression however is incorrect. The total removed risk of system failure is highest if the available budget is spent preferentially on replacing pipeline sections forming an entire cooling branch (Figure 1c), as opposed to replacing randomly selected sections inside the system (Figure 1b).

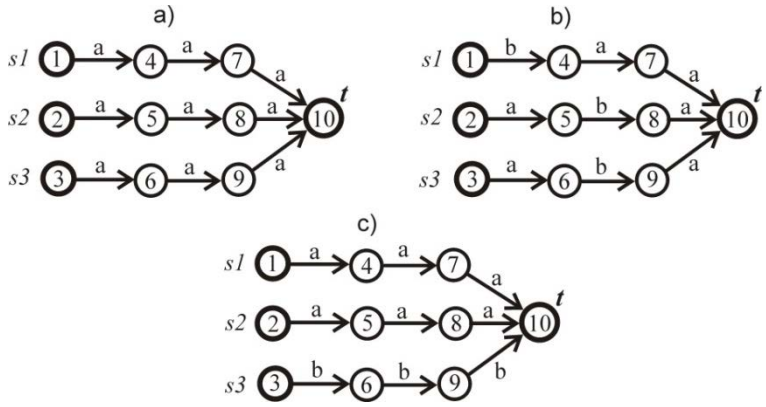

Figure 1. A safety-critical cooling system consisting of three parallel branches

Indeed, the reliability of the parallel-series arrangement in Figure 1b is:

$$R_b = 1-(1-0.4^2\times 0.9)^3 = 0.373 \tag{1}$$

while the reliability of the parallel-series arrangement in Figure 1c is significantly higher:

$$R_c = 1-(1-0.4^3)^2\times(1-0.9^3) = 0.76 \tag{2}$$

The variant presented in Figure 1c is an example of *a well-ordered parallel-series system*. A well-ordered parallel-series arrangement is obtained if the available components are used to build the branch with the highest possible reliability, the remaining components are used to build the next branch with the highest possible reliability and so on, until the entire parallel-series arrangement is built.

If there are three types of branches with different age: new, medium and old branches, the maximum reliability is achieved if all new components are arranged in a single branch, the medium age components in another branch and all old-age components are grouped in a separate branch (Figure 2).

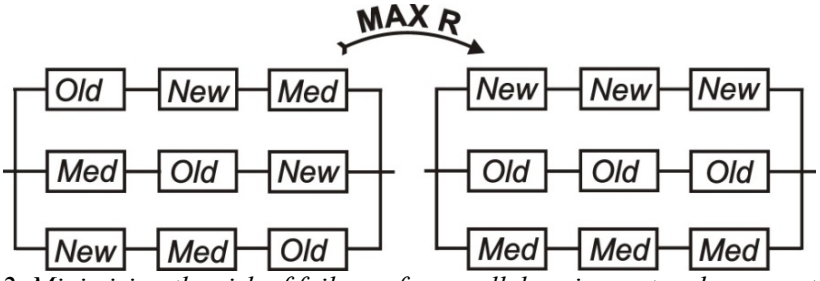

Figure 2. Minimising the risk of failure of a parallel-series system by permutation of interchangeable components

Parallel-series arrangements are very common. Consider a safety-critical system for detecting the release of toxic gas, based on *n* detectors working in parallel. Upon a toxic gas release, the system detects the critical event if at least one of the detectors working in parallel detects the toxic gas release. This system is often a parallel-series system because the parts building the separate detectors are logically arranged in series.

The result stated for the maximum reliability of well-ordered parallel-series systems has been verified by a computer simulation. The computer simulation consisted of specifying the reliabilities of the interchangeable components in the branches and calculating the reliability of the well-ordered system. The second phase of the validation program is a "random scrambling" of the interchangeable components in the branches, by generating random indices of components from different branches and swapping their reliability values. The swapping guarantees that any resultant system includes exactly the same set of interchangeable components as the initial system. After the 'random scrambling', the reliability of the scrambled system was calculated and compared with the reliability of the well-ordered system. If the reliability of the well-ordered system was greater than or equal to the reliability of the scrambled system, the content of a counter was increased. At the end, the probability that the well-ordered system has reliability not smaller than the reliability of the scrambled system was calculated. In all of the conducted simulations, this probability was always equal to one, which confirms that well-ordered systems are indeed characterised by the largest reliability/availability.

These results can be summarised by stating a generic risk-reduction principle: ***The well-ordered parallel-series system is characterised by the smallest possible risk of failure.***

Proof. This principle will be proved by a contradiction and the extreme principle. Suppose that there is a system which is not well-ordered and which possesses the highest possible reliability. Without loss of generality, suppose that the branches in this system have been re-arranged in such a way that for any two branches '*i*', '*j*' for which $i < j$, the branch with index '*i*' is equally reliable or more reliable than branch '*j*' ($R_i \geq R_j$). If the system is not a well-ordered system, then there will be two branches *a* and *b* with reliabilities $R_a \geq R_b$, where there will be at least one component in branch *b* with a larger reliability than the reliability of the analogous interchangeable component in branch *a*. Suppose that $R_a = a_1 a_2 \times ... \times a_{na}$ and $R_b = b_1 b_2 \times ... \times b_{nb}$ are the reliabilities of branches *a* and *b* and na, nb are the number of components in branches *a* and *b*, correspondingly. Without loss of

generality, suppose that the two analogous interchangeable components mentioned earlier, are the last components in the branches *a* and *b* ($a_{na} < b_{nb}$).

The reliability of the initial system can be presented as

$$R_{sys1} = 1 - (1 - a_1 a_2 \times ... \times a_{na})(1 - b_1 b_2 \times ... \times b_{nb}) \times [1 - R_{rest}] \quad (3)$$

where R_{rest} is the reliability of the rest of the parallel-series arrangement.

After swapping components a_{na} and b_{nb}, the reliability of the resultant system becomes

$$R_{sys2} = 1 - (1 - a_1 a_2 \times ... \times a_{na-1} b_{nb})(1 - b_1 b_2 \times ... \times b_{nb-1} a_{na}) \times [1 - R_{rest}] \quad (4)$$

Subtracting (4) from (3) yields:

$$R_{sys1} - R_{sys2} = (a_{na} - b_{nb})(a_1 a_2 \times ... \times a_{na-1} - b_1 b_2 \times ... \times b_{nb-1}) \times [1 - R_{rest}] \quad (5)$$

Because $R_a = a_1 a_2 \times ... \times a_{na} \geq R_b = b_1 b_2 \times ... \times b_{nb}$ by the way the branches have been arranged in descending order according to their reliability ($R_a \geq R_b$), and because $a_{na} < b_{nb}$ (by assumption), the inequality

$$a_1 a_2 \times ... \times a_{na-1} > b_1 b_2 \times ... \times b_{nb-1} \quad (6)$$

holds, which means that in equation (5) $a_1 a_2 \times ... \times a_{na-1} - b_1 b_2 \times ... \times b_{nb-1} > 0$.

Since $1 - R_{rest} > 0$, the right hand side of equation (5) is negative, which means that the resultant system (after the swap of components) has a higher reliability. This contradicts the assumption that the initial system (before the swap) was the system with the highest possible reliability. Therefore the reliability of a system which is not well-ordered, can be improved by swapping components between parallel branches until a well-ordered system is finally obtained. A well-ordered system is unique and there can be no two well-ordered systems. Because a parallel-series system can either be a well-ordered or not well-ordered system, the well-ordered system has a higher reliability compared to any other arrangement. The risk-reduction principle has been justified.

This principle provides an opportunity to remove the maximum amount of system risk *by concentrating the available budget on renewing single parallel branches as opposed to randomly replacing aged components in the system.*

This result also provides the valuable opportunity to improve the reliability of common systems with parallel-series logical arrangement of their components *without the knowledge of their reliabilities and without any investment*. Unlike all traditional approaches, which invariable require resources to achieve reliability improvement and risk reduction, a system risk reduction can also be achieved by appropriate permutation of the available interchangeable components in the parallel branches.

Components of similar level of deterioration (reliability levels) should be placed in the same parallel branch (see the example from Figure 2).

The risk reduction principle based on permutation of interchangeable components has wide applications reaching far beyond its initial engineering context.

Consider a common example where three groups of people (teams) 1, 2 and 3, each of which includes three independently working team members. The teams work in parallel towards achieving the same goal (Figure 3a). The goal is achieved if at least one of the teams succeeds in achieving the goal. Within each team, the task of achieving the goal is divided into subtasks among the team members. Every single person in a team must accomplish their sub-task successfully, in order for the team to achieve the goal. The level of training of each team member is from one of the categories: Strong (S), Weak (W) and Medium (M). A person with strong level of training has a better chance of accomplishing a task successfully compared to a person with medium training or weak training. A person with medium training has a better chance of accomplishing the task successfully compared to a person with weak training.

Separating the people in groups with similar level of training (Figure 3b) yields the highest chance of achieving the goal. Note that reducing the risk of not achieving the goal has been achieved at no extra cost.

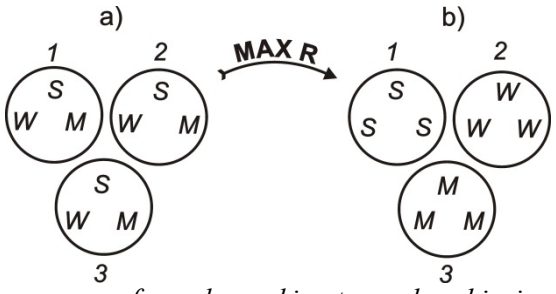

Figure 3. Three groups of people working towards achieving the same goal

TIME OF EXPOSURE AND SPACE OF EXPOSURE

Reducing risk by reducing the time of exposure

Suppose that a particular hazard is characterised by a cumulative distribution function $F_H(x)$ of its magnitude (impact). The hazard appears during a finite time interval with length *t* and the times of hazard occurrence follow a homogeneous Poisson process with intensity ρ. Suppose that the resistance to the hazard (the strength) is characterised by a probability density distribution $f_S(x)$. It is also assumed that the hazard and the resistance are statistically independent. The probability that during the finite time interval with length *t,* there will be no critical hazard impact exceeding a specified resistance *x*, is given by $\exp(-\rho_{cr} t)$, where ρ_{cr} is the number density of the hazard occurrence. The number density of the critical hazard occurrences (the occurrences where the hazard magnitude exceeds the resistance *x*) is given by $\rho_{cr} = \rho \times [1 - F_H(x)]$, where ρ is the number density of the hazard occurrences and $1 - F_H(x)$ is the probability that a hazard occurrence will be critical (will exceed the resistance *x*). Consequently, the probability $p_0(x)$ that during the finite time interval with length *t,* there will be no critical hazard occurrences for which the resistance *x* is exceeded, is given by

$$p_0(x) = \exp(-\rho \times [1 - F_H(x)]t) \tag{7}$$

The probability that the resistance will be in the infinitesimal interval *x*, *x*+d*x* is given by $f_S(x)dx$. Assuming that the hazard occurrence does not depend on resistance, the probability that an entity with resistance in the interval *x*, *x*+d*x* will survive all hazard occurrences on the time interval 0,*t* is given as a product of the probabilities of statistically independent events

$$P(resist.\ x\ will\ survive\ all\ hazard\ occurences) = p_0(x) \times f_S(x)\,dx \tag{8}$$

where $p_0(x)$ is given by equation (7).

If $S_{\min}$ and $S_{\max}$ are the lower and upper limit of resistance, the probability of surviving all random hazard occurrences during the time interval 0,*t* is given by

$$R(t) = \int_{S\min}^{S\max} \exp[-\rho\, t\,(1 - F_H(x))]\, f_S(x)\, dx \tag{9}$$

The probability of no failure (the reliability) associated with the finite time interval (*0,t*) can then be calculated from the integral in equation (9).

The probability of failure p_f is simply

$$p_f = 1 - \int_{S\min}^{S\max} \exp[-\rho\, t\,(1 - F_H(x))]\, f_S(x)\, dx \tag{10}$$

The term $\exp[-\rho\, t(1 - F_H(x))]$ in the integral gives the probability that none of the hazard occurrences in the time interval *0,t* will exceed resistance with magnitude *x*.

With reducing the time of exposure t, the probability $\exp[-\rho t(1-F_H(x))]$ that the hazard magnitude will exceed resistance (the probability of failure) decreases significantly, which leads to a significant reduction of the probability of failure p_f.

If the process or action is conducted within a very small time interval $t \approx 0$, the probability of failure tends to zero.

$$\lim_{t\to 0} p_f = 1 - \int_{S\min}^{S\max} f_S(x)dx = 1 - 1 = 0 \tag{11}$$

This principle has a wide application to processes and operations affecting all areas of the human activity. Thus, if the overall duration of an operation is reduced, it is less likely to be disrupted (failed) by a random cause. Reducing the length of operation reduces significantly the probability of encountering a critical hazard.

Consider an example from transportation. For road accidents following a homogeneous Poisson process with intensity ρ, along a road with length L_1, the probability of no road accident associated with the length L_1 is $p_1 = \exp(-\rho L_1)$. If the road length is increased by a factor of m ($L_2 = mL_1$), the probability of no accident will be $p_2 = \exp(-\rho m L_1)$. Taking the ratio of the logarithms of these probabilities gives

$$\frac{\ln p_1}{\ln p_2} = 1/m$$

from which $p_2 = (p_1)^m$. From the last expression, if for example, the probability of no road accident associated with road length L_1 is $p_1 = 0.9$, increasing the length of the road four times decreases the probability of no road accident to $0.9^4 = 0.66$. Unlike short journeys, long journeys are likely to be affected by delays caused by road accidents. Consequently, for long journeys, a delay reflecting potential accidents should be added to the estimated overall time of the journey.

Reducing risk by reducing the space of exposure

Suppose that V is the volume of a component with complex shape subjected to a complex loading. Let λ be the number density of the flaws which follow a homogeneous Poisson process in the stressed volume V. The probability that in the stressed volume V, there will be no critical flaws capable of causing failure is given by $\exp[-\lambda_{cr} L]$. The number density of the critical flaws is $\lambda_{cr} = \lambda \times F_c$, where F_c is the probability that a single flaw will be critical, given that it resides in the stressed volume V. Consequently, the probability that the stressed component will not contain a critical flaw becomes $\exp[-\lambda V \times F_c]$ and the probability of failure of the component will be

$$p_f = 1 - \exp[-\lambda V \times F_c] \tag{12}$$

The conditional probability F_c of failure of the stressed component given that a single flaw with random size resides in the volume V can be determined easily from a finite elements solution.

Suppose that the stressed volume has been discretized into finite elements for each of which the maximum tensile stress is known. Suppose that the size distribution of the flaws is given by the cumulative distribution $F(d)$. The function $F(d)$ gives the probability that the size of a random flaw will not exceed a particular value d. The probability F_c that a single flaw will be critical, given that it resides in the stressed volume V, can be determined by the following algorithm:

Algorithm

$F_c = 0$;

For each finite element *i* **do steps 1-5**

1. Determine the maximum tensile stress σ in the *i*th finite element;
2. Determine the critical flaw diameter $d_{\sigma,i}$ beyond which failure will be initiated by the maximal tensile stress σ;
3. Determine the probability $p_{f,i} = 1 - F(d_{\sigma,i})$ that a flaw with random size will cause failure if it resides in finite element *i*
4. Determine the probability that a flaw with random size residing in the volume *V* will actually reside in the *i*th finite element and will cause failure:

$$p_{c,i} = (v_i / V) \times [1 - F(d_{\sigma,i})]$$

5. Accumulate the probability $p_{c,i}$ into the total probability F_c: $F_c = F_c + p_{c,i}$

The probability $p_{c,i} = (v_i / V) \times [1 - F(d_{\sigma m,i})]$ that a flaw with random size (given that it is in the stressed volume *V*) will reside in the *i*th finite element and will cause failure is given by the product of the probability v_i / V that the flow will reside in the volume v_i of the *i*th finite element and the probability $1 - F(d_{\sigma,i})$ that its diameter will be larger than the critical flaw diameter $d_{\sigma,i}$ causing failure under the maximum stress σ characterising the *i*th finite element. Because a single flaw cannot be in the volume of more than one finite element, the probabilities $p_{c,i}$ are added as probabilities of mutually exclusive events.

At the end, the probability of failure of the component with complex shape and volume *V* is given by equation (12).

If the volume is small ($V \approx 0$), the probability of having a critical flaw in the volume *V* tends to zero:

$$\lim_{V \to 0} [p_f = 1 - \exp(-\lambda V \times F_c)] = 0 \tag{13}$$

Now assume that the volume *V* has been discretised into *n* elements with the same volume $v = V / n$. Adding all elementary probabilities $p_{c,i} = (v / V) \times [1 - F(d_{\sigma,i})]$ gives:

$$F_c = \sum_{i=1}^{n} [(v / V) \times (1 - F(d_{\sigma,i}))] = \frac{1}{n} \sum_{i=1}^{n} [1 - F(d_{\sigma,i})] \tag{14}$$

for the conditional probability that a flaw with random size will be critical, given that it resides in the stressed volume *V*. In words, the conditional probability of failure given that the flaw resides in the volume *V* is equal to the average of the conditional probabilities of failure given that the flaw resides sequentially in each of the elementary volumes. Substituting F_c in equation (12) gives the probability of failure of the component with complex shape.

A typical example of limiting the technical risk of failure by reducing the space of exposure is reducing the length of a piece of wire in order to reduce the probability that a critical defect will be present.

Another example is limiting the risk of an error in a long chain of the same type of calculations. Assuming that the errors follow a Poisson distribution, if λ is the number of errors per unit number of calculations, the probability of an error associated with the total number of calculations *N* is given by $p_f = 1 - \exp(-\lambda N)$. Reducing the number of calculations *N*, dramatically reduces the probability p_f of a calculation error.

INTRODUCING INVERSE STATES

Inverse states cancelling the anticipated state with a negative impact or acting as counter-balancing forces

An inverse state of the anticipated negative impact state can be used to compensate the negative effect. The two states superpose and the result is an absence or a significantly attenuated negative effect.

In acoustics, this principle works in noise-cancellation headphones designed for reducing the risk of hearing damage caused by noise. A sound wave is emitted with the same amplitude but with inverted phase to the noise. The result is a significant attenuation of the harmful noise and reduced risk of hearing damage.

This principle also underlies active methods of controlling vibration. The active vibration control involves suitable vibration sensors (e.g. accelerometers), controllers and actuators for vibration control. The signal from the vibration sensors is fed to a controller and through an actuator, a spectrum of cancellation vibrations are generated in response. The advances in the sensor, actuator and computer technology made active methods of control cost-effective and affordable.

Deliberate inverse states cancelling negative anticipated effects is at the heart of many temperature-compensation circuits designed to mitigate the impact of generated heat on the parameters of the electronic devices. For example, creating inverse states through a Wheatstone bridge cancels the negative effect from temperature and is the foundation of various measurement techniques characterised by a small error.

In mechanical engineering, a typical example is the compensation clock pendulum in which the temperature elongation of the pendulum rod is counteracted by an opposite expansion so that the period of oscillation remains the same**.** A typical application of an inverse state as a counterbalancing force are the counterweights in cranes which reduce the loading on the lifting motor and improve the balance and stability of the crane. Another example is the gate valve which is maintained open by a hydraulic pressure acting against a counterbalancing compression spring. Upon failure of the hydraulic system, the counterbalancing spring expands and returns the valve in closed (safe) position.

Inverse states buffering the anticipated state with a negative impact

Introducing an inverse state which serves as a buffer can be done in many cases where a negative effect has been anticipated and the inverse state is provided for buffering the impact of the anticipated negative effect.

This technique underlies reducing the risk of failure of zones generating heath. Components working in close contact (e.g. piston-cylinder) and moving relative to each other generate heat which, if not dissipated, causes intensive wear, reduced strength and deformations. The risk of failure of such an assembly is reduced significantly if one of the parts (e.g. the cylinder) is cooled to dissipate the released heat which reduces the friction and wear.

The *cold expansion*, used in aviation for creating compressive stresses at the surface of fastener holes (Figure 4) is another example of using buffering inverse states.

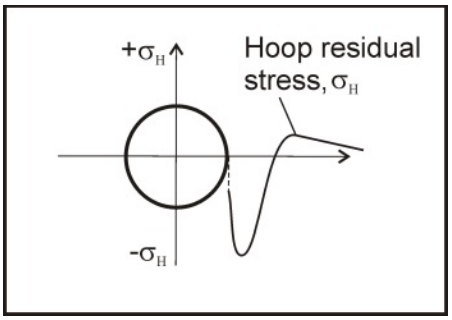

Figure 4. Countering the stress-concentration effect of a hole by creating compressive stresses through cold expansion.

This is done by passing a tapered mandrel through the hole. The inverse state created in the vicinity of the hole (compressive residual stress field), counters the tensile loading stresses during operation and impedes the formation of fatigue cracks at the edge of the hole and their propagation which reduces the risk of fatigue failure.

Another way of introducing an inverse state through plastic deformation is the process of *pre-setting* used in spring manufacturing. For compression springs for example, pre-setting consists of inducing a permanent plastic deformation of the spring, which reduces the length of the spring and results in a compressive residual stress at the surface of the spring wire. During loading, the compressive residual stress at the spring surface is subtracted from the tensile stress from loading. The result is a smaller stress range and increased fatigue life of the pre-set spring.

In order to counter the tensile stresses from loading at the surface and improve fatigue resistance, *shot-peening*, introducing compressive stresses at the surface, has been used as an important element of the manufacturing technology (Niku-Lari, 1981; Bird & Saynor, 1984). As a result of this operation, the fatigue life of leaf springs for example, can be increased up to 10 times.

In the construction industry, the pre-tensioned concrete is a typical example of buffering the negative effect by introducing an inverse state. The tensile stresses from bending of concrete beams can be reduced if pre-loaded in tension tendons (steel cables or rods) are inserted in the beam to provide a clamping load. After the concrete sets, the beam is pre-loaded in compression. The compressive stress from pre-loading is an inverse state which compensates the tensile loading stresses. Since the tensile stresses from bending superpose with the compressive residual stresses, the effective stress during service is compressive or a significantly reduced tensile stress. Pre-stressed concrete is the main material for floors in high-rise buildings. In addition, pre-stressing makes it possible to construct larger spans in bridges and buildings with large column-free spaces.

An inverse state of compressive residual stresses at the surface, acting as a buffer compensating the tensile service stresses from loading, can also be created by a special heat- and thermochemical treatment such as *case-hardening*, *gas-carburising* and *gas-nitriding*.

The *corrosion, erosion* and *wear allowances* added to the computed sections of pipes are other examples of inverse states anticipating the loss of wall thickness. They act as buffers compensating for the loss of wall thickness and decrease significantly the risk of failure.

The use of inverse states as a buffer has a wide application in many other areas of human activity. In project management, providing time buffers for certain critical tasks reduces the risk of a delay should particular risks materialise. Similarly, in managing stock in the presence of random demands, increasing the reserve of a particular safety-critical stock (e.g. particular life-saving medicine) reduces the risk of running out of stock in case of clustering of random demands.

Increasing the financial reserves of a bank or a company makes it less vulnerable to depleting its reserves due to materialised credit and market risks.

Inverting the relative position of objects the motion of objects and the direction of flows

There are cases where inverting the relative position of objects eliminates a detrimental effect from a third factor. Drilling vertical blind holes in components, by a robot, on a manufacturing line for example is associated with the need for cleaning the blind holes from metal chips. If the hole is drilled on the component positioned upside-down, the problem associated with cleaning the hole from chips is eliminated because gravity now helps to clean the hole.

Often making a moving object stationary and moving the stationary object in the opposite direction (inverting the motion) results in a significantly improved performance and risk reduction. This is the idea behind the Cosworth® sand casting process where the molten metal is never poured down into the sand mold as is the case in the classical sand casting process. It flows into the opposite direction (uphill) into the mold which eliminates turbulence and reduces the risk of trapping oxides into the metal which would reduce significantly the fatigue strength of the product.

Inverse states can even be used to reduce the costs associated with multiple source-destination connecting paths in transportation networks, supply networks, communication networks and support networks. Reducing these costs is associated with reducing the waste of energy, the cost of delivering a particular commodity or a service, the level of congestion and environmental pollution. The next example in the area of logistic supply is a unexpected application of this principle.

Figure 5 features a logistic supply network where a particular commodity is delivered from the

three interchangeable sources *s*1,*s*2 and *s*3 to the destinations *t*1,*t*2 and *t*3.

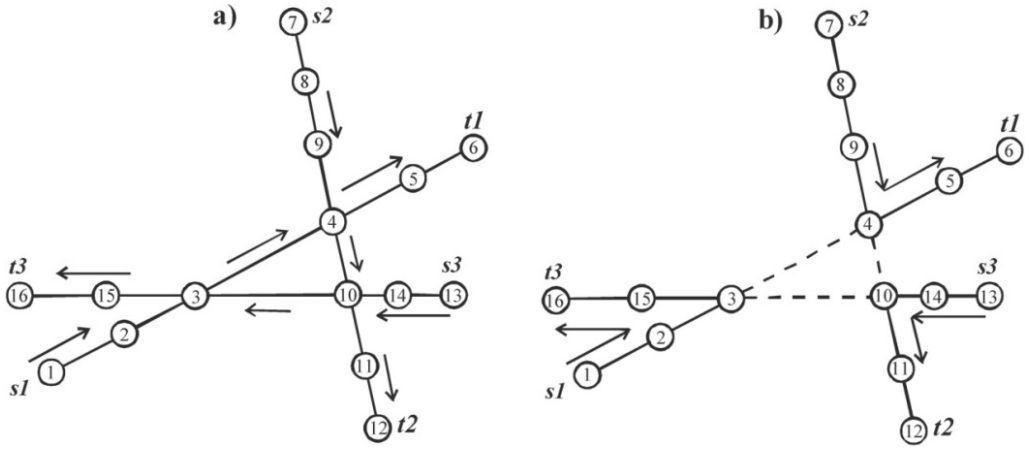

Figure 5. Draining closed parasitic flow loops in logistic supply networks

As a result, a closed parasitic flow loop essentially appears between nodes 3,4,10 and 3 despite that the transported commodity does not travel along a closed contour. Closed parasitic flow loops are cyclic paths where the flow essentially travels in the direction of traversal (Figure 5a, the flow loop 3,4,10,3). By draining the parasitic flow loop by augmenting the cyclic path 3,4,10,3 with flow in opposite direction, the flow loop is eliminated (Figure 5b). As a result, value is derived from significantly reducing the transportation losses and risk of congestion without affecting the throughput flow from the interchangeable sources *s*1,*s*2 and *s*3 to the destinations *t*1,*t*2 and *t*3.

Parasitic flow loops are associated with increased risk of congestion and accidents, big wastage of energy, time, and increased levels of pollution to the environment. Parasitic flow loops exist in real transportation networks with a very high probability. Optimizing supply networks by draining highly undesirable parasitic flow loops derives significant value by reducing the transportation costs, the risk of congestion and accidents, and the environmental pollution. The result is billions of dollars saved to the world economy.

The existence of parasitic flow loops in networks remained unnoticed by scientists for nearly 60 years. Ironically, despite the years of intensive research on static flow networks, closed parasitic flow loops appear even in the "network flow solutions" from all published algorithms (including the famous Ford-Fulkerson algorithm; Ford and Fulkerson, 1956) for maximising the throughput flow in networks, since the creation of the theory of flow networks in 1956.

The parasitic flow loops are not necessarily closed flow loops only. Flow loops, for which more than half of the cyclic path contains flow along a particular direction of traversal are also associated with significant transportation losses and risk of congestion.

In Figure 6, three interchangeable sources s1,s2 and s3 are supplying a particular commodity or service to three destinations d1,d2 and d3. As a result, a parasitic flow loop 4,5,6,7,2,3,4 appears. The parasitic flow loop can be eliminated by augmenting the cyclic path 4,5,6,7,2,3,4 with flow in the opposite direction of the direction of the dominating flow. As a result, the parasitic flow loop disappears (Figure 6b) without affecting the throughput flow from the interchangeable sources to the destinations.

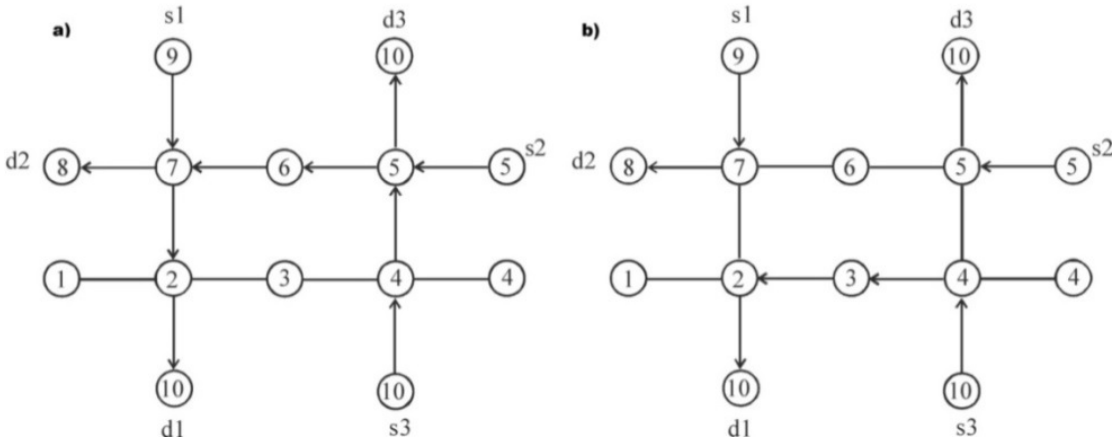

Figure 6. Draining a parasitic flow loop (2,7,6,5,4,3,2)

Interestingly, *selecting the nearest available source does not guarantee an absence of parasitic flow loops*.

In the network in Figure 7a with interchangeable sources s1,s2,s3 servicing destinations d1,d2,d3, the nearest source to the destination d1 is s1, the nearest source to the destination d2 is s2 and the nearest remaining source to the destination d3 is s3. Each source can service not more than one destination. In addition, all of the source-destination pairs have been connected with the shortest paths. Despite the shortest-path selections, the obtained solution is far from optimal. A parasitic flow loop 12,8,3,6,5,11,12 is present and by augmenting it with flow in the opposite direction of the direction of the dominant flow, the new set of connections in Figure 7b appear where no parasitic flow loops are present. The throughput flow from the interchangeable sources s1,s2 and s3 to the destinations d1,d2 and d3 remains the same while the transportation costs have been reduced significantly.

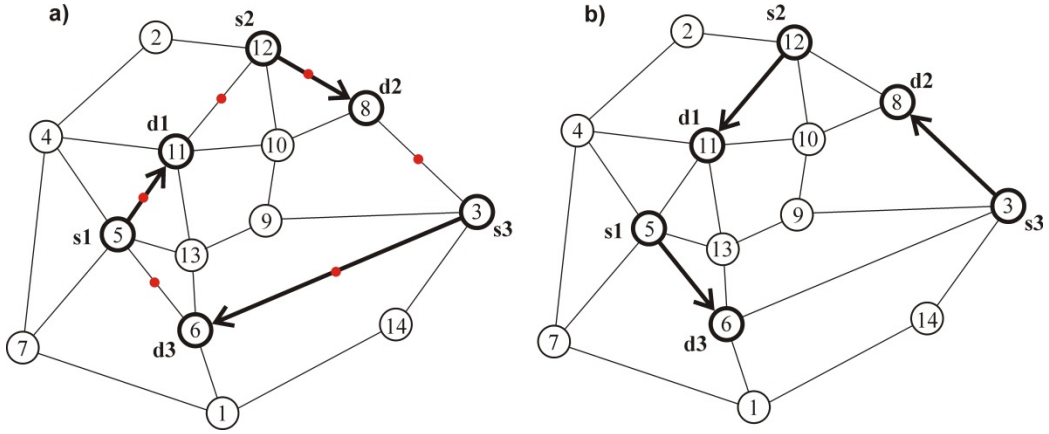

Figure 7. Selecting the nearest available source does not guarantee an optimal solution.

REDUCING RISK BY SEPARATION

In general, it is difficult to optimise a single component carrying many functions with regard to every single function. Separating critical functions and properties is often the key to improving reliability and reducing technical risk. The separation principle can be discovered in the design of flexible pipes carrying hydrocarbons. The different layers in the flexible pipe are designed for 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 loads. The separate layers building the flexible pipe however can be optimised with respect to the function they carry. The result is increased

reliability of the pipe. Separating functions to different components relieves the load on components and reduces the risk of failure.

The separation of functions is also used for mutual compensation of the deficiencies of manufacturing methods. A typical example is the hybrid joint combining an adhesive joint and mechanical fixing. There is clear separation of functions: the adhesive reduces the stress concentration along the joint while the mechanical fixing increases the peel resistance of the adhesive joint and its stiffness.

This principle is also very useful in cases where reliability is balanced against weight and cost. Improving reliability only locally, where it matters saves resources and results in light-weight designs.

Separating critical properties is often present in the design of complex alloys where some of the microstructural components provide resistance to wearout, while other components provide toughness (resistance to crack propagation).

Often, the need for reducing the risk of failure requires different properties from the different parts of the component. Guaranteeing different properties at different parts of the components is the underlying principle behind coatings improving the wear resistance and corrosion resistance. The case hardening of components, which consists of a local induction heating of the surface layers followed by quenching, improves the surface resistance to large contact stresses and wear, while leaving the core tough which is necessary to withstand impact loads.

This principle is often used in composite materials combining structural constituents with different properties in different directions. Concrete used in the construction industry is a material with good compressive strength but small tensile strength. The steel bars in reinforced concrete are placed in areas loaded in tension where the concrete cannot resist tensile stresses.

The separation principle can also be used for mitigating the consequences from failure because an increased local reliability delays the propagation of damage to the rest of the component/structure. Thus, increasing the reliability of a fire door by additional fire-proof coating delays the fire escalation and limits the fire damage.

The separation principle has a wide application. Reliable operation often depends on critical properties, events or factors not being present at the same time or in the same space region. Separating people from hazards is an important measure for reducing the damage if the control over the hazards is lost. A typical example of time separation of risk-critical factors is the traffic lights, preventing collision between intersecting flows of traffic and flows of pedestrians. 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.

Typical examples of space separation of risk-critical factors is the separation of intersecting flows of traffic and flows of pedestrians at different levels which seliminates the risk of collisions and accidents.

Limiting the spread of infection by urgent quarantine measures isolating infected individuals is another example of the space separation principle.

Application of the separation principle for blocking a common cause

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.

Maintenance and operating actions common to different components is a major source of common-cause failures. Software routines written by the same person/team often exhibit common faults.

Acceleration stresses leading to accumulation of damage and fast wearout are typical examples of common causes. Examples of acceleration stresses are the temperature, humidity, cycling, vibration, speed, pressure, voltage, current, concentration of particular ions, etc. This list is only a sample of possible acceleration stresses and can be extended. Because acceleration stresses lead to a faster wearout, they entail a higher propensity to failure for groups of components which reduces the overall system reliability.

A typical example of this type of common cause failures is the high temperature which increases the susceptibility to deterioration of a group of electronic components. By simultaneously increasing the hazard rates of the affected components, the probability of system failure is increased. Humidity, corrosion or vibrations increase the joint probability of failure of the affected components and shorten the system's life. Even in blocks with a high level of built-in redundancy, in case of a common cause, all redundant components in the block may fail within a short period of time and the advantage from the built-in redundancy is lost.

Failure to account for the acceleration stresses acting as common causes usually leads to optimistic reliability predictions - the actual reliability is smaller than the predicted.

In many cases, the separation principle 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 from 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 control 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 will permit a safe landing of the plane. Separating the redundant components by insulating 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 the basis of 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 cooling pumps (a main pump and an emergency pump) participate in cooling of a chemical reactor, failure of both pumps creates an emergency situation. If the two cooling devices are from different manufactures or operate on different 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 different 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.

Sundararajan (1991) suggests preliminary common-cause analysis which consists of identifying all possible common causes to which the system is exposed and their potential effects. The purpose is to alert risk analysts to potential problems.

Avoiding common links which can be affected by a common cause is an efficient way of blocking out common causes. Such are for example the common location for components, the common storage of data, etc. The destruction of all communication lines due to accident or vandalism can be avoided by avoiding placing all of the communication lines in a common conduit. A common cause failure due to a software bug for example, can be avoided if an alternative 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 failure which reduces 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.

REDUCING RISK BY SEGMENTATION

Segmentation reduces risk by (i) improving the load distribution, (ii) reducing the vulnerability to a single failure, (iii) reducing the damage escalation and (iv) limiting the hazard potential.

Segmentation improves the load distribution

Consider a flange with very few fasteners. A flange connection with a very small number of fasteners leads to excessive stresses in some of the fasteners. Segmentation involving an increased number of fasteners improves the load distribution and reliability.

A significant reliability increase can be achieved if the load is distributed upon many load-carrying units. Thus, the load capacity of a V-belt cannot be increased by increasing its thickness because of increased bending stresses and big hysteresis losses overheating the belt. The load-carrying capacity and reliability however can be increased significantly by multiple parallel V-belts.

Segmentation reduces the vulnerability to a single failure

Segmentation also decreases vulnerability to a single failure. Consider again a flange with very few fasteners. Failure of a single fastener is very likely to cause a loss of containment. A flange with a larger number of fasteners will not be vulnerable to a single failure or even several failures.

Failure of one of the multiple parallel V-belts will not cause failure of the transmission system. Similarly, failure of a single wire in a rope built by twisting many wire strands will not normally cause failure of the rope.

Segmentation reduces the damage escalation

Segmentation also helps reduce the damage escalation and the consequences given that failure has occurred. Segmenting a pipe into many separate sealed segments helps limit the damage from a propagating crack within a single segment only, which reduces significantly the consequences from failure.

Crack arrestors can be strips or rings made of tougher material (Figure 8a). The mechanism of crack arrest consists of reducing the strain energy flow to the crack tip upon encountering a tougher material strip. In Figure 8b, the crack is arrested at the edge of the pipeline section. Segmentation in this case does not prevent cracks from becoming unstable, it only limits the extent of damage once the damage has started escalating. In this case, segmentation reduces risk by limiting the consequences of failure.

Another example of the segmentation principle can be given with buckling of a pipeline subjected to a high external hydrostatic pressure. Buckling could be eliminated by increasing the thickness of the pipeline but this option is associated with significant costs. Control of buckling propagation achieved by using buckle arrestors is a cheaper and more preferable option. Buckle arrestors are thick steel rings welded to or attached at regular intervals to the pipeline in order to halt the propagating buckle and confine damage to a relatively small section (Figure 8a). In this way, the losses from buckling are limited to the length of the section between two buckle arrestors. In case of failure, only the buckled section will be cut and replaced. The spacing between buckle arrestors can be optimised on the basis of a cost-benefit balance between the cost of installation of the arrestors and the expected cost of intervention and repair.

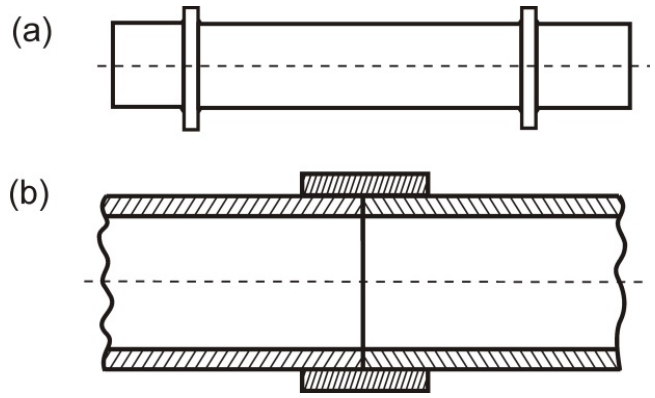

Figure 8. A segmented pipeline with crack arresters of type stiffened welded rings

The segmentation principle can be used for reducing risk in a wide range of applications.

Segmentation is used to increase the resistance of a ship to flooding. The volume of the hull is divided into watertight compartments. If the flooding is localised, only one or very few compartments are affected which allows the ship to retain buoyancy. Segmentation of the corridors in a building, with fireproof doors, protects against the fast escalation of fire.

Segmentation can be used to prevent the spread of infectious diseases. Such is the purpose of preventing formations of large gatherings of people in case of infectious disease.

.

Segmentation limiting the hazard potential

Segmentation can be applied with success to limit the amount of energy possessed by hazards which limits their potential to cause harm. Thus, processing small (segmented) volumes of toxic substances at a time reduces the hazard potential of the substance and the risk of poisoning, in case of accidental spillage. Preventing the formation of large build-ups of snow, water, overheated water vapour etc., reduces both the likelihood of an accident and its destructive power should it occur.

CONCLUSIONS

System risk can be reduced at no extra cost by appropriate permutations of interchangeable components/operations. To achieve a maximum risk reduction in systems with a series-parallel logical arrangement, components or operations of similar reliability must be located on a single branch.

The risk associated with processes and operations can be reduced significantly by reducing the time and space of exposure. Closed-form expressions with a wide application domain have been presented to quantify the impact of the time of exposure and space of exposure.

The risk of failure in many areas of the human activity can be effectively reduced simply by introducing inverse states countering anticipated negative effects during service. In logistic supply networks, the application of this principle leads to a significant reduction of the risk of congestion and delays. The reduction of transportation costs and the environmental pollution has the potential to save billions of dollars to the world economy.

Separation is a generic principle for reducing risk which is particularly efficient in separating functions to be carried out by different components. This relieves the load on components, permits optimisation with respect to the carried function and reduces the risk. The separation principle is also very important for blocking out a common cause.

Segmentation is a generic principle for risk reduction that can be applied with success to reduce risk in many areas of the human activity. It is particularly efficient in improving the load distribution, reducing vulnerability to a single failure, reducing the hazard potential and damage escalation.

REFERENCES

Altshuller, G.S. (1999). *The Innovation Algorithm, TRIZ, Systematic Innovation and Technical Creativity*, Technical innovation Center, Inc. Worcester.

Altshuller, G.S. (1996). *And suddenly the inventor appeared, TRIZ, the theory of Inventive Problem Solving*, Translation from Russian, 1996.

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

Anderson, T.L. (2005). Fracture Mechanics: Fundamentals and Applications, Taylor & Francis.

Aven T. (2003). *Foundations of risk analysis*, Wiley.

Barlow, R.E. & Proschan F. (1975). *Statistical theory of reliability and life testing*, Rinehart and Winston, Inc..

Bedford, T. & Cooke R. (2001). *Probabilistic risk analysis, foundations and methods*, Cambridge University Press.

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

Bird, G.C. & D. Saynor. (1984). 'The effect of peening shot size on the performance of carbon-steel springs', *Journal of Mechanical Working Technology.* 10(2), 175-185.

Collins, J.A. (2003). *Mechanical design of machine elements and machines*, John Wiley & Sons.

Dhillon, B.S., C.Singh (1981). *Engineering reliability: New techniques and applications*, John Wiley & Sons.

Ebeling, C.E. (1997). *An introduction to Reliability and Maintainability Engineering*, McGraw-Hill.

Ewalds, H.L. & Wanhill, R.J.H. (1984). *Fracture Mechanics*, Edward Arnold, London.

Ford, L.R. & D.R. Fulkerson, Maximal flow through a network (1956). *Canadian Journal of Mathematics*, 8(5), 399-404.

French, M. (1999). *Conceptual design for engineers*, 3rd ed., Springer-Verlag London Ltd.

Hertzberg, R.W. (1996). *Deformation and fracture mechanics of engineering materials*, 4th ed., John Wiley & Sons, Inc.

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

Pahl G., W.Beitz, J.Feldhusen, K.H.Grote (2007). *Engineering design*, Springer.

Phadke, M.S. (1989). *Quality engineering using robust design*, Prentice Hall, Englewood Cliffs.

Niku-Lari, A. (1981). 'Shot-peening', *In* the *First International Conference on Shot Peening*, Paris, 14–17 September, 1-27, Pergamon Press.

O'Connor, P.D.T (2003). *Practical reliability engineering*, 4th ed., John Wiley & Sons.

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

Sundararajan, C (Raj) (1991). *Guide to reliability engineering: Data analysis, applications, implementations and management*, Van Nostrand Reinold.

Todinov, M.T. (2004). Reliability governed by the relative locations of random variables in a finite interval, *IEEE Transactions on Reliability*, 53(2), 226-237.

Todinov, M.T. (2007). *Risk-based reliability analysis and generic principles for risk reduction*, Elsevier.

Vose D., (2002), *Risk Analysis:A quantitative guide*, John Wiley & Sons.

Zahavi E., & Torbilo, V. (1996). *Fatigue design, Life expectancy of machine parts*, CRC Press, 1996.

AUTHOR'S BIO

AUTHOR ADDRESS

Michael Todinov
Oxford Brookes University
Department of Mechanical Engineering
and Mathematical sciences
UK

Email: mtodinov@brookes.ac.uk
Tel: (++44) 1865 48 3546

Michael Todinov holds a PhD related to mathematical modelling of thermal and residual stresses and a higher doctorate Doctor of Engineering (DEng) which is the engineering equivalent of Doctor of Science (DSc) in the area of new probabilistic concepts and models in Engineering.

M.Todinov's name is associated with creating the foundations of risk-based reliability analysis (driven by the cost of failure) and the theory of repairable flow networks and networks with disturbed flows. A sample of M.Todinov's results includes: the discovery of closed and dominated parasitic flow loops in real networks; the proof that the Weibull distribution is an incorrect model for the distribution of breaking strength of materials and deriving the correct alternative of the Weibull model; a theorem regarding the exact upper bound of properties from random sampling of multiple sources; a general equation for the probability of failure of brittle components with complex shape, the formulation and proof of the necessary and sufficient conditions of the Palmgren-Miner rule and Scheil's additivity rule and deriving the correct alternative of the Johnson-Mehl-Avrami-Kolmogorov equation. M.Todinov's research has been funded by research councils, the automotive industry, the nuclear industry and the oil and gas industry.