Mark Horton

Real Consequences

Applying the principles of Reliability-centred Maintenance to protective systems and hidden functions

Real Consequences

Table of Contents

Preface

Important Note

Any protective system that is worth maintaining is designed to protect you from a hazard. That hazard may be the cost of damaged equipment or production downtime. The system may protect you from environmental damage. It may even be the final defence against injuring or killing your own personnel or innocent members of the public.

The first chapter of this book shows how the failure of protective systems can lead to disaster; but even before you read that chapter, you need to understand that disasters have occurred not just because of negligence and lack of maintenance, but also from active interference and maintenance of those devices. Before implementing any failure management solution, you need to be certain that you understand the systems, the tasks you have put in place, and the level of risk that your company or organisation is assuming.

I wrote this book because modern industry relies so heavily on protective systems that are sometimes poorly understood and often badly maintained. I have been privileged to work with hundreds of technicians, operators, maintainers and managers in production, manufacturing and utility organisations. These individuals are massively committed to improving safety and reliability, and have been brutally honest in discussing both incidents that happened and those many more that were “near misses”. None of these incidents features in this book, but I am hugely grateful to those who persuaded me that this text needed to be written.

Although the reasons for writing this book may be clear, I am also aware that the author of a book has no control over its application. That concerns me. I cannot look over your shoulder to explain why a system should be treated in one way rather than another. I cannot ask you supplementary questions that might suggest a completely different approach from the one you are considering. I have no way of ensuring that the task you ask your maintainers to do is safe. Worse, I am human, and so there are mistakes in this book. If I were sitting next to you, it is probable that one of us would find them, but I am not. Therefore I need to draw your attention to this disclaimer.

Neither the author nor the publisher accepts any responsibility for the application of the information presented in this book, nor for any errors or omissions. The reader accepts full responsibility for the application of the techniques described in this text.

Review

Each chapter of the book ends with a section called *Key Points and Review*. If you have some experience in reliability analysis or in Reliability-centred Maintenance, you may already be familiar with the material covered in some chapters. You may want to use these sections to ensure that you understand the material covered in the chapter before skipping it.

Acknowledgements

I was first introduced to Reliability-centred Maintenance by John Moubray, who applied the techniques developed by civil aviation in every conceivable industry sector around the world. His enthusiasm for RCM was unstoppable. His uncompromising quest for safety and relentless curiosity to a great extent drove the development and documentation of many of the techniques described in this book. At the same time, John approached failure-finding training with his usual good humour, named the whole subject "Fifi", and transformed what might otherwise have been an intimidating subject into an unforgettable experience for anyone fortunate enough to be in his classroom. The reliability community lost a passionate champion with John's death in 2004, and we miss him still.

My colleague and friend Chris James originated the idea of economic optimisation that is described in this book. Over nearly thirty years, our discussions have been long and sometimes challenging, but always rooted in practical engineering and the core principles of safety and environmental integrity. Chris continues to champion the management of hidden failures through RCM.

I was fortunate to meet Nancy Regan when she visited the UK Navy RCM team in the 1990s. I would like to thank her for time spent discussing the material in this book and for bringing so many real-world examples to my attention. Nancy's infectious enthusiasm gives her a unique ability to teach RCM concepts at every level.

I am grateful to my former colleagues at ISC Ltd (now owned by BAe Systems), who cultivated a grounded, open, real-world environment for the application of RCM principles. They include Noel Clarke, Dion Hoare, Colin McLewee, Kevin Weedon and Malcolm Yaldron.

Finally, special thanks go to those I worked with in the UK Royal Navy over nearly two decades to develop and apply RCM principles, including Roger Crouch, Ted Main, Andy Matters, Nigel Morris, Nigel Newling and Malcolm Regler.

Section 1

Principles

1 Hidden failures, Real Consequences

1.1 Functions and Failures

Up to the middle of the twentieth century, the focus of maintenance was the prevention of failure. Lubrication, overhauls and scheduled replacement of equipment were intended to prevent failures from happening. When failures did occur, often the response was to do more maintenance or to do it more frequently in the hope of preventing them in future.

By the 1960s the inadequacies of this approach were becoming obvious. The aviation industry discovered that doing more maintenance, or reducing maintenance task intervals, very often made no difference to failure rates. Far worse, and more surprisingly, more maintenance could sometimes increase failure rates rather than reduce them. A survey by United Airlines (Nowlan and Heap, 1978) found that 14% of items showed no relationship between age and chance of failure, but that 68% of items failed predominantly early in their life. At this point it was recognised that maintenance—or, at least, scheduled overhaul and replacement—is exactly the wrong way to prevent equipment failure.

If overhaul and replacement is the wrong solution, what is the right way to prevent failure? Reliability-centred Maintenance (RCM) was developed from the 1970s onwards in order to answer this question. The technique starts by focusing on the functions of equipment rather than on its failures, in other words, on what it *does* rather than what it *is*.

RCM is a systematic technique for generating a maintenance schedule. It begins by listing all the functions of the equipment under analysis, and then moves on to list all the ways in which it can fail (failure modes) and what happens when each failure occurs (failure effects). It then uses all the information collected to select an appropriate maintenance task to deal with the effects of the failure. Here is the difference between RCM-based maintenance and what preceded it: RCM deals with the *effects* of failure and focuses on maintaining *functions*; older maintenance methodologies try to prevent failure and focus on maintaining equipment.

The process used in RCM to select maintenance tasks begins by asking the question

"How does it matter if this failure occurs?"

Evident failures can matter in four ways.

Category	Description	Examples
Safety	A failure that could hurt or kill people	Leaking gasoline causes an explosion A worker falls from a corroded ladder An aircraft rudder failure results in a crash
Environmental	A failure that breaches an applicable environmental regulation	Oil leaking from an oil platform pollutes the sea Untreated effluent escaping into a river kills wildlife
Operational	A failure that affects production	A seized aircraft brake prevents it from moving Turbine failure shuts down a power station
Non-operational	The only effect is the cost of repair and secondary damage	A cooling pump fails, but a standby pump takes over immediately

There is one more category which differs fundamentally from the four categories above. Some failures have absolutely no effects at all when they happen. In fact, we have absolutely no idea that a failure has occurred. They almost all involve protective devices of some sort: fire alarms, trips, gas detectors, proximity alarms, pressure relief valves, standby pumps and generators, and so on. Failure of a simple fire alarm, for example, has no effects at all when it happens; it only matters if a second failure occurs (a fire).

Failures like this are called *hidden* failures because they only become evident if another abnormal event or failure occurs. The associated function is called a hidden function, and what it protects us against is called the initiating event, or sometimes the protected system. The initiating event may be the result of equipment failure, human error or negligence, a natural event (rain, earthquake and so on), or external failure (e.g. the power supply).

If the protective system fails and the initiating event occurs, the result is a *multiple failure*. A few examples of multiple failures are listed below.

Initiating Event	Protective System	Multiple Failure
Fire breaks out	Fire alarm	Fire occurs and the fire alarm does not sound. Occupants of the building are given less warning of a fire and may not be able to escape.
Boiler overpressure	Pressure relief valve	Excess pressure not relieved and continues to rise. Boiler may explode.
Personnel inside moving packing line	Emergency stop button	Someone is inside the moving machine and it cannot be stopped quickly. Personnel may be seriously injured.
Fan motor high vibration	Vibration trip	Fan motor vibration is high and it is not shut down automatically. Motor may be damaged or destroyed.

1.2 Why do Hidden Failures Matter?

As we have already seen, the truth is that a hidden failure does not matter at all. It matters so little, in fact, that by definition no one knows that the hidden function has failed. If a fire alarm fails, it doesn't matter; if a pressure relief valve is stuck closed, it doesn't matter. If a tank's ultimate level switch is stuck, it doesn't matter at all. Unless, of course, the event occurs that the hidden function is intended to protect us against. Then the hidden failure can make the difference between a minor embarrassment and a major disaster.

The purpose of this chapter is to learn about the maintenance and design of protective systems by analysing accidents and disasters in which they are somehow implicated.

The incidents described in this chapter were chosen because their consequences were so severe that they became global news and are still remembered years later. There is a very specific reason for including them in a text on hidden failures, because each of these incidents would not have happened, or at least would have been far less severe, if protective systems had worked as they were intended. The intention of this chapter is to provide some context for the theory of hidden failures outlined in the remainder of the book.

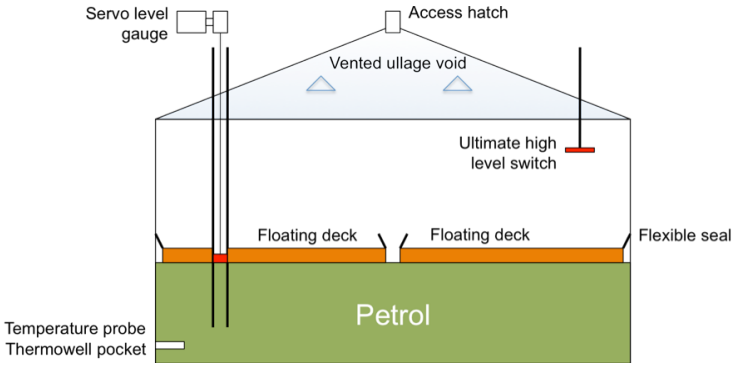
It is very easy to be seduced by theoretical models, but the accidents listed here should provide a sobering lesson. In each case, the equipment and systems were analysed in depth. In each case someone, somewhere in the organisation decided that their design, maintenance and operation provided acceptable protection for staff within the plant and for those living nearby. And in each case that analysis and sign-off was completely and absolutely wrong because something happened that did not fit the theoretical model.

1.3 Buncefield Storage Terminal

The first major incident is also one of the simplest. Atmospheric pressure storage tanks are found everywhere, and although the technology used for level measurement and remote valve operation has changed over the years, the basic principles have not. What differentiates the Buncefield incident from others is the scale of the ensuing consequences, and that but for extremely fortunate timing, tens or hundreds of people could have been injured or killed.

The Buncefield Hertfordshire Oil Storage Terminal is located about 3 miles from the town of Hemel Hempstead, England and 25 miles (40 km) north west of London. When the terminal was built in 1968 there was little development in the immediate area, but an industrial estate was built next to the plant, and by 2005 the area was surrounded by commercial and residential property.

The oil storage terminal supplied fuel to the London area and south east England. Fuel was delivered to the terminal in batches through three pipelines, then separated into tanks at the storage depot. Fuel left the depot by road tanker and through two pipelines, one to London Heathrow airport and another to London Gatwick airport.



Buncefield storage tank 912 layout

The primary means of measuring the level of fuel in the tank was a servo gauge which fed level information to the Automatic Tank Gauging (ATG) system. The ATG enabled operators to monitor tank levels, temperatures and valve positions throughout the site, and to operate tank valves remotely. The system stored several months' sensor and valve data in a large database.

The ATG provided high and high-high alarm levels which were intended to provide a visible and audible warning of high tank levels. Critically, the alarms did not have independent sensors: they derived their signal from the level control system. An additional, independent ultimate high level switch was designed to shut off delivery if the fuel reached a maximum tank level. Operation of the ultimate high level switch generated an audible and visual alarm in the local control room, and the trip was also transmitted to the pipeline operators.

Late on Saturday 10 December 2005, the terminal began to accept a pipeline delivery of unleaded petrol (gasoline) to tank 912. At about 03:00 the next morning, the ATG showed that the level was static at about 67% full, although post-incident review of SCADA records shows that delivery was continuing at a rate of about 550 cubic metres per hour.

The tank level continued to rise until it was above the ATG high and high-high alarm levels; because the level gauge was stuck, no alarm was raised in the control room. Later analysis estimated that the ultimate high level switch would have been reached at about 03:55, and the tank would have been full by about 05:20.

The tank began to overflow, forming a cloud that eventually extended over an area of about 80000 square metres, awaiting a source of ignition.

At 05:50 on 11 December, a tanker driver reported a strong smell of petrol at the loading bay. A few minutes later at 05:59, a supervisor contacted the control room to report a tank spill; by this time around 300 cubic metres of petrol would have escaped from the tank. Before any significant action could be taken, the vapour cloud encountered an ignition source, possibly a running vehicle engine, and ignited.

Seismographic sensors record a major explosion at 06:01:32 followed by a series of smaller explosions. The initial explosion was heard over 100 miles away from the site in much of southern England and northern France. The fire that followed engulfed 23 storage tanks on the site; it burned for five days and destroyed most of the site. There was serious structural damage to nearby homes and businesses, and buildings up to five miles away from the incident were damaged. 2000 residents were evacuated from their homes. 650 businesses on the adjacent Maylands Industrial Estate were severely disrupted.

Loss of the oil storage depot caused temporary disruption to fuel supplies in the area. London's Heathrow Airport was badly affected, losing 50% of its daily fuel requirement.

The total estimated cost of the incident was £900 million, with £625m in compensation claims and £245m impact on aviation.

No one was killed as the result of the incident. The legal judgement which apportioned damages and costs for the incident observed:

"*The failures which led in particular to the explosion were failures which could have combined to produce these consequences at almost any hour of any day. The fact that they did so at one minute past 6 on a Sunday morning was little short of miraculous.*" (Judiciary of England and Wales, 2010)

If the explosion had occurred during the working week, it is possible that tens or hundreds of people might have lost their lives.

An inquiry was opened in January 2006 to identify the causes of the incident. Its final report was published in 2008 and demonstrated the critical importance of the correct design and maintenance of protective systems.

The most obvious failure of the system's design is that the initial tank level alarms depended on the same servo sensor that transmitted tank level readings to the ATG system. When the level gauge stuck, it was guaranteed that these alarms would also be disabled. But the ultimate high level trip was designed to be independent of the level gauge. Why did it not operate when the tank level reached it?

Testing high level trip switches thoroughly can be difficult. A complete test would include raising the tank level until it reaches the trip switch, then observing that all the expected shutdown systems operate correctly. The test is likely to be disruptive because of the time taken to fill the tank and to return it to normal levels after the test. Worse, simulation of high tank levels might lead to unintended overfilling of the tank if the trip system does not operate correctly. The switch used in the Buncefield tank provided a plate or lever which allowed a technician to simulate a high tank level and to test the shutdown system without needing to fill the tank. However, using the switch in its test mode disabled its normal function: it was essential to return it to the "normal" position after the test.

The switches on tanks 911 and 912 had been replaced, but the maintenance contractor did not appreciate that they were not like-for-like units. These switches included a test mechanism and a padlock which was to be used to lock the mechanism during normal operation. Instructions about unlocking and locking the padlock were not routinely supplied by the switch manufacturer; even when the were supplied, they did not point out the critical importance of the padlock. Users were not told that the switch would not work at all if the lever was left even slightly below the horizontal.

It seems likely that tank 912's ultimate level switch was disabled because its padlock had not been put in place after testing. Of course, that did not matter at all until the level gauge stuck.

Initiating Incident

Failure of a petrol (gasoline) storage tank's level control system.

Protective Device Failures

Protective Device	Failure	Consequence
ATG High level alarm	Not functional because its signal was provided by stuck level gauge	Tank level rose above the high level
ATG High high level alarm	Not functional because its signal was provided by stuck level gauge	Tank level rose above the high high level
Ultimate high level switch	Disabled: test plate probably left in "test" position or its padlock was not used	Tank level rose above its ultimate high level, allowing petrol to escape through the tank breather holes and finally ignite

1.4 Three Mile Island

The Buncefield incident represents failure of a very simple protective system with spectacularly severe consequences.

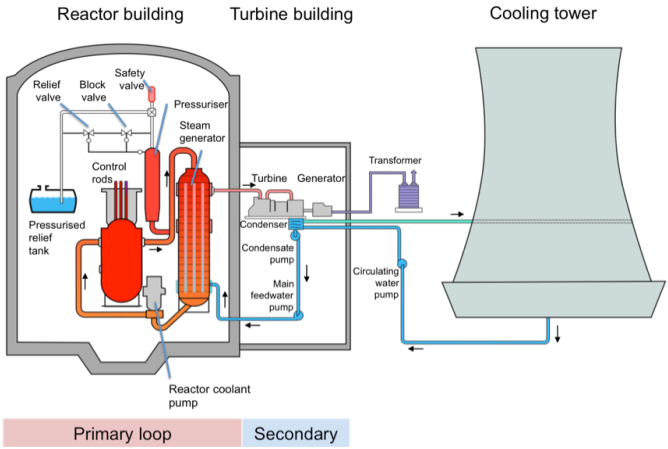
Nuclear reactors bring together a vast number of potential safety and environmental hazards: high power output in a small space; high pressure superheated water and steam; and, of course, radioactivity and the possibility of a runaway nuclear reaction. Nuclear power plants are designed with multiply redundant systems, comprehensive instrumentation, and alarms and trips to provide the best possible defence against human and equipment failure.

In many ways the Three Mile Island incident is similar to that at Buncefield, because at its core is the hidden failure of a single relief valve. Where it differs is in the complexity of the reactor design, with interconnected redundant systems, instrumentation, alarms and trips so complex that the operators struggled to understand and control the crisis. With hindsight, the cause may seem obvious; but pay particular attention to the timeline and it becomes evident just how much pressure they were facing.

On 8 March 1979, an incident at reactor TMI-2 on Three Mile Island, Harrisburg, Pennsylvania, cut through all its levels of defence and made the name synonymous with nuclear near-disaster.

The installation at Three Mile Island consisted of two pressurised water reactors TMI-1 and TMI-2. On the day of the disaster, TMI-1 was shut down for refuelling and TMI-2 was operating at close to full power.

In a pressurised water design, heat from the nuclear reactor produces steam to generate power in two steps. In the primary loop, water enters the reactor at about 275°C and is heated to about 315°C. The water remains liquid because the primary coolant loop is held at a pressure of about 150 bar (2200 psi). After leaving the reactor, water flows through the steam generator, which heats water in the low pressure secondary loop to generate steam, driving a turbine and turning the generator. The steam is then condensed, cleaned and the condensate is returned to the secondary loop.



Three Mile Island Reactor TMI-2 Simplified Schematic (US NRC, 2009)

At about 04:00 on 8 March 1979, TMI-2's condensate polishing system pumps stopped running for reasons that have never become clear. The polishing system in the secondary loop filtered and removed ions from condensate, maintaining the water at close to the purity of distilled water. Loss of water from the polishers set off an automatic cascade of trips: first of the steam generator main feed water pumps, then of the turbine itself.

When the turbine tripped, auxiliary feed pumps started automatically to provide water to the steam generator. However, the valves to the auxiliary pumps had been left closed after maintenance, so no water flowed. As a result, the secondary loop was no longer able to remove heat from the primary loop, and the temperature and pressure in the loop began to rise quickly. Finally, eight seconds after the initial trip, the reactor shut down ("scrammed") automatically.

In a non-nuclear generation system, that would probably have been the end of the incident: embarrassing, certainly, but easily recovered. But shutting down a nuclear reactor is not so simple, because it continues to produce heat even after the basic reaction has stopped. Scramming a reactor reduces the uranium fission rate by inserting neutron-absorbing material so that fission processes are halted. While uranium fission produces most of a reactor's heat output, there is a second source, because breaking uranium nuclei produces radioactive fragments that in turn generate heat as they decay. This is why spent nuclear fuel rods need to be cooled for months or even years after they are removed from a reactor. The power generated is significant: immediately after shutdown, a reactor can continue to generate 7% of its rated power because of decay heat, and it still produces 1%–2% of its full power after an hour. Even after the reactor has been shut down, continued coolant circulation is essential.

Now that no heat was being removed from the core, the primary loop pressure and temperature continued to rise because of decay heat. The pilot operated relief valve (PORV) in the primary coolant loop opened to relieve the excess pressure that had been generated. A few seconds later, when the pressure and temperature had fallen, the PORV should have closed, but instead the valve stuck open.

As we will see, this failure was central to the drama that was about to unfold.

At this moment the reactor operators were faced with a mass of instrument readings, alarm and trip warnings, including a light that showed the open PORV. What they did not recall—or perhaps did not know—was that the PORV light did not reflect the position of the valve, but just the presence of power on the PORV solenoid. Under normal circumstances, of course, absence of power on the solenoid meant that the valve was closed; but on 8 March the valve was stuck open while the operators assumed that the lamp meant that it was closed. From this point on, no one knew that water was being lost continuously from the primary coolant circuit through the PORV.

The operators' assumptions about the PORV position proved to be critical. They were now faced with seemingly contradictory information about the primary loop: although the reactor pressure was low, the water level in the pressuriser was high. The pressuriser controlled the primary loop pressure, and it was important that it contained both water and steam. The staff on duty seem to have been concerned that, if pressuriser water levels rose too far, they would lose control of the primary loop pressure. While they thought that the water level was high, what was really happening was that coolant was flowing *through* the pressuriser and out of the PORV. Two minutes into the incident, while the operators were trying to reconcile contradictory readings from the primary loop, the emergency water injection pumps cut in automatically to maintain the core coolant level.

The operators were still focused on the apparent rise in water levels, and they were now even more concerned about the coolant level. With the emergency injection pumps operating, they assumed that more coolant would be flooding into the primary loop, and that the pressuriser water level would continue to rise. At four and a half minutes into the incident, a supervisor turned off one of the injection pumps and cut back flow from the other.

After eight minutes one of the operators noticed that the secondary loop auxiliary pump valves were closed and he opened them; the secondary loop was now working normally, but coolant was still flowing out of the primary loop.

Escaping coolant from the primary loop filled the quench tank that collected the PORV discharge overfilled and then filled the containment building sump. This was an obvious sign of coolant loss, but operators ignored it because they still firmly believed that the PORV was closed. At 04:15 the quench tank relief diaphragm ruptured and coolant leaked into the containment building. It was then pumped from the containment building sump to auxiliary building outside containment until the sump pumps stopped at 0439.

After an hour and twenty minutes, the primary loop circulation pumps began to vibrate and two were switched off; twenty minutes later, the remaining two were stopped. Unknown to the operators, vibration was caused by steam passing through the primary loop. With no circulation, water now boiled in the core and continued to escape through the open PORV.

After 130 minutes, the water level dropped far enough to expose the reactor core. Steam reacted with Zorcalloy fuel rod cladding to produce hydrogen; fuel pellets were damaged and radioactive fission products escaped into the coolant and from there through the open PORV. Still the operating crew was unaware of the coolant loss.

The shift change at 06:00 brought fresh minds to the problem. An operator noticed that the temperature downstream of the PORV was high, diagnosed a coolant leak, and shut the block valve. The leak was now over, but 130 cubic metres of radioactive water had been lost. After 165 minutes the radiation alarms activated, and at 06:56 a site emergency was declared.

Even now, the emergency was far from over. The core had sustained extensive damage; it was later estimated that about 50% of the core had melted. High pressure now prevented coolant from being pumped into the core, so after 7 hours a backup relief valve was opened to allow the loop to be filled with water. After 16 hours, the primary loop pumps were started and the core temperature started to fall. For days afterwards, the threat of a hydrogen explosion remained; in the worst case, such an explosion might have breached the primary containment vessel and spewed fuel and fission products into the environment. On the third day hydrogen was vented to atmosphere and the immediate crisis was finally over.

Cleanup of the TMI-2 reactor took 14 years, from 1979 to 1993 and cost $975m.

Initiating Incident

Condensate polishing pumps stopped for reasons that are not known, causing a cascade of equipment trips.

Protective Device Failures

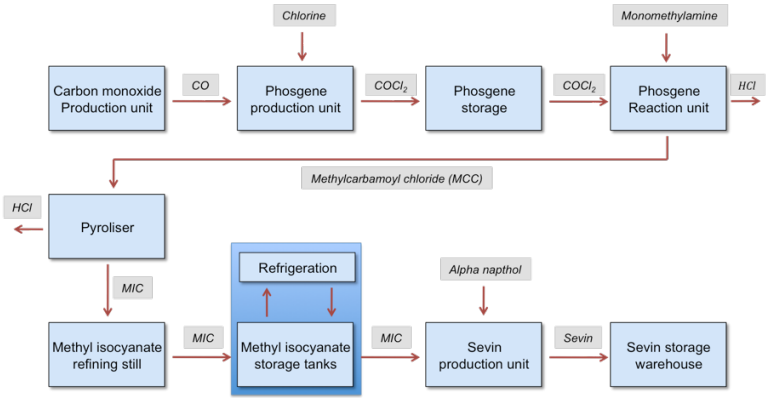
Protective Device	Failure	Consequence
Auxiliary feed water pumps	Valve closed because of maintenance	Secondary loop circulation failure. Reactor scram.
Pilot Operated Relief Valve (PORV)	Did not seat after relieving primary circuit pressure[1]	Severe operator confusion. Loss of 130 cubic metres of primary coolant

1.5 Bhopal

The Bhopal pesticide plant was operated by a Union Carbide of India Ltd to produce Sevin and other carbamate pesticides from components supplied to the plant. The plant design was based on optimistic projections of Asian demand. Its capacity on opening in 1980 was 5250 tonnes per year, but it was soon recognised that the market for its products was more challenging than had been expected, and the plant was modified to produce many of the precursor chemicals needed for Sevin synthesis in order to reduce costs. Low demand continued to threaten the viability of the plant and by 1984 it was operating at only about 25% of its full capacity (Fortun, 2001).

The plant was built in the northern part of Bhopal in what was at the time a relatively unpopulated area. By 1984, uncontrolled development had brought slum housing right up to the southern plant perimeter.

[1] The Three Mile Island incident is not unique. Another incident involving a relief valve held open by a control system is described in *Normal Accidents* by Charles Perrow (1984)

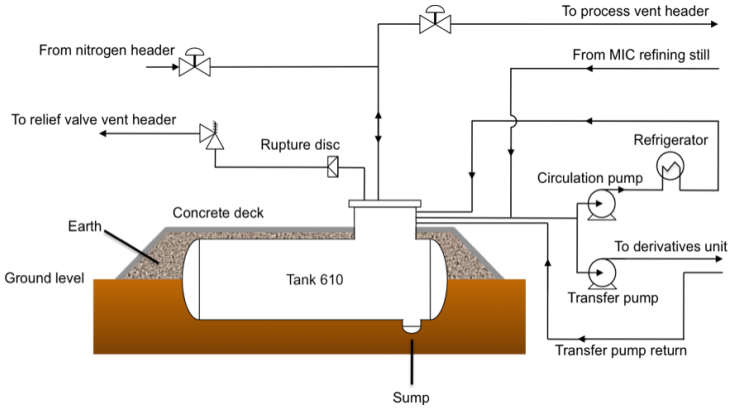


Bhopal Sevin synthesis route
The methyl isocyanate (MIC) storage system is highlighted

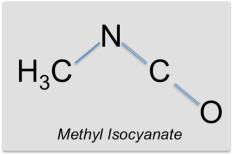
Sevin pesticide was produced in a series of steps.

- Carbon monoxide, produced on site, was reacted with chlorine to produce phosgene
- Phosgene and methylamine reacted to produce methylcarbamoyl chloride and hydrogen chloride
- Methylcarbamoyl chloride was pyrolised (decomposed at high temperature in the absence of oxygen) to produce methyl isocyanate
- Methyl isocyanate was distilled and then stored
- Batches of methyl isocyanate were fed to the Sevin production unit, where they were reacted with alpha napthol to produce the final product

Methyl isocyanate (MIC) is a colourless, volatile liquid. It is unstable and liberates large amounts of heat when it breaks down, so it is usually stored at around 0°C. The effects of MIC exposure on humans are very unpleasant: it attacks skin, eyes, the lungs and internal organs. MIC is more lethal than phosgene, which is well known for its use in World War I poison gas attacks.



Schematic diagram of Bhopal methyl isocyanate storage tank 610



Methyl isocyanate was stored in three identical stainless steel tanks, each with a volume of about 55 cubic metres. The tanks were partly buried, an earth mound covered the upper part of the tank, and a concrete deck was constructed on top.

Because of the extreme instability of methyl isocyanate and the possibility of a runaway reaction, each tank included a refrigeration unit and circulation system that were intended to maintain a liquid temperature of 5°C. One tonne batches of MIC were transferred to the Sevin production area by pressurising the storage tank to about 1 bar (14 psi) with nitrogen gas. The operating manual stated that the MIC level should be kept below 60% of tank capacity, apparently to allow for the possibility of pressure excursions.

A number of safety systems provided defence against venting MIC to the atmosphere. If the tank pressure rose, for example because of unexpected decomposition of MIC, a rupture disc and relief valve allowed the vapour to pass through to a scrubber and flare stack before opening to the atmosphere. The vent gas scrubber was a 1.7m diameter tower 18m high which constantly circulated a solution of caustic soda that would neutralise the gas. If the caustic soda solution flow dropped an auxiliary pump started automatically.

The flare tower burned vent gases from the carbon monoxide unit, MMA vaporiser safety valve and MIC refining still. It also burned gas from MIC storage tanks arriving directly or through the vent gas scrubber. The flare tower included a shielded pilot flame and flame front generator so that pilot could be re-lit.

The MIC storage tank refrigeration system was shut down in June 1984, apparently in order to save money, so the MIC temperature was now between 15°C and 20°C instead of the usual 0°C-5°C. To avoid the inevitable alarms, the high temperature alert was disconnected rather than being reset to a higher temperature. The MIC refrigerator's coolant was used elsewhere on the site.

On 23 October the MIC production unit was shut down. The vent gas scrubber circulation pump was set to standby with the result that caustic soda circulation would only restart under manual control.

At some time in October maintenance started on the flare stack to replace a section of corroded pipe.

By 1 December 1984, all the elements were in place for the ensuing disaster. Methyl isocyanate tank 610 contained about 41 tonnes of liquid, well above the maximum 60% tank level. The refrigeration system had been shut down for months, so the liquid was warm and there was no possibility of controlling a runaway reaction. With the exception of the bursting disc and safety valve, *all* the protective safety systems were disabled or missing. The tank temperature alarm was disabled; the scrubber system required manual intervention to start it; and the flare stack was still dismantled because maintenance started in October had not been completed. Finally, and perhaps worst of all, the plant was now close to crowded, poor quality housing.

Before the evening shift change on 2 December, tank 610 contained about 41 tonnes of MIC at a pressure of 1.1 bar. At some point between 500 and 1000 kg of water were introduced into the tank. Exactly how this happened has never been determined with any certainty. It may have been the result of water washing production piping (a standard procedure) carried out at 21:00 on the same day; it is known that on this occasion no slip blind was used to prevent water entering the MIC storage area. Deliberate sabotage has been suggested. However it happened, we do know that water entered tank 610 and started to react with the methyl isocyanate.

At 23:00 on 2 December 1984, just after the shift change, tank pressure had increased to 1.7 bar. Because this was still within the normal limits of 1.1-2.7 bar, it seems that the new shift did not recognise that pressure had increased fairly rapidly. There was no equipment that gave the operators a history of temperature and pressure readings.

At about 23:30 workers noticed a smell of methyl isocyanate and found a leak near the scrubber. Dirty water and MIC was leaking from a branch of the relief valve pipe downstream of the safety valve.

Tank pressure continued to rise, and at 00:15 on 3 December a supervisor started the vent gas scrubber circulation pumps. There was no flow indication. What the operators did not know was that even a working scrubber would have been incapable of neutralising completely the volumes of gas coming from the tank.

At 00:30 the tank pressure gauge reached its maximum reading of 3.8 bar. The control room operator walked to the tank area to check local indicators on the tank. He heard rumbling from the tank, a screeching relief valve, and felt radiated heat.

The safety relief valve had opened at 3.5 bar, as it had been designed to do. With no other protective systems operational, a jet of methyl isocyanate shot up the scrubber tower and escaped to atmosphere from the disabled flare stack.

The external alarm was sounded to warn the local neighbourhood, but it was then turned off to avoid panic. At 00:50 the plant alarm sounded and workers escaped upwind. A fire squad arrived and began to spray the flare tower, but the water fell well short of the top of the stack. They then sprayed the tank hoping to cool it.

Tank 610 expanded, burst its concrete casing and toppled over. A second pipe ruptured and released MIC to the atmosphere.

Between 01:30 and 02:30 the tank pressure began to drop and the safety valve reseated; by 04:00 the gases were finally brought under control.

At around 02:30 the plant external siren, used for warning local residents, had been sounded again. By then the smell of gas had been obvious for over an hour.

Methyl isocyanate vapour is twice as dense as air, so when the tank began to vent a cloud drifted down to the ground. Unfortunately there was a light north-westerly wind which blew the cloud toward the city. The composition of the escaping gases is not certain, because MIC should have decomposed at high temperature into metylamine and hydrogen cyanide. People ran from the local area; by this time many were suffering from chemical burns to their eyes and lungs. Some were simply trampled in the stampede to escape. Local medical services were overwhelmed, and in any case doctors had little or no information on how to deal with the effects of MIC inhalation.

About 3800 people in the slum colony around plant died in the immediate aftermath of the disaster. It has been estimated that 20000 or more premature deaths occurred in the following 10 years and 100000-200000 people sustained permanent injuries. In a settlement reached in 1989, Union Carbide paid $470m in damages to claimants.

More than twenty-five years after the Bhopal disaster, no one knows exactly how a cubic metre of water entered tank 610. What is absolutely certain, however, is that the consequences could have been very different if any of the associated protective systems had been working.

Initiating Incident

Up to one cubic metre of water entered the methyl isocyanate storage tank causing a runaway reaction. How the water was able to enter the tank is unknown.

Protective Device Failures

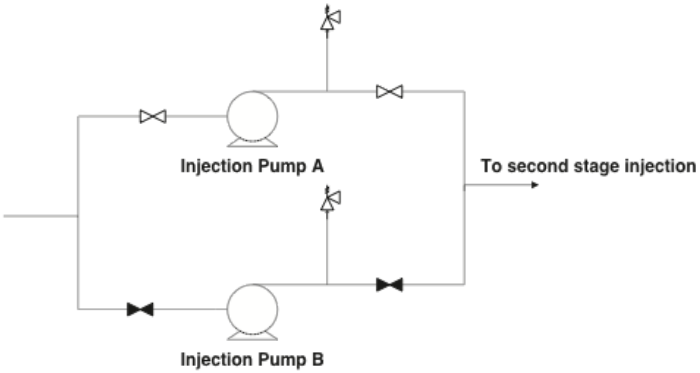
Protective Device	Failure	Consequence
High temperature alarm	Deliberately disabled because the refrigeration unit had been deactivated	Operators had no warning of the MIC reaction with water until tank pressure started to rise
Vent gas scrubber	Unavailable at the beginning of the incident because set to manual start	MIC vented directly to atmosphere
Vent gas scrubber	Incapable of neutralising gas	Even with the vent gas scrubber operating, all the gas could not be neutralised
Flare gas stack	Partially dismantled	Not available to flare gas; MIC vented directly to atmosphere.

1.6 Piper Alpha

Over twenty years after the platform was destroyed, Piper Alpha is still remembered as one of the worst ever incidents to occur in the offshore oil industry. Not only are faulty protective systems largely responsible for the scale of the disaster, but maintenance of a pressure relief valve is a central cause of the incident. For anyone who believes that "more maintenance is better", it is worth considering that 167 men would not have lost their lives if the relief valve had *not* been removed for maintenance.

Piper Alpha was a fixed offshore oil production platform operated by Occidental Petroleum in the North Sea, about 120 miles (200 km) north east of Aberdeen, Scotland. Oil production started in 1976, and the platform was responsible at one point for around 10% of all UK oil production. Initially Piper Alpha produced only oil; in 1978 it was modified to export small quantities of gas. The non-methane gases (mainly butane and propane) were compressed and injected into the oil export pipeline.

The incident that destroyed Piper Alpha began on 6 July 1988 when the first stage condensate injection pump A was isolated in preparation for maintenance on its coupling. Condensate production continued using the second pump B. While pump A was isolated, an opportunity was taken to remove its associated relief valve for routine maintenance. It is likely, but not absolutely certain, that a flange was fitted in place of the missing relief valve.



Simplified Piper Alpha first stage injection process and instrumentation

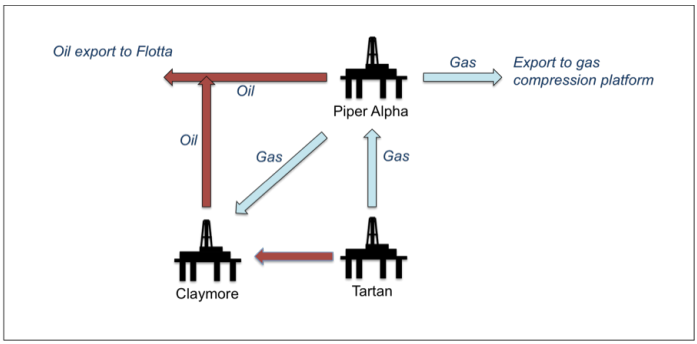
Later in the evening of 6 July, injection pump B tripped. The operators tried several times to restart the pump but were unsuccessful.

The platform's design meant that failure of condensate injection would eventually lead to a shutdown of both oil and gas production, so the operators knew that it was critical to restart injection if at all possible. Maintenance on pump A's coupling had not been started, so they isolated the faulty pump B and restarted pump A. At this point Piper Alpha's permit system plays a key role because it was organised by physical location, and the relief valves were in a different compartment from the injection pumps. As a result he permit for pump maintenance was separated from the permit that would have shown that the relief valve was missing. Additionally, although the pump's status had been mentioned at shift changeover, it appears that nothing was said about the relief valve. The operators seem to have been completely unaware that there was no relief valve on the line.

Pump A was restarted at about 21:55. With no relief valve to contain the condensate, it escaped under high pressure from the flange where its relief valve should have been.

Six gas alarms were triggered, but so much condensate escaped that it ignited before any preventative action could be taken. The resulting explosion blew through the firewall and started more fires. The Custodian operated the emergency stop button, halting Piper Alpha's production and isolating the platform. The control room was abandoned a few minutes later.

The fire deluge system should have started automatically to fight the fire, but it did not operate at all. The incident inquiry later discovered that it had been set to manual mode earlier in the day in order to protect divers who had been working under the platform. It had not been switched back to automatic mode when the work was completed.



Piper Alpha oil and gas export network

If Piper Alpha had been the only source of fuel, the fire would eventually have burnt itself out when its production had been isolated. However, it was part of a network of gas and oil pipelines from other platforms, and their operators assumed that Piper Alpha would request them to halt production in an emergency. The high cost of shutting down and restarting production meant that the operators on the Tartan and Claymore platforms were reluctant to shut down production. What they did not know was that the explosion on Piper Alpha had destroyed its communications, so they continued to export oil and gas. This forced fuel back out of the ruptured pipework on Piper Alpha and fed the fires. Within the next half hour, gas pipelines ruptured and massive explosions destroyed the platform. By midnight about three quarters of the platform had sunk.

Of the 224 staff who were on the platform on 6 July, 165 lost their lives; two men aboard a support vessel were also killed in the incident.

A detailed incident inquiry under Lord Cullen began in 1988 and produced a detailed report in 1990. The report made 109 recommendations whose implementation changed fundamentally the safety culture of the UK offshore industry.

Initiating Incident

Failure of a standby condensate pump, causing the operators to switch to a pump whose relief valve was missing.

Protective Device Failures

Protective Device	Failure	Consequence
Pump A Relief valve	Device removed for maintenance	Condensate leak at flange
Gas alarms	Functioned correctly, but insufficient time available to prevent an explosion	Gas cloud spread and ignited
Manual emergency shut down	Functioned correctly	Halted Piper Alpha's production, but flow from other platforms continued
Fire deluge system	Did not function. Incorrectly left in manual mode after earlier diving work	Platform fire spread unchecked
Emergency inter-platform communication	Disabled by the initial incident	Oil and gas from other platforms continued to feed the fire on Piper Alpha even when local production had been shut down

1.7 Chernobyl

Chernobyl is now synonymous with nuclear disaster, and the 1986 incident remains one of the most serious in the industry. At the heart of the accident was testing of a protective system.

The reactor was cooled by water flowing through the reactor core. If the reactor were to be scrammed (i.e. shut down in an emergency), it would still require coolant flow to remove heat, and there was concern that external power might not be available to run the pumps. The reactor had three backup diesel generators, but they would need over a minute to run up and supply enough power to run a cooling pump.

To supply power while the diesel generators were running up to speed, engineers proposed using energy from the steam turbine, which would be running down after the reactor scram. Tests carried out in 1982, 1984 and 1985 had been unsuccessful because the turbo-generator had been unable to provide enough power, and a further test was scheduled before shutting down reactor 4 for maintenance. The test was not intended to simulate exactly a loss of external power; instead, the reactor would be run at low power with the steam turbine running at full speed. The steam supply would be turned off, and the generator output measured during the turbine's free wheel.

During the night shift on 25-26 April 1986, the power output from the reactor was reduced to 700-1000 MW in preparation for the test. At low power, xenon 135 gas built up in the fuel rods, absorbing neutrons and depressing the nuclear reaction; as a result, the reactor power dropped further*. The operator noticed the power reduction and for reasons that are not fully understood, inserted the control rods too far and reduced the power to an almost complete shutdown. The output power was now far too low to carry out the test safely, so operators decided to extract the control rods and increase the reactor's power output. Running the reactor at low power had led to accumulation of xenon in the fuel rods, so many of the control rods had to be fully withdrawn to restore power output.

After some time, reactor thermal power output was stabilised at about 200MW. Although this was far less than the 700MW specified in the test schedule, preparations were made for the test. At 01:05 operators increased the coolant flow rate through the reactor core. Since water is a neutron absorber, the effect was to reduce reactor power output again. Now reactor output was suppressed by two factors: by accumulated xenon 135 and by additional coolant. The operators do not appear to have understood that the reactor's output was suppressed by xenon accumulation, because they withdraw almost all the control rods to maintain reactor power.

* Xenon is produced from iodine 135, one of the common fission products. Its half-life is 6.7 hours and decays into xenon 135, with a half life of 9.2 hours. Xenon 135 has a very large cross-section for neutron absorption (3 million barns, compared with about 500 barns for uranium). A high neutron flux is needed to "burn away" the xenon 135, so it accumulates at low reactor power levels.

The test was started at 01:23. Steam to the turbine was shut off and the number of feed water pumps reduced from eight to four. The reduced water flow rate caused water in the reactor to boil, forming steam bubbles. Because steam is so much less dense with water, the process has an inherent instability: fewer neutrons are absorbed by steam than by water, which increases the number of steam bubbles, which In turn causes the reactor output to rise. Boiling of the cooling water was expected in this reactor design, and the control system was designed to insert the control rods automatically to compensate for rising output power. However, in this case the power rose for two reasons: first, the water was beginning to boil; and second, the higher neutron flux was "burning off" the accumulated xenon 135. Both of these caused positive feedback.

At some point it appears that the operators reacted to the rapidly rising power levels by initiating a manual reactor scram. Scramming this reactor was not an instantaneous process: the control rods took around 20 seconds to achieve full insertion, compared with less than four seconds for a typical European or US reactor. Unfortunately one of the peculiarities of the reactor design was that water coolant was displaced by the control rods before neutron-absorbing material was inserted, so the initial effect of inserting the control rods was to *increase* the power output of the lower part of the reactor.

The reactor power rose very quickly and an explosion occurred, breaking fuel rods broke and preventing movement of the control rods. With reactor output at around 30GW, is then thought that a steam explosion destroyed the reactor casing and blew off the upper shield, which weighed about 2000 tonnes, and exposing the reactor core. A second explosion is thought to have been caused by a nuclear transient limited to part of the core.

In the immediate aftermath of the event, the reactor crew seems to have been oblivious to the loss of reactor containment, choosing to believe that "off the scale" dosimeter readings were the result of faulty measuring equipment. Fire fighters were unaware of the immediate danger, but extinguished fires on the roof and around the building to protect the number 3 reactor. The fire inside the number 4 reactor continued until 10 May when it was extinguished by helicopters dropping neutron absorbing materials from helicopters.

31 people died within the first three months; they were mostly reactor staff, fire and rescue workers. 135000 people were evacuated from the local area and approximately 131000 square kilometres were contaminated by radioactive material. There is considerable uncertainty about the long term effects on life expectancy and health, but UN estimates suggest 8000-10000 cases of thyroid cancer may result (UNDP and UNICEF, 2002).

Initiating Incident

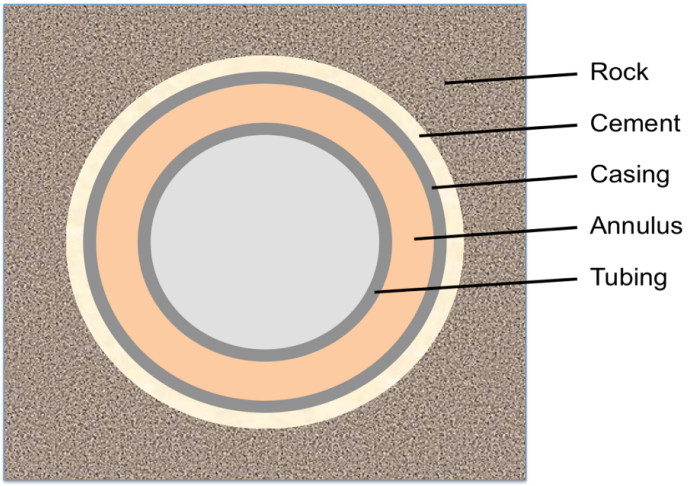
Test of emergency power system with the reactor in a low output state.

Protective Device Failures

Protective Device	Failure	Consequence
Steam turbine generator residual power	Reactor in unstable state during test	Reactor power rose uncontrollably
Reactor scram	Slow scram by design. Graphite displaced water, increasing the lower reactor power output	Rapid increase in output power; explosion; containment lost

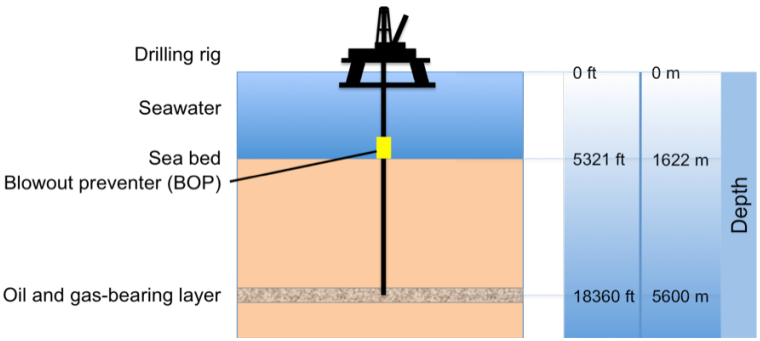
1.8 Deepwater Horizon

In February 2010 Deepwater Horizon, a semisubmersible drilling platform owned by Transocean and under lease to BP, started exploratory drilling for oil about 40 miles off the coast of Louisiana in over 5000 feet of water. The site was not only difficult because of the sea depth; including the depth under the sea floor, the total drill bore was expected to be over 19000 feet in length. After completion of the exploratory well's casing and cementing, it would normally have been tested and plugged before being abandoned to await future production activity.



Simplified cross-section through a production well

Oil and gas reservoirs can be under very high pressure, and controlling flow to the surface is one of the greatest challenges facing oil exploration. If a hole were cut through rock into a pressurised well with no control, oil and gas would escape under very high pressure through the bore hole to the surface. Well pressure can eject piping and tools at high speed, and escaping gas poses an obvious and immediate explosion hazard.



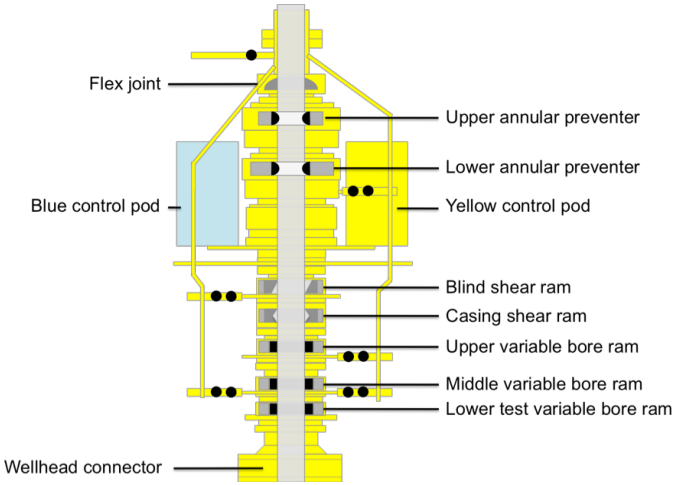
Schematic layout of the Deepwater Horizon and the Macondo well

Liquid "mud" is pumped down the inner bore to the drill head during drilling. It returns through the annular space between the tubing and casing, bringing with it rock cuttings that are removed at the surface. Drilling mud is actually a complex mixture consisting of a base fluid (water, oil or synthetic) with clays and chemicals. As well as bringing cuttings to the surface, the mud flow lubricates and cools the drill bit. It also plays a key role in controlling well pressure: mud density is chosen so that its weight balances the well pressure, preventing uncontrolled escape of oil and gas from the reservoir. A badly behaved well can turn drilling into a constant battle between the wellbore and the weight of mud above it.

Drilling into hydrocarbon-bearing layer is sometimes compared with puncturing a balloon or a car tyre, but in reality well behaviour is far less predictable than that of an air-filled rubber tube. As drilling progresses, well pressure can vary widely. The drilling crew tries to balance the well as its pressure varies, but sometimes the pressure changes very rapidly; a short high pressure transient is generally called a "kick". A sustained pressure excursion can result in a blowout, where drilling fluids and even equipment may be ejected from the borehole and the uncontrolled escape of gas and oil may lead to fire and explosion hazards.

A blowout preventer protects drillers from sudden pressure changes by limiting flow or by closing off the well completely. The blowout preventer installed on the Macondo well included three levels of protection.

Blind shear ram	Capable of cutting the drill pipe and sealing the well
Casing shear ram	Capable of cutting the drill pipe, casing and tool joints, but not able to seal the wellbore
Upper, middle and lower variable bore rams	Able to close the annulus and seal against the inner tubing



Macondo well blowout preventer

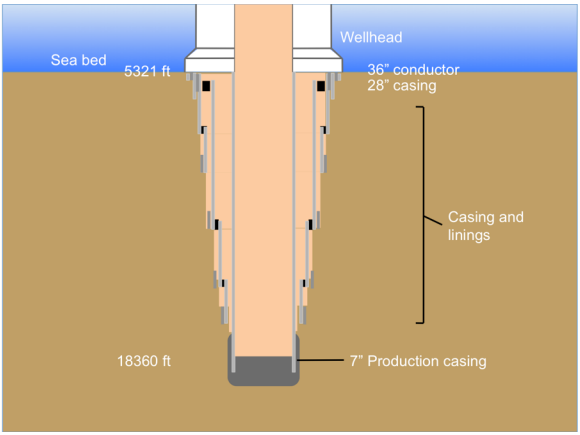
Hydraulic power was provided from the surface to one of the blowout preventer's control pods. Operation of the blowout preventer was controlled from two locations on the rig: the driller's cabin and the bridge. Modules in the preventer received commands from the surface through two independent cables, one to each control pod, and activated the appropriate solenoid valves. There were two power supplies in each electronic module and battery backup in case the surface power supply failed.

The blowout preventer contained eight 80-gallon (360 litre) 5000 psi accumulators which should have been capable of providing hydraulic power during normal or fail-safe operation. The preventer could be operated manually from one of the control panels, but it also had three emergency modes.

- A manual emergency disconnect sequence initiated from the rig
- The automatic mode function (AMF), a fail-safe system which operated automatically if communication, electrical power and hydraulic power from the surface were lost. This was intended to seal the well automatically if the rig were disabled or if it drifted off position.
- An auto-shear function which had to be initiated from a remotely operated vehicle (ROV) on the sea floor

The first two emergency modes should have prevented or mitigated the effects of a blowout; ROV operation could also shut off the well, but would only be used to stop the uncontrolled flow of oil and gas into the sea after a serious incident had already occurred.

In summary, the blowout preventer was designed to be a multiply-redundant, fail-safe protective device with multiple levels of protection. As we shall see, "fail-safe" does not mean failure-free.



Deepwater Horizon Macondo well

The Deepwater Horizon incident occurred when work on the well was substantially complete and rig crew were preparing to abandon the well. On 19 April 2010 cement was pumped down the production casing and into the annulus to prevent oil and gas from the reservoir from entering the wellbore. Before abandoning the well, it was necessary to test that the cement sealing the production casing and annulus was secure. The seal was tested under positive and negative pressure to ensure that a complete seal had been made at the end of the production casing. The negative pressure test entailed replacing heavy drilling mud with lighter sea water; if the seal were not effective, hydrocarbons would enter the bore or annulus. According to the BP incident report, pressure and volume readings indicated that the barriers were not effective; however for some reason the rig crew and BP staff incorrectly assumed that well integrity had been proven. Having carried out the negative pressure test, sea water was replaced with drilling mud in order to overbalance the well; this temporarily hid any problems with the cement barriers.

At 20:02 on 20 April, as part of the process leading up to abandoning the exploratory well, drilling mud was again replaced with seawater. At 20:52 there was evidence of flow from the well, but this appears to have been masked by emptying of a trip tank (a small tank used to measure the amount of mud needed to keep the wellbore full). After 21:00, drill pipe pressure continued to increase with pumps shut off, indicating flow from the reservoir into the well.

Oil and gas were now moving up the well, and at about 21:40 displaced mud overflowed onto the rig floor and then shot up through the derrick. At this point it is believed that the drilling crew tried to close the BOP's lower annular preventer. Mud from the well was diverted to the mud-gas separator, which normally removed relatively small quantities of dissolved gas from drilling mud. Drill pipe pressure continued to increase, and the mud-gas separator was overwhelmed by the flow rate. At about 21:46, high pressure gas began to escape from the mud-gas separator vents toward the deck, setting off gas alarms. A minute later the drill pipe pressure increased rapidly from 1200 to 5730 psi which may have been the result of the BOP sealing around the pipe.

Large volumes of gas were now spreading over the rig and into electrically unclassified areas where they could find a source of ignition. The gas cloud caused entered the main power generation engines' air intake and caused an over-speed; electrical power was lost. A few seconds later at 21:59 the first explosion occurred, followed almost immediately by a second.

After the explosions, at 21:52, the subsea supervisor attempted to operate the BOP's emergency disconnect sequence to seal the well. It is likely that the attempt was unsuccessful because communications had been destroyed by the explosions. A mayday call was transmitted.

115 personnel were transferred to a rescue vessel. 17 were injured in the incident and 11 killed. The consequences did not end at this point because the blowout preventer had not sealed off the well; oil and gas continued to flow freely into the sea. Attempts were made during the period from 21 April to 5 May to engage the BOP's third emergency shutdown function from a remotely operated vehicle.

Engineers intuitively assumed that the blind shear ram had partially operated, but had been obstructed or that it had crimped the pipe but not sheared it. In response, pressurised hydraulic fluid was injected by a submersible, but the hydraulic system leaked and needed multiple attempts to seal it. Failure of the hydraulic system shocked the engineers because it had been subject to very frequent, strict leak tests. Finally, with the hydraulic leaks fixed, the submersible was able to apply the full 5000 psi hydraulic pressure to the blades, but with no sign of movement. Gamma ray imaging of the blowout preventer showed the true internal picture of the blowout preventer: one blade had deployed, but there were no remaining options for forcing the other closed.

Oil continued to flow until the well was finally capped on 4 August. Up to 4 million barrels of oil flowed into the ocean, closing 86000 square miles of fisheries in the most severe US environmental incident. The total financial loss has been estimated at $30 billion.

Initiating Incident

Defective well cement.

Protective Device Failures

Protective Device	Failure	Consequence
Blowout preventer annular preventer	Operated by crew after uncontrolled mud spill on rig floor. Did not seal immediately around drill pipe.	Mud and gas escape onto rig Gas escapes outside electrically classified areas. Explosion and fire. 11 personnel killed, 17 injured.
Fire and gas system	General audible and visible gas alarm may have been inhibited (BOEMOE, 2011)	Less time for personnel to respond
Blowout preventer Emergency Disconnect Sequence	Operated by subsea supervisor after the initial explosion but did not function	Continued gas and oil escape on rig
Blowout preventer automatic mode function	Fail-safe function failed because of a solenoid fault and battery low charge	Continued gas and oil escape feeding the fire and resulting in release of oil into the ocean.
Blowout preventer auto-shear operation initiated by ROV	May have partly closed the blind shear ram but did not seal the well	Most severe US environmental incident ever with up to 4 million barrels lost. Widespread pollution of water and beaches. Closure of 86000 square miles of fisheries. Wildlife severely affected. Total losses up to $30 billion.

1.9 So what?

By now it is easy to believe that maintenance of protective devices only matters in complex environments where multiple factors can lead to the death of tens or hundreds of people. So far this section has described and analysed incidents that have gained global media coverage. If you look through national safety authority reports, the picture is different: incidents and near misses are happening every day that involve smaller numbers of individuals. The causes are very similar: neglected maintenance, misuse, poor understanding of protective systems and inappropriate design. The final examples in this section are just some of hundreds.

Crane Limit Switch, Rotherham, England

On 2 July 2003 at a Corus plant in Rotherham, England, a crane was used to lift a 260kg block. The crane's limit switch failed, allowing the hoist rope to be over-tightened. The rope snapped. The block fell from a height of 7 metres and killed a worker who was below it.

Dormitory Fire Alarm, Longwood College, Virginia

At about 06:50 in the morning of 28 April, 1987, a student woke to find an electrical fire under way in his dormitory room (US Fire Administration, 1987). The fire quickly spread to died textiles used as decoration in the room.

Smoke and fire began to spread through the dormitory and the hall fire alarm was pulled by a student. It failed to operate. At about 07:00 a boiler plant employee saw smoke and flames coming from a third floor window and called the Campus Police dispatcher. A resident assistant activated the fire alarm manually, but many students ignored it thinking that it was "just another drill". Finally an announcement over the public address system persuaded the remaining students to evacuate the building.

Fifteen students were treated for injuries: 12 for smoke inhalation, one for second degree burns, one for a broken ankle and one for severe respiratory problems caused by an existing illness.

The investigation found that the original cause was probably a light duty six-outlet extension cord. The fire alarm did not operate because its main breaker switch located in the basement was in the "off" position. A follow-up inspection found that 85% of smoke detectors in student rooms were either disconnected or failed to operate; the detector in the room where the fire started did not work.

Interlock switches, Bury, England

44-year-old Paul Palmer had a 20 year career as a paratrooper serving in Iraq and Bosnia before joining a specialist chemical company in Radcliffe near Bury in northern England. The company makes sealants, adhesives, surface treatments and other chemicals for the building industry.

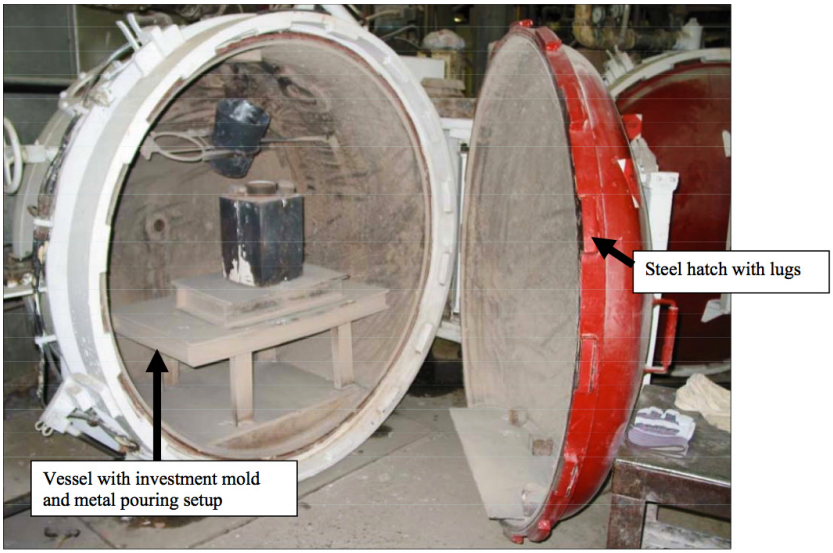
Low speed industrial mixer
Photograph courtesy of the UK Health and Safety Executive

In August 2005, Mr Palmer climbed into a low-speed industrial mixer in order to clean it. Shortly afterwards a colleague started the machine, unaware that anyone was inside. Although the machine ran for only a few seconds, Paul Palmer was killed by the mixer blade.

The subsequent inquiry found that the guards provided were inadequate, and that two switches that should have prevented the machine from operating when its lid was open had failed because of "faults from lack of maintenance." (UKHSE, 2010)

Pressure Relief Valves, New Jersey, USA

Three pressure vessels were used in a small foundry in New Jersey to pressurise and depressurise aluminium to eliminate porosity. The interior of the vessel was accessed through a large hinged hatch at the front secured by metal lugs and sealed by a large O-ring. Two of the three vessels were in use, but there were problems with the third vessel's hatch seal.



One of the surviving pressure vessels, similar to the unit destroyed

A new O-ring was installed and two workers tested it for leaks. One worker operated the pressure controls on the side of the vessel while the second worker listened for leaks at the front. A leak was found when the pressure was set at 80 psi (5.5 bar). It would have been possible to depressurise the vessel at this point and reseat the ring, but in the past gaskets had sometimes been forced to seat by increasing pressure further. The pressure was increased to 112 psi (7.7 bar); at this point the vessel exploded. The hatch was blown off and landed 35 feet (11 metres) away, instantly killing the worker who had been standing in front of it. Nine workers were injured in the accident.

The final pressure that was used in an attempt to seat the O-ring was above the rated vessel pressure, and of course it should have raised the relief valves. After the incident it was discovered that the relief valves were not working because they were clogged with aluminium from the production process. (NJ FACE, 2009)

1.10 Summary

Event	Causes
Buncefield	Level switch left in non-functional state after routine test
Three Mile Island	Auxiliary feed pump valves left closed after maintenance Primary coolant loop relief valve stuck open Operators confused by alarms and instrumentation
Bhopal	Refrigeration system switched off Temperature alarm disabled Scrubber system left in manual mode Flare stack partially dismantled for maintenance
Piper Alpha	Missing relief valve on standby pipework Poor maintenance of deluge system and associated pipework Opportunistic maintenance of relief valve Grossly inadequate permit to work system Poor design of fire deluge system Deluge system left in manual mode after earlier diving work Loss of communication to satellite platforms Inadequate preparation for emergency evacuation
Chernobyl	Test of emergency shutdown power system with reactor in low power state
Deepwater Horizon	Gas alarm may have been disabled Blowout preventer (BOP) annualar preventer failed to seal BOP emergency disconnect sequence failed BOP automatic mode failed BOP ROV mode failed

What are protective systems for?

Protective systems generally fulfil one or more of five roles.

Role	Examples
1 *Provide a warning of unwanted conditions*	Any alarm: high or low temperature, pressure, level, flow, current, voltage, speed, vibration Fire alarms, burglar alarms Gas alarms Aircraft stall warning and ground warning systems Airport explosive detectors
2 *Shut down equipment*	Trips: high or low temperature, pressure, level, flow, current, voltage, speed, vibration Limit switches Emergency stop buttons Electrical residual current detectors; fuses
3 *Reduce the risk of a hazard*	Guards, warning signs Electrical equipment earth bonding Computer network firewalls Firearm safety catch Safety interlock switches
4 *Reduce the effects of failure*	Fire fighting equipment Fire escapes Vehicle traction control, anti-lock braking systems Lifeboats Emergency breathing equipment Pressure and vacuum relief valves; rupture discs Bunds Defibrillator
5 *Provide a standby capability*	Any standby equipment: pumps, generators, lighting Uninterruptible power supply

How can they fail?

1 The protective device has failed since installation or since it was tested

By definition, failure of a hidden function on its own has no effects. So failure of a protective system does not become evident until it is tested or until another failure happens. This is the central reason why protective devices are subject to regular testing: if a device has been tested, we assume that the chance of it failing is reduced compared with a device that is never tested. So pressure relief valves, level switches and electrical system interlocks are checked frequently to ensure that they operate correctly. Calculating how frequently they should be tested is something that will be dealt with in detail by the later parts of this book.

But maintenance of hidden functions is not just about when to test them: it is also about *how* to test.

Maintenance tends to focus on ensuring that protective systems will operate when a real hazard occurs, but it is also important to remember that a protective device may fail to operated *during* a test. So if a level switch is tested by pumping liquid into a storage tank, safeguards need to be in place to prevent overfilling if the level switch does not operate. Similarly, if pressure relief valves are tested by pressurising a vessel, checks must be in place to ensure that the vessel is not overpressurised if the relief valves fail to operate.

Finally, and again the details will have to wait for a later chapter, maintenance needs to ensure that all functions of the protective system operate correctly. For example, the primary function of a pressure relief valve is to relieve excess pressure above a specified level. On Three Mile Island, the primary function of the pilot operated relief valve (PORV) operated perfectly. What contributed to the disaster was its secondary function, to reseat after relieving excess pressure.

2 The protective device never functioned

If a non-functional protective device has been installed, the level of risk is exactly the same as if the device were not there at all. The most obvious way to prevent these failures is to test the device immediately after it has been installed, and the test should be part of the commissioning process.

There is a particular problem with devices that cannot be tested without destroying them, such as fuses and bursting discs.

3 The device has been deliberately disabled

Devices are sometimes disabled in order to test or maintain them: a relief valve isolation valve may be closed for testing; a level switch could be left in its “test” mode and unable to detect high liquid levels.

Devices may also be deliberately disabled because they generate too many trips during normal operation. Worse, as at Bhopal, a device may be disabled because the process that it protects is being deliberately run outside its normal operating envelope.

4 The protective device is not present

Absence of the protective device matters in two ways. Most obviously, if the process can operate without the protective device in place, then the level of risk is increased. If the process is knowingly operated without protection in place, other arrangements (such as manual monitoring) should be made to provide an equivalent level of protection.

Second, the process may not be capable of operating safety without the device. In general this failure would be evident: absence of a level switch would shut down the associated process, or a missing relief valve would cause immediate loss of containment. However, this is exactly the failure of protection that was at the root of the Piper Alpha incident. This failure was *doubly hidden*: it only became evident when the duty condensate pump failed, causing the operators to start the standby leg. When the standby pump was started (hidden function 1), the missing relief valve became evident (function 2).

5 The device operates when it is not required

Unwanted or unintended operation of a protective device is usually evident: the process shuts down, gas escapes, or an unexpected alarm sounds when equipment is running correctly. The consequences of an unexpected alarm may be trivial (nuisance and repair costs) or economic (lost production due to shutting down a process).

The table below summarises the role that protective devices played in the incidents that have discussed in this chapter.

Incident	Protective Device Failed	Protective Device Poor Design	Failed During Test	Protective Device Disabled	Protective Device Missing
Buncefield				☝	
Three Mile Island	☝			☝	
Bhopal	☝	☝		☝	
Piper Alpha	☝	☝		☝	☝
Chernobyl		☝	☝		
Deepwater Horizon	☝	?			

1.11 How can we ensure the availability of protective systems?

One thing becomes obvious from the table above: that design and maintenance play a core role in the availability of protective devices.

In summary, the options for maintenance include the following.

1. Use preventive maintenance (also called *proactive maintenance*) to prevent the protective device from failing. Maintenance may include monitoring the device to anticipate a failure (condition monitoring), scheduled overhaul or scheduled replacement of the device.
2. If preventive maintenance is not applicable, test the device at regular intervals to check if it is working. Repair or replace the device if it is not functional. Scheduled testing or failure-finding is applicable to a wide range of devices where failure cannot be anticipated or prevented.
3. If failure cannot be prevented or detected, determine whether the system design is robust enough to reduce the risk of failure to a tolerable level. If it is not, consider redesigning the protective system or other equipment to reduce the risk of failure.

But while maintenance of protective devices is important, it is also clear that maintenance or any invasive action can be responsible for *disabling* protective systems. Because more maintenance does not necessarily mean improved safety or availability, the maintenance of protective devices is a subtle art that will occupy most of the remainder of this book.

1.12 Key Points and Review

Protective system failure has been implicated in a wide range of incidents.

Protective systems can fail because of lack of maintenance, but poor design and deliberate or accidental disabling of devices have been implicated.

2 Hidden Functions

2.1 Introduction

Having established the critical importance of hidden functions and failures in real-world incidents in the previous chapter, this section goes on to define the general concepts used in the analysis of hidden functions and failures.

It also looks at some of the subtleties of hidden and evident functions and tries to answer an apparently simple question: when is a function evident, and when is it hidden?

2.2 When is a Function or Failure Hidden?

Hidden functions are conditional

A hidden function is *conditional*: it only comes into play on condition that a second event occurs. For this reason, typical hidden function statements can be recognised by words similar in meaning to those in the list below.

if	To shut down the turbine **if** its rotational speed exceeds 15000 rpm
capable of	To be **capable of** sounding an audible alarm if the storage tank liquid level rises above 2.5m from the tank base
in the event that	To bring the train to a safe stop **in the event that** the driver fails to respond to the audible and visual alarms

A protective device carries out its hidden function if a second event occurs; this is the *trigger event* or *initiating event*. The most obvious trigger is the failure of other components or equipment, but it could be the result of anything that does not occur during normal operation, including the following.

- Human error
- Loss of an external service such as electrical power, gas, cooling or heating services
- Failure of a control system
- External factors such as vehicle impact, severe weather, earthquakes and so on

The table below lists a number of typical protective systems, their associated functions and the trigger events that cause the protective system to operate. The final column is the overall function statement for the protective device; the trigger event is shown in *italics*.

Protective System	Carries out this function	...if this trigger event occurs	Function statement
Emergency stop switch	To stop the can filling line	*Any one of 10 emergency stop buttons is pressed*	To stop the can filling line if *any one of 10 emergency stop buttons is pressed*
Carbon monoxide gas alarm	To raise an audible and visible alarm	*The carbon monoxide concentration exceeds 400 ppm for 10 minutes*	To raise an audible and visible alarm if *the carbon monoxide concentration exceeds 400 ppm for 10 minutes*
Boiler pressure relief valve	To relieve excess boiler pressure	*Boiler pressure exceeds 10 bar*	To be capable of relieving excess boiler pressure if *it exceeds 10 bar*
Residual current device (RCD) or ground fault circuit interrupter (GFCI)	To interrupt the power supply within 40 milliseconds	*The imbalance between live and neutral line currents exceeds 10mA*	To interrupt the power supply within 40 ms if *the imbalance between live and neutral line currents exceeds 10mA*

Failure of the hidden function by itself has no consequences

First we need to be clear what "consequences" are. In this context, consequences include anything that could be observed by the equipment operators, not just the failure's direct effects on production output or safety.

Because the trigger event is not expected to occur during normal operation, the hidden function can never be activated unless something unusual happens. As a result the hidden function is never triggered in normal circumstances, and the hidden failure *by itself* has absolutely no consequences at all.

If a protective device is in a failed state when an initiating event occurs, then of course the outcome is very different. The resulting consequence is a *multiple failure*, the event that the protective device was intended to prevent.

No one will notice the effects of a hidden failure

It follows from the last section that when a hidden function fails, no one who is involved in operating the equipment notices any effects. As we have already said, these are not only effects on production or safety; they include any effects, including "fail safe" features that may have been designed to make the hidden failure evident.

This part of the definition can be confusing if you think about it hard enough. How can a device or system that has failed have absolutely no effects at all? How would we ever be able to diagnose a problem? To take a real example, could the failure of a pressure relief valve really be considered hidden if I could just walk past and see solidified product around it that would prevent it from operating correctly?

This is where the definition needs to be more precise. Of course, hidden failures do have some consequences: at very least, some part of the protective device has failed, and perhaps we could work out that the failure had occurred by inspection, by shaking the device or by dismantling it. But we are not talking about whether the failure can be found through maintenance intervention: the question is whether the failure would be noticed during normal operation, without equipment maintenance and without an engineer specifically looking for the problem. If there would be no effects under normal conditions, the failure is hidden.

The importance of time

There is one last factor to take into account: time.

Failure effects do not have to appear immediately for a failure to be classified as evident.

For example, if the filter in a cooling water supply is blocked, its effects may not become evident until there is a demand on the cooling system. It could take some time for the process that uses cooling water to overheat; in fact, it could be hours or even days before the problem comes to light. Is the filter blockage hidden? No, because its effects become evident *eventually*, even if the immediate effects are negligible or non-existent.

This rule may seem contrived, but it is not difficult to remember: a failure is evident if the operating staff *eventually* become aware of its effects when *everything else is operating normally*. So the filter blockage is evident, because eventually it causes the downstream process to overheat. On the other hand, failure of a fire alarm to detect fires is hidden because fires are not part of normal operating conditions.

2.3 Hidden Failures: a Definition for RCM Users

The previous sections have laid down these principles for defining a hidden failure.

1 A hidden failure by itself has no effects

2 The effects of a hidden failure only become evident if a trigger event occurs which would normally cause the hidden function to operate

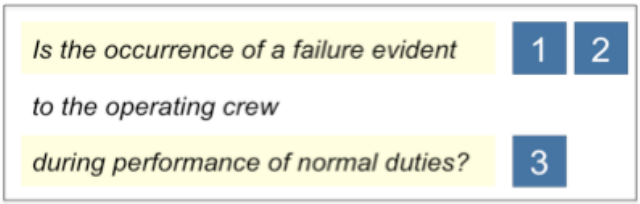
3 The only failure effects that count are those observed by the operations staff carrying out their normal duties

4 Even if it is not possible to diagnose exactly which failure has occurred from its effects, the failure is still evident. To be hidden, a failure must have no effects at all when it occurs on its own.

5 A failure whose effects appear eventually under normal circumstances is evident, not hidden

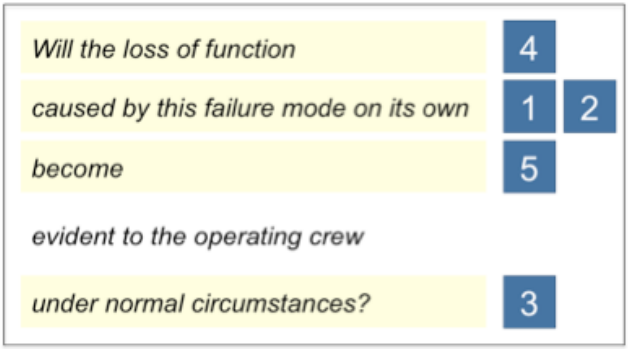
Every Reliability-centred Maintenance decision diagram includes a flowchart that identifies hidden failure modes. Finding the question is easy: it is usually the longest and most complex because it tries to embody all five of the principles above in a single sentence.

First, here is the original question from the Nowlan and Heap (1978) decision diagram.



This embodies principles 1, 2 and 3, but it does not capture them all, and it does not capture principles 1 and 2 as well as it could.

The most complete and carefully considered definition of an evident failure in RCM is probably that in the RCM 2 Decision Diagram (Moubray, 1997).



The RCM 2 definition introduces the idea of time ("*become* evident…") and focuses on the loss of function rather than the failure itself; it is the effects of failure that have to be evident, not the ability to diagnose the failure. Even so, it is probably impossible to compress all the subtleties of hidden failures into one sentence. Whichever definition you use, most failures are easy to categorise as either hidden or evident; you will only need to use the whole checklist for a tiny proportion of failure modes.

2.4 Failure Modes

So far this chapter has considered only one aspect of protective device failure: the loss of the primary protective function. However, even the simplest device can usually fail in at least two ways. One failure mode causes loss of the protective function (and is therefore hidden), while the second failure mode incorrectly triggers the protective action in normal circumstances, usually resulting in evident consequences.

The list below shows some simple examples.

Protective System	Failure Mode	What happens if the failure occurs
Emergency stop switch	Is incapable of stopping the canning line when an emergency stop switch is pressed	No effects in normal circumstances. Someone could be seriously injured if the emergency stop were needed to protect someone from running equipment
	Shuts down the canning line when no one has pressed an emergency stop button	Interrupts production and may result in significant product loss.
Carbon monoxide (CO) gas alarm	Cannot raise an audible alarm when CO concentration exceeds 400 ppm for 10 minutes	No effects if the CO concentration is normal. People could be injured or killed by high undetected CO levels if a burner or flue malfunctioned.
	Raises an audible alarm when CO concentration is normal	A spurious alarm could cause evacuation of personnel and a shut down of equipment until it has been inspected.
Boiler pressure relief valve	Is incapable of relieving boiler pressure above 10 bar	No effects unless the boiler pressure rises to abnormal levels, when it could explode
	Relieves at normal boiler pressure, allowing steam to escape	Allows steam to escape at normal boiler pressure, affecting production
Residual current device (RCD/ GFCI)	Is incapable of interrupting the power supply within 40 milliseconds if live and neutral currents are out of balance	No effects under normal conditions. If an unintended short to earth occurs, personnel could be seriously injured or equipment damaged.
	Interrupts the power supply when the live and neutral currents are balanced	Cuts power and shuts down production equipment

These are only simple examples; in practice, protective devices can fail in many ways. Some of those failure modes will be hidden, and some will be evident. Properly designed protective systems take into account the level of protection required and the impact that spurious alarms and trips may have on normal operation. The analysis of maintenance requirements—including periodic testing—also needs to take into account both hidden and evident failure modes.

The following chapters deal with these different failure modes in more detail, including methods for evaluating the availability of protective systems and the expected rate of spurious operation. They also cover techniques for combining failure modes by "black boxing" to reduce the analysis overhead.

2.5 Making a Hidden Function Evident

Hidden failures are potentially dangerous because there is no indication that the failure has happened unless the protective device is checked or a multiple failure occurs. So designers sometimes add features that monitor the protective device and take action (usually raising an alarm) if the protective function is disabled for some reason.

For example, failure of a car's traction control or anti-lock braking system could be hidden, because under normal circumstances the system does not need to operate to prevent skidding. However, manufacturers have recognised for some time that drivers need to be aware when the system is not working, and so modern units incorporate sophisticated monitoring of the control unit and its sensors to make the driver aware of most failures.

Two further examples are shown in the table below, with a description of failure effects for the unmodified and modified protective devices.

Protective System	What happens when it fails?	Hidden?
Smoke detector connected to a simple alarm system	Nothing happens under normal circumstances. If a fire occurred in the area covered by the sensor, no alarm would sound	Yes
Smoke detector connected to a more complex alarm system	Under normal circumstances, the alarm polls the sensors every 60 seconds to ensure that they are capable of sending an alarm signal. Most sensor failures would cause a "fault" light to illuminate on the alarm panel and a signal would be sent to the remote monitoring station. If a fire occurred in the area covered by the failed sensor, no alarm would sound	No
12000 rpm overspeed alarm warning lamp	Nothing would happen under normal circumstances. If the turbine entered an overspeed condition, no alarm would be displayed and an uncontrolled shutdown would be initiated at 15000 rpm. If the alarm had worked, the operator could have taken measures to reduce turbine speed or to initiate a "soft" equipment shutdown.	Yes
12000 rpm overspeed alarm warning lamp with intelligent monitoring system	Under normal circumstances, the control system detects an open circuit lamp and displays a warning on the operators' main control screen. The operator schedules lamp replacement. If the warning lamp were non-operational and the turbine entered an overspeed condition, no alarm would be displayed and an uncontrolled shutdown would be initiated at 15000 rpm. If the alarm had worked, the operator could have taken measures to reduce turbine speed or to initiate a "soft" equipment shutdown.	No

While the designer of the protective device has made the hidden function evident, it is important to remember that the new layer of protection has introduced an additional hidden function. So in the examples above, the smoke detector monitor would need to be checked to ensure that it can identify a failed detector, and similarly we need to ensure that the function of the lamp monitor is properly maintained. Both of these are hidden functions.

2.6 Into the Grey: Hidden or not?

Before starting this section, let me say first that it is easy to classify almost all failures as hidden or evident. A very small proportion—well under one per cent—cause any difficulty, and only a very few of those are genuinely ambiguous.

A very small decrease in performance or increase in operating costs

Most ambiguities arise because the effects of a failure are small and, under normal circumstances, almost unobservable.

For example, a very slow leak of water from a pipe joint would obviously result in higher utility bills. If the leak were into a drain, and the loss was automatically made up by the feed water system, would the leak be hidden or evident?

Would the leak become evident eventually? If the leak is likely to grow and become evident, perhaps because of pools of water or increasing water usage, then the failure is evident; otherwise it is genuinely hidden.

Frequent activation of a protective device

A hoist includes a protection system to stop the motion of the load if it is lifted too high. Investigation shows that the protective device is tripped on average about once per shift, or three times per day.

Given the high rate of usage, tripping the over-hoist protection appears to be part of "normal operations", so the failure appears to be evident. In any case, testing the device more than once a shift would be impractical, so failure-finding does not really seem appropriate. However it is very unlikely that the designers intended the switch to be operated so frequently; they almost certainly intended it to be a rarely used protective function. Rather than meekly accepting the current state of affairs, this example suggests that design and operation of the hoist should be reviewed. Classifying the failure as hidden or evident is probably irrelevant.

Extended period between the failure and its consequences

A sunken oil storage tank develops a leak. Over time, oil percolates through the soil, but the rate of loss is not enough to alert operations staff. Remember that the analysis is zero-based, so we assume for the moment that no maintenance is being carried out; no one is going around looking for leaks. After a period of years, the oil reaches a river that is used as a local source of fresh water, and its presence is detected by analysis of samples. Is the leak hidden or evident?

A theoretical approach says that the leak is evident, because it becomes evident eventually. If you want to stir up an argument, you could say that the period between a leak starting and anyone noticing the consequences is so long that the plant could have closed down by then. So isn't the failure hidden after all?

Dealing with ambiguity

For the very few failures that are genuinely difficult to classify, it is helpful to take a step back and ask the question: "*What difference will it make if the failure is classified as hidden or evident?*"

The objective of RCM is to manage failures appropriately, and classifying them as hidden or evident is just part of that process. The ultimate goal is to put in place maintenance tasks that are effective or to identify where redesign is necessary.

The table below takes the three examples above and lists the likely maintenance task selection assuming the failure is treated either has hidden or evident.

Failure	Possible maintenance task selection if hidden	Possible maintenance task selection if evident
Slow water leak from pipe joint into drain	Visually check joint for leaks once per day	Visually check joint for leaks once per day
Overhoist protection switch fails	Change operating procedures or redesign system	Change operating procedures or redesign system
Slow oil leak from underground tank	Take soil samples from area around tank at an appropriate interval	Take soil samples from area around tank at an appropriate interval

In this case it makes no difference: the responses are the same whether the failures are classified as hidden or evident.

2.7 Key Points and Review

A hidden failure has no observable effects unless another event occurs, usually a second failure.

The only failure effects that count are those observed by the operations staff carrying out their normal duties

A failure is evident even if it is not possible to diagnose exactly which failure has occurred from its effects.

The effects of an evident failure appear eventually under normal circumstances

Typical protective devices can fail in at least two ways. Failure to provide the protective function is generally hidden, but unintended operation of the protective device is usually evident.

In a real world analysis, most failures can easily be classified as hidden or evident. Ambiguous failures are rare.

If you find a failure that is difficult to classify, focus on the maintenance outcome: does it make any difference if the failure is classified as hidden or evident?

3 Managing Hidden Failures

3.1 Introduction

Hidden failures need to be managed because of the severity of the consequences of a multiple failure. Managing hidden failures poses two specific challenges. First, and by definition, there are no observable effects when a hidden failure occurs by itself. This is precisely what makes the failure hidden rather than evident. Secondly, many of the devices used in protective systems rely on electronics and other technologies that predominantly fail at random, with no predictable pattern or age of failure. These two factors together appear to make the management of hidden failures an impossibility.

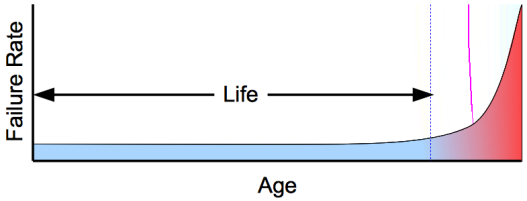
A management policy that focuses on the *effects* of a hidden failure is doomed, simply because there are no effects to manage. The key to management of hidden failures is to focus on the characteristics of the failure itself rather than on its effects. What are the characteristics that could provide the basis of a successful failure management policy? This chapter examines the factors that affect the selection of maintenance tasks in general, and with an emphasis on hidden failures in particular.

3.2 One Part, Several Failure Modes

One part can have several failure modes. Some of them may be hidden, others evident. A failure maintenance policy is needed for each failure mode, not just one policy for the whole part.

3.3 Scheduled Overhaul and Discard

Many failure modes have a characteristic life. Their life is not a point at which all failures will occur, but it is a time when the probability of failure starts to increase rapidly. If the part remains in service, it becomes more and more likely to fail.



An age-related failure pattern

In the pattern illustrated above, there is a small, roughly constant chance that the failure occurs at any time after the part is installed. Later on, the chance of failure begins to increase rapidly; this point is marked as the item's life.

The obvious response to items that have an identifiable life is to replace them before that life is reached. This type of task is known as *scheduled discard*, *scheduled replacement* or *lifed* task.

An item's life is almost never known exactly unless formal reliability trials have been carried out. In general an estimate is made of the likely minimum life, and the replacement task is scheduled before that life is reached. Life is not always measured in terms of calendar time: it may be expressed in run hours or some other measurement of the part's usage. The common characteristic of all scheduled discard tasks is that they are carried out at fixed intervals.

This is obvious when considering evident failures, since these are the drivers for cyclic replacement of components.

Examples of "lifed" items include the following.

Component	Failure Mode(s) driving replacement cycle
Vehicle tyre	Tread wear Material degradation
Pipe	Erosion by impact of particles in fluid Corrosion by fluid External corrosion
Pump impeller	Erosion
Aircraft wings	Fatigue

Some components are subject to several "lifed" failure modes. This may lead to an "either/or" maintenance policy. For example, in the case of tyres, the material of which they are made wears off (lifed failure mode 1) and also degrades over time (lifed failure mode 2). This leads to a task that could be written as follows.

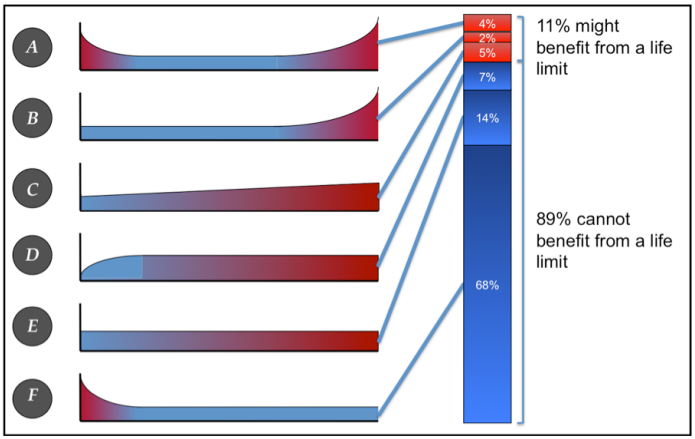
"Replace tyres every 50000 km travelled or every 5 years"

If a pipe would fail after five years as the result of internal erosion, but after 15 years because of external corrosion, it might be replaced every five years. However, as we shall see below, it is often possible to devise a maintenance policy for some items which is both safer and extends a component's useful life compared with fixed interval replacement.

An important point to take from this section is that the "life" which drives any task is that of the *failure mode*, not that of the item or part. As with the tyre, a single part may be subject to several failure modes, each of which has a different characteristic life. A separate task is needed to manage each failure mode, and the tasks are then combined when the maintenance schedule is constructed.

3.4 Condition-based Maintenance

Research in civil aviation during the 1960s and 1970s revealed a fundamental problem with maintenance management that relies on scheduled replacement: most failure modes do not have a predictable life. Of the items studied by United Airlines, fewer than 11% had failure patterns for which lifed tasks would have been a plausible management strategy (Nowlan and Heap, 1978).

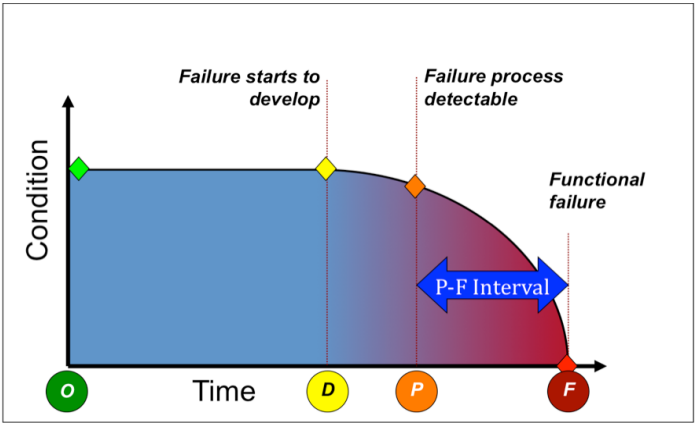


Equipment failure patterns from Nowlan and Heap (1978)

How, then, could the remaining 89% of failures be managed?

Although relatively few failure modes have a definite characteristic life, many failures give some warning that they are about to happen. The length of the warning period may vary widely, from seconds to months or years, but often it is long enough to prevent the failure from occurring, or at least to reduce or eliminate the consequences of failure.

The diagram below illustrates the failure development process from a point where the equipment is running acceptably through to the point of failure.

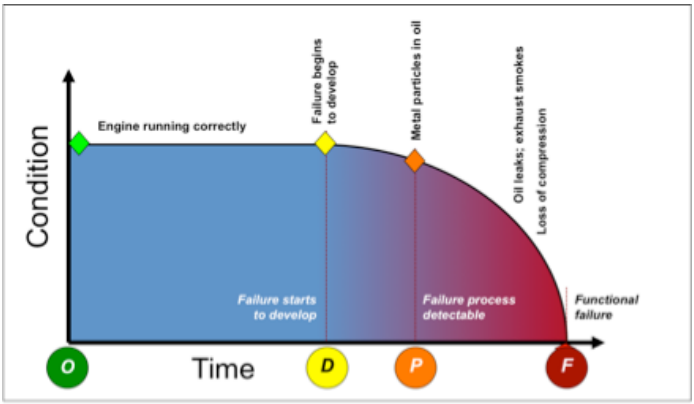


Failure development process through to functional failure

Initially the equipment is operating acceptably (point O above). At some point *D*, not necessarily related in any way to the equipment's age, the failure begins to develop; it may not be possible to detect any symptoms immediately, but if nothing is done, the item will deteriorate all the way to functional failure (point *F*). Between points *D* and *F*, it becomes possible to detect the deterioration, perhaps by sight, by sound, or by using some form of sensor. This is the point of potential failure (P), sometimes called incipient failure.

The example below demonstrates how this might apply to failure of an engine's piston rings.

Stage	Description
O	Engine running normally
D	Piston ring begins to wear. Metal particles are present in the oil but are not detectable by normal oil sampling techniques.
P	Detectable debris present in the oil. Oil leaks as wear increases. Loss of compression.
F	Engine is unable to sustain the required load. Serious oil leakage. Severe engine damage is possible.



Failure development and potential failure conditions

A failing piston ring shows more than one warning sign before the engine stops: first, the presence of small metal particles in the engine oil, then oil leaks and lack of compression. Any of these symptoms might be used as a potential failure condition. The key for managing the failure is the interval between a detectable symptom of failure (particles, oil loss and so on) and final functional failure (the engine stops). This interval is known as the P-F interval.

Which symptom is chosen as the potential failure condition depends on the expected interval between the P-F interval and how much notice is needed in order to schedule appropriate maintenance or to mitigate the consequences of failure. Depending on how the engine is used, oil sampling might have a P-F interval of weeks or even months; compression testing or looking for a smoky exhaust might give days' or weeks' notice of failure.

When a potential failure condition has been identified, the condition monitoring task that manages the failure can be written. Because P-F intervals are often not known with any certainty, it is common to schedule the condition monitoring task to take place at half the P-F interval, although there is no absolutely definite rule. If we assume that the minimum P-F interval for oil sampling is eight weeks, then the task chosen to manage the failure might be written like this.

> "Take an oil sample every four weeks [i.e. half the P-F interval] and send sample for analysis. If analysis indicates piston ring wear, schedule replacement of all rings or substitution of engine."

3.5 Failure-Finding

Failures that occur after a definite life can be managed by scheduled component replacement; failures that provide some form of warning or potential failure can be managed by condition monitoring. What can be done if the failure has no life and no identifiable potential failure?

If we were trying to manage an evident failure mode, the answer would be simple: maintenance could do nothing. If the consequences of doing no maintenance were unacceptable, our only remaining option would be to redesign the equipment or to change the way in which we use it.

The situation is different if we are dealing with a hidden failure, simply because in normal circumstances we do not know that a failure has occurred. If we do nothing at all, the protective system will fail, then remain in a failed state until whatever event it is supposed to protect us against happens. And then, of course, it will do nothing because it has already failed.

The difference between hidden and evident failures gives us an alternative to doing nothing. Although we may not be able to predict that a failure is going to occur, we can detect it once it has happened. This is not going to prevent the failure, but we may be able to find out that the protective system has failed before its failure has any consequences.

For example, a fire alarm consists of many components, the failure of any one of which could prevent it from working if a fire started. Very few of these failures have a definite life or give any warning of failure before they happen. However, we can test the alarm, perhaps by pressing a "test" button, or by simulating a fire, and repair the alarm if it does not operate. If the test shows that the alarm is not working, then it can be repaired so that it is capable of detecting a real fire.

This testing or checking task, usually carried out at regular intervals, is called a *failure-finding* task. There is a very important distinction between failure-finding tasks and the scheduled discard, scheduled refurbishment and condition monitoring tasks discussed above. If a scheduled discard, scheduled refurbishment or condition monitoring task works as intended, the failure that it is managing never happens. Failure-finding tasks are different because they potentially allow the protective system to fail. More than that, if the protective system has failed, it remains in a failed state until the next failure-finding task is carried out.

Why does this matter? Because if the protective system fails, there is no protection *at all* until the next failure-finding task is carried out and reveals the failure. So if the fire alarm fails, it is incapable of protecting us from the consequences of a fire until it is next tested.

So failure-finding tasks are fundamentally different from the range of scheduled tasks that can also be applied to evident failures because of the possible gap between failure of the protective system and the next failure-finding task. The gap means that however often we check the protective system, there is always a small probability that it could be in a failed state when the event against which it is protecting us occurs.

3.6 Do nothing

To do nothing might appear to be an unlikely failure management strategy, but it is a logical choice under certain conditions.

- The failure has no safety or environmental consequences, and
- There is no applicable preventive task, or
- The cost of carrying out any applicable preventive task is more than the cost of allowing the failure to happen

Doing nothing, or more formally *no scheduled maintenance*, may therefore be a positive decision based on the reliability and cost data that are available.

Doing nothing is an unusual decision for hidden failures, because if the design of equipment requires a protective device, it is likely that maintenance is required to ensure that it is available when needed. Section 3.8 below describes the circumstances under which failure-finding may not be applicable to hidden failures.

3.7 Redesign Options

While redesign is not strictly part of maintenance management, it is an important aspect of failure management. If no maintenance task can prevent, predict or detect the failure, and doing nothing is not an acceptable option, the final choice is to redesign the equipment. "Redesign" does not necessarily mean a high cost, physical redesign; in practice, redesign may mean changing the way in which equipment is used or changing operating instructions.

Redesign as a management strategy for hidden failures is discussed fully in a later chapter. Examples might include the following.

Redesign Strategy	Examples
Make the hidden failure evident	An uninterruptible power supply checks itself three times per day and raises an alarm if its battery has failed Add a circuit to check the continuity of normally-off incandescent warning lights
Add more protection	Add a second relief valve to a vessel that has a single valve Add a tank ultimate high-level shutdown switch in addition to its existing alarm system
Improve the reliability of the protective system	Upgrade a tank's ultimate level switch to a new model that prevents the entry of dust, dirt and product
Reduce the rate of initiating events	Improve a crane's control system and retrain operators to reduce the rate of demand on the overhoist protection switch

3.8 When is Failure-Finding not Feasible?

If you believe that every hidden failure can be successfully managed by carrying out some form of scheduled test, then think again. Failure-finding is only a practical management policy if the following conditions are met.

- It is possible to check whether the protective system has failed, and
- It is practical to carry out the failure-finding task at the required interval

Both points probably seem obvious, but you need to be aware of some important issues that have to be considered.

When is it impossible to check whether a protective device has failed? Devices that cannot be tested without destroying them belong in this category. They include fuses, rupture discs, shear pins and automobile air bags among others. None of these devices can be fully tested without operating and destroying the device, and as a result failure-finding is impossible.

It may be infeasible to test a protective system if there is a significant chance of causing the multiple failure while carrying out the test. For example, testing a turbine overspeed trip by deliberately defeating the normal control system, or a tank high level trip by overfilling the tank may result in an unacceptable risk of the multiple failure occurring during the test. In this case it may be possible to bring the risk to within a tolerable level by careful wording of the task or by employing additional protection during the test. "Additional protection" does not need to be additional equipment: it may be that a second technician could monitor conditions and be ready to shut down the system if required. Although it can be good practice for a failure-finding task to replicate the abnormal conditions as closely as possible, it is essential to ask the following question before selecting the proposed failure-finding task.

"If the protective system fails to operate correctly during this test, is there a risk that the test will result in the multiple failure occurring?"

Even if it is possible to test a protective system, the required failure-finding interval could be impractical for two reasons: it may be too long, or it may be too short.

Long failure-finding intervals are common if the protective system is very reliable, demands on it are infrequent, and the consequences of failure are insignificant.

> A review group analysing a section of a chemical plant needs to set the failure-finding interval for a motor overload trip. The motor drives a water pump. The best estimates available to the group show that the mean time between failures of this type of trip is at least 200 years and demands are likely to occur no more than once every 20 years on average. The cost of checking the trip would be $30. If the trip failed to operate when required, the motor would burn out, but its replacement cost is no more than $250.
>
> The group determines that the trip's optimum failure-finding interval is 31 years.

When a properly calculated failure-finding interval is longer than the probable life of the equipment that it protects, the message is simple: be sure that it works today, then leave it.

Short failure-finding intervals are more challenging, and whether a specific interval is practical depends on the details of the system under analysis.

Incorporating failure-finding tasks in equipment start up or shut down procedures often provides the best opportunity for high frequency checks.

> Engine start up checks
>
> …
>
> After applying power, but before starting the engine, check that the following lamps are illuminated on the control panel: battery charging alarm; low pressure oil warning alarm;…

If a failure-finding task is required too frequently to be practical, it can mean that the design of the system is no longer able to deliver an acceptable level of risk. The redesign options discussed in the previous section should be considered.

3.9 Important Note

By now it should be obvious that the subject matter of this book is a very small part of a complete maintenance strategy. We have limited ourselves to hidden failures and mostly ignored those that are evident. The remainder of the book further assumes that the hidden failure cannot be managed by preventive or predictive tasks, so failure-finding is the only remaining option.

Remember that failure-finding allows the protective system to spend time in a failed state, unable to provide protection. For that reason it is important to consider options that *prevent* the failure before considering failure-finding. See other texts such as John Moubray's book (Moubray, 1997) for further information on failure management through fixed interval replacement, overhaul and condition monitoring.

3.10 Key Points and Review

Failure-finding is a task that checks whether a protective system is in a failed state. The protective system is allowed to run to failure, but its function is checked at fixed intervals to determine whether it has failed.

Failure-finding is not the only maintenance policy that can be used to manage hidden failures.

Because the protective device is allowed to run to failure, there is always a finite chance that it is in a failed state when a demand occurs on it. If failure-finding is chosen as a maintenance policy, there is a finite chance that a multiple failure will occur.

In general it is possible to manage the chance of a multiple failure by increasing the frequency of a failure-finding task, decreasing the demand rate on the protective device or both.

If condition monitoring, fixed-interval replacement or overhaul is technically feasible, it may reduce the risk of a multiple failure below the level that can be practically achieved through failure-finding.

If no maintenance policy can achieve a tolerable level of risk, the system may need to be redesigned to improve the availability of the protective system, reduce the demand rate on it, or to make the hidden function evident.

4 Failure-Finding Basics

4.1 Introduction

This chapter builds the foundations that you will need to apply failure-finding and other failure management policies to real equipment.

The techniques presented in this book are like tools in a toolbox. Before using them, you need to be able to understand the terminology and to identify the protective device, the demand, and the ultimate multiple failure. Even if you already have some background in risk analysis, risk-based inspection or Reliability-centred Maintenance, you should spend some time becoming familiar with the terminology used in the following chapters.

4.2 Protective Devices and Systems

The terms *protective device* and *protective system* are used interchangeably in this book.

A protective device is intended to operate if an initiating event or trigger event occurs. In general the term "protective device" is used for a small, self-contained component such as a sensor or a relief valve, while "protective system" is applied to a whole item of equipment such as a fire alarm. The terms are often used interchangeably in this book, and there is not usually any significance in the use of "device" rather than "system".

Examples of protective systems are listed below.

Protective System
Fire alarm
Pressure relief valve
Pump motor trip
Car anti-lock braking system (ABS)
Hospital emergency generator

4.3 Demand and Initiating Event

The protective device operates when a *demand* is placed on it by an *initiating event* or *trigger event*. These three terms are used interchangeably.

Examples of typical demands on protective systems are listed below.

Protective System	Demand (initiating event)
Fire alarm	A fire breaks out
Pressure relief valve	Steam boiler overpressure
Pump motor trip	The pump motor stalls
Car anti-lock braking system (ABS)	Need to brake in an emergency or in slippery conditions
Hospital emergency generator	Main electric power supply failure

Protected Function

The term *protected function* is used by Moubray (1997) and in other published work derived from RCM 2. This book avoids using the term for a number of reasons.

- "Demand" and "event" are far more widely accepted, and they clearly describe the relationship between the protective device and the events that should cause it to operate
- The terms "protected function" and "protective device" are so similar that they often cause confusion
- It is the *failure of the protected function* that actually places a demand on the protective device
- It is sometimes unclear what the protected function actually is

If the protective system is a backup system such as a standby water pump, it is obvious that the protected function is something like this: "To pump water at a specified rate", a function that is probably part of the RCM analysis. It is far less clear if the device is a fire alarm, where the function could be "Not to catch fire", which would probably not appear in the analysis. Overall, the term *protected function* has been avoided to improve clarity.

4.4 Multiple Failure

The multiple failure is what happens if the demand occurs while the protective system is in a failed state. The effects of the multiple failure need to be recorded clearly so that your can set up a consistent maintenance schedule for the protective system.

Failure Effects and Consequences

Before beginning to analyse a protective system, ensure that the following components are clearly identified.

- The protective system
- The demand or initiating event
- The multiple failure

Do not be tempted to continue with the analysis until you can clearly define each of the above elements. If you are facilitating an RCM review group, consider writing them down so that no one is in any doubt.

The following table shows some examples of protective systems, the associated demands and a definition of the multiple failure in each case.

Protective System	Demand	Multiple Failure
Fire alarm	Fire	An undetected fire occurs, resulting in increased risk of death, injury and physical damage.
Pressure relief valve	Steam boiler overpressure	Excess steam pressure is not relieved and the boiler explodes resulting in death and injury of personnel.
Pump motor trip	Motor stall	The motor stalls and burns out.
Car anti-lock braking system (ABS)	Need to brake in an emergency	ABS does not operate when brakes are applied in an emergency, and the vehicle skids out of control.
Hospital emergency generator	Main power supply failure	Emergency generator does not start during a power outage.

4.5 Failure Modes

In Chapter 2 we saw that one protective device can fail in a number of ways; in other words, it displays a number of *failure modes*. The primary function of the device may be hidden, but that does not mean that all of its possible failure modes are also hidden.

Protective devices can fail in two distinct ways: fail to operate when required, and to operate when there is no demand (spurious operation). A device can be subject to both hidden and evident failure modes. *Examples*.

4.6 Availability

Beware: "Availability" is a deceptively simple word. A protective device is *available* if it is capable of performing its function if a demand occurs. If it is incapable of correct operation, it is *unavailable*. From this point of view, a protective system is either available or it is not, so its availability is either 100% or 0%. In the real world, availability could hardly be a simpler concept.

Mathematicians, statisticians and reliability engineers learn over a period of many years' training to make simple ideas far more complex. To a reliability engineer, the availability of a protective device could be 0%, or 100%, or any number in between. To see how this picture differs from the simple all-or-nothing, 100% or 0% picture of availability, consider the following question.

"Did the fire alarm operate when we had that electrical fire last week?"

This is a simple question, and the answer is equally simple: either it worked or it didn't. The question could be rephrased in availability terms like this:

"What was the availability of the fire alarm when we had that electrical fire last week?"

The answer is either 100% or 0%, not 80% or 99.5% or 5%. It worked or it didn't.

Now look at a different question.

"If a fire were to occur now, would the fire alarm be capable of detecting it and annunciating an alarm?"

The truthful answer to this question is that we have no idea. In availability terms, the question is:

"What is the availability of the fire alarm now?"

There are two different ways in which we could try answer this question. Since the real world answer to the question is either 100% or 0%, we could start a fire (or preferably simulate one) and see whether the fire alarm operates. If it does, it was available; if it doesn't, it was unavailable. Although that gives us a definite answer, it's of no real use to us. Truthfully we don't want to know whether the alarm works now; we are far more concerned about whether it would operate when no one is around to test it, perhaps in the middle of the night. What we want to know is:

"What is the chance that the fire alarm would work if a fire occurred?"

An analysis of the system and its maintenance (perhaps by a reliability engineer) might be able to tell us the *probability* that the alarm would work correctly if a fire occurred. Although the "all-or-nothing" picture of availability represents what happens when a fire occurs, this probability is of far more use to us. It tells us how likely our protective systems are to operate when they are needed. The probability of operation is a number between 0% and 100% and it is known as the availability.

The probability of the alarm working correctly depends on a number of factors that we will investigate in the following sections.

In order to be effective, a protective system not only needs to exist, it needs to be available when it is required. For example, a simple fire alarm system could be unavailable for a number of reasons when a fire occurs.

- A component has failed in such a way that it is unable to detect a fire and annunciate an alarm
- The system has recently tested in a way that involves disabling part of the system during the test, and the technician forgot to enable the system afterwards
- The system's power supply has failed and no backup power supply is available

Any one of these failures is sufficient to ensure that the fire alarm's function is unavailable when a fire occurs. While it is possible to influence a system's availability through scheduled testing, failure of its components is only one root cause of unavailability. The simple example above demonstrates that unavailability may also arise from human intervention (testing) and external factors (the power supply) and even its design. When analysing a protective system, ensure that you understand and take into account all the factors that might disable it, not just those which maintenance can influence.

For discussion

The fire alarm in this example is a simple system. Most commercial systems incorporate a battery back-up power supply so that they can operate for extended periods without mains power; the alarm may also signal its monitoring centre when power supply problems occur.

What additional maintenance requirements could arise because of the increased complexity of a fire alarm which includes a back-up power supply and signalling, compared with the maintenance of a simple alarm?

4.7 Availability: a Practical Example

What does availability mean for a real system? How does availability depend on the maintenance policy chosen for the protective system?

To answer these questions we will calculate the availability of a fire alarm system during one calendar year. The alarm is known to be working at the start of the year, but it fails a few moments after midnight in the morning of 1 April. In this first example, the alarm is not checked again until the end of the year. What is its availability over the year?

Let us be clear about the sense of the word "availability" in this section. We do not mean, for example, "Does the alarm function when a cigarette starts a fire on 18 July?" The availability that we want to determine is the probability that the alarm would operate if a fire occurred on a randomly chosen day during the year. We assume that no fires actually occur during the year.

In this first example, the alarm system is operational from 1 January to 31 March. It fails, but the failure is hidden because no fire occurs. The failure is discovered at the end of the year and the alarm system repaired.

What is the availability of the alarm over the year? Because we have the benefit of perfect knowledge, we know that the alarm was operational from 1 January to 31 March, or 90 days. The system availability is therefore

$$\frac{90}{365} = 24.7\%$$

In the year we have chosen, the availability of the alarm system is poor. What effect can a different maintenance policy have on the availability achieved?

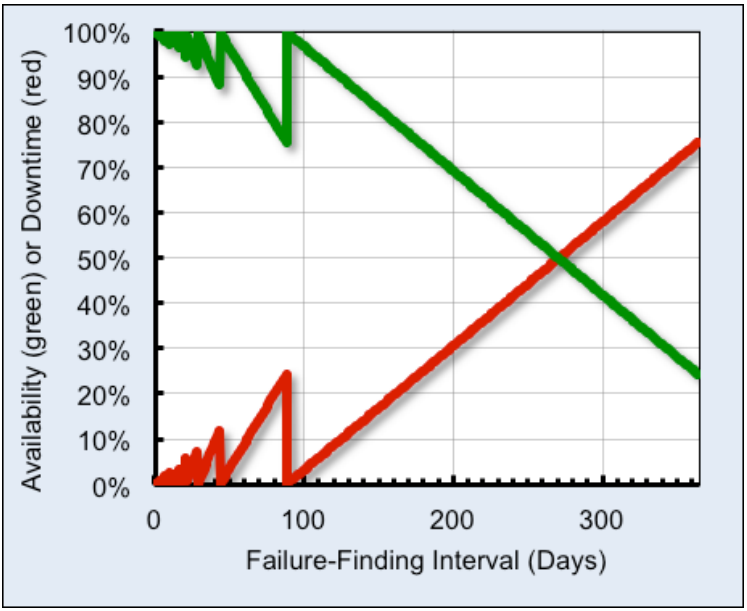
What is the availability if it is tested on 1 January and 1 July rather than just once per year? The system is now operational from 1 January to 31 March when it fails. Since the failure is hidden, it remains in a failed state until it is tested on 1 July. It is tested, found to be failed and repaired. For the sake of simplicity we assume that no further failures occur during the remainder of the year. What is the overall availability achieved?

The system is again available for 90 days to 31 March; from 1 April to 30 June it is unavailable (91 days); it is repaired on 1 July and is operational for the rest of the year (184 days). It is therefore available for 274 out of 365 days, or 75% of the time.

The table above summaries the availability achived for a range of task intervals.

Test interval	Operational Days	Non-operational days	Availability
1 Year	90	275	25%
6 Months	274	91	75%
1 Month	335	30	92%
1 Week	358	7	98.1%
1 Day	364	1	99.7%

The graph below shows how the alarm availability changes as the testing interval is increased from 1 day to 1 year. It may be surprising that the graph is not a smooth curve, but remember that we have made a number of assumptions. First, the device is checked on 1 January. We assume that there is a single failure on 1 April, when in a real situation we would have no idea when the failure might happen, because the failure is random and hidden. Finally, availability is calculated over the year to the end of 31 December, not over a long—or possibly infinite—period, as it might be in the models that we will use shortly.



Availability (green) and downtime (red) for failure-finding intervals from one day to one year, assuming that the device device has been checked on 1 January and that it fails on 1 April

The availability achieved has peaks and troughs depending on how close to 1 April the task is carried out. So if the task interval is 89 days, the first task after 1 January just misses the failure, so the failure is not found until the second task, resulting in downtime of nearly 25%. If the task interval is 91 days, the first testing task catches the failure, and downtime is only a single day over the year.

In a similar way, the table below examines the effect of increasing the testing frequency.

This exercise demonstrates that there is a relationship between availability and testing frequency: ignoring for a second the peaks and troughs shown on the graph, the protective device spends less time in an undetected failed state if it is tested more often, and so a higher overall availability is achieved.

As has already been pointed out, this section is in some ways a fraud because real life is very different from the simple example above.

First, the assumption that the alarm fails on 1 April is unrealistic. If we knew that the device would fail on 1 April, we would intervene in some way to provide continuous alternative protection or to repair the alarm as soon as possible. There are two reasons why this assumption is unrealistic. First, unless the device has a very well-defined lifetime, we have no idea exactly when it will fail. Second, because the failure is hidden, there is no way for us to know that the failure has occurred except to test it.

Second, we have assumed that we can "re-run" the same year's history with different task intervals, certain that the failure will occur on 1 April every time. If failures of the protective device occur at random, then history is absolutely no guide to the future, and no one year will be like the one before or the next.

Unrealistic as it is, the example does demonstrate one fundamental principle very clearly: that protective device availability is not a property that is fixed by the manufacturer and over which we have no influence. In summary,

If the protective device works when it is first installed, its availability is entirely controlled by our maintenance policy.

This is why it is vital to pay close attention to the failure-finding interval and to the way in which the task is carried out. Calculating the failure-finding interval is the core subject matter of section 2.

4.8 Key Points and Review

A protective device is designed to initiate a response if an unusual condition (the demand) occurs.

The protective device is usually designed so that, if it performs correctly, it reduces or eliminates the consequences of the demand.

A multiple failure occurs if a demand arises when the protective device is in a failed state or disabled in some other way (a fire occurs but the fire alarm is broken or turned off, so people are at increased risk of death or serious injury).

Protective devices can fail to operate when required (the multiple failure); they can also operate when they are not required (spurious operation).

Choosing failure-finding as a maintenance policy for a protective device means that the device can be in a failed state for an extended period, so a multiple failure could occur.

In the simple model presented in this chapter, the availability of a device can be increased by checking its operation more frequently.

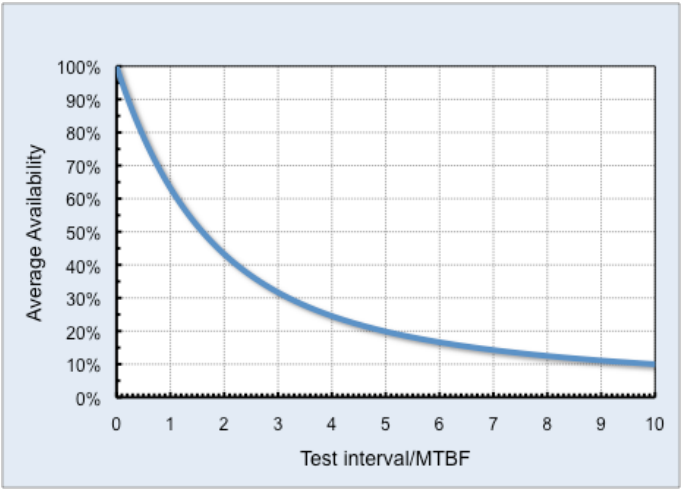
The rate at which multiple failures occur can be managed in two ways: by increasing the availability of the protective device (for example, by checking it more frequently); and by reducing the demand rate (possibly by maintenance on or redesign of the system which causes the demands).

The objective of any management policy is to reduce the chance of a multiple failure to a tolerable level.

5 The basis of decision-making

5.1 Introduction

In earlier chapters we saw that a protective device's availability is not fixed, but it depends on its reliability and how frequently it is tested. For an idealised single device that is expected to fail at random, its average availability falls continuously from 100% as the testing interval is increased.



This relationship enables us to choose a task interval that delivers the minimum average availability that is needed. This chapter focuses on the question

How do we decide what protective device availability is needed?

The following sections show that there are three different approaches:

- Use an availability figure determined by detailed quantitative modelling
- Specify the minimum allowed mean time between multiple failures
- Choose the availability that delivers the lowest cost to the organisation

5.2 Availability

If it is possible to set a target availability for the protective device, then it can be used directly to calculate the required failure-finding interval. The formula used to work out the device's availability must take into account its configuration and technical characteristics, but in principle it is easy to define the right testing interval.

First we are going to look ahead to a later chapter where we find that the average availability of a simple, single protective device that fails at random is given by the following formula.

$$A = \frac{M_{dev}}{T}\left[1 - exp\left(-\frac{T}{M_{dev}}\right)\right]$$

The terms in this equation are

A	The device's average availability over time
M_{dev}	The device's mean time between failures
T	The failure-finding interval

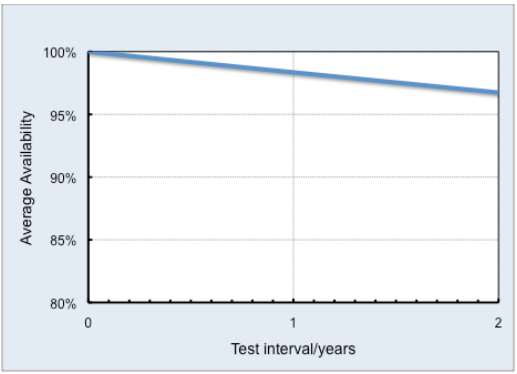
If we know the required average availability and the device's mean time between failures, the required test interval is easily calculated, or it can be found by drawing a graph of availability against test interval and finding the interval that gives the required availability.

Example

A single level switch is used to sound an alarm if the liquid level in a storage tank rises above the permitted high level. The switch is thought to fail randomly and its mean time between failures is at least 30 years. The average required availability is 99%.

Based on these figures, we need to find the failure-finding interval T for which

$0.99 = (30/T)\ (1–\exp(–T/30))$



The required interval is 0.6 years, or about 7 months. In practice the task would probably be carried out every six months to simplify maintenance scheduling.

The calculation is simple and it only requires two items of information: the mean time between failures of the protective system and the required availability.

Although the device's mean time between failures may not be known with absolute certainty, it is usually possible to find a worst case, lower bound by using maintenance records, manufacturers' data, or information that is available from generic industry databases. But where does the required availability come from?

Sometimes the required availability can be found in equipment or system documentation, particularly if the analysis involves an asset that has been subject to a rigorous, quantitative risk analysis using techniques such as fault tree analysis (FTA). Sadly, no easily accessible availability target exists for most industrial equipment, and we have to do a little more work before we can calculate the failure-finding interval.

Availability is the simplest criterion that can be used to derive a failure-finding interval, but it should not be used unless a robust quantitative model is available which justifies the chosen value.

5.3 Tolerable Risk

In the previous section we saw how easy it can be to calculate a failure-finding interval using just two pieces of data:

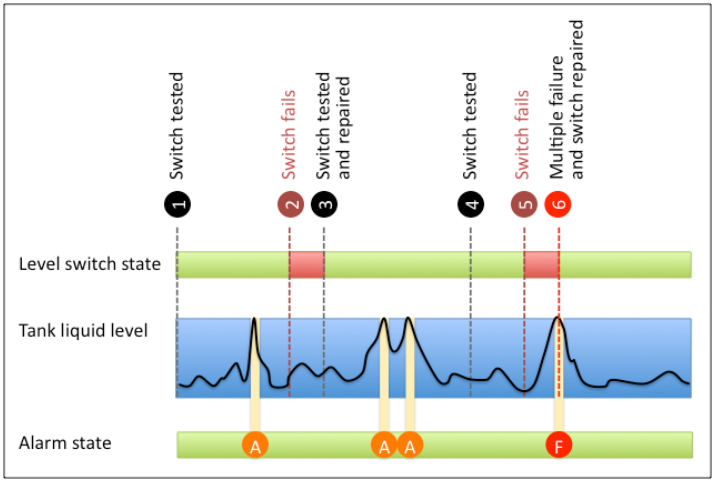
- The device's mean time between failures
- The required average device availability

Unfortunately we also saw that the required availability is not usually known with any certainty. How can we calculate it?

Chapter 4 introduced the concept of *multiple failures*. A multiple failure occurs if a demand is made on the protective system while it is in a failed state; in other words, while it is unavailable.

In the example above, a demand on the protective device occurs if the liquid level in the tank rises above the alarm level. An alarm sounds if the level switch is working; if it has failed, then we have a multiple failure that might lead to a process trip or liquid escaping from the tank.

The diagram below shows a possible history of the tank alarm system and tank liquid level.



The switch is tested regularly (at points 1, 3 and 4 in the diagram). It is working at the start of the time line, but at point 2 it fails. When it is tested at point 3 it is repaired; during the interval between points 2 and 3 it is in a failed state and could not sound an alarm if the liquid level were to rise, but there is no abnormally high level and so there is no multiple failure.

The tank level rises above the alarm level at four different times. The first three times, the switch is working (available) and an alarm is sounded each time.

The switch is tested at point 4, but it fails at point 5. From this time onward it is unavailable, and before it is tested again, the liquid level rises but no alarm is sounded. The multiple failure occurs at point 6.

We know that the availability of the level switch can be increased by testing it more often, but it is impossible to make it function continuously unless we also check it continuously. One way to decide what level of availability is needed is to set a maximum tolerable rate of multiple failures.

Example

Suppose that the tank is overfilled on average three times a year. These occasions occur at random, so it is not possible to know when the alarm system might be needed. If the availability of the level switch is 90%, then the probability that the alarm sounds each time is 90%, and the average number of alarms per year is

90% x 3 = 2.7 per year

Conversely, the average number of multiple failures—tank overfills that do not result in an alarm—is

10% x 3 = 0.3 per year

Of course this is just the average number of multiple failures per year. In reality there may be zero, one, two, three or even more per year, but the average over a long period of time should be 0.3 per year.

We can use this multiple failure rate (0.3 per year in the example above) to set the required availability. If we increase the availability of the protective device, then the number of multiple failures per year is decreased. So instead of choosing the availability of the protective device directly, we ask

How often, on average, are we willing to tolerate a multiple failure?

If the mean time between demands (the average time between tank high levels in this example) is M_{dem}, and the average protective device availability is A, then the average number of multiple failures (undetected tank high levels) per unit time is

$$\frac{1}{M_{dem}}(1-A)$$

If we decide that the minimum mean time between multiple failures that we are willing to tolerate is M_{mf}, then by rearranging the equation above, the required availability is

$$A = 1 - \frac{M_{dem}}{M_{mf}}$$

Now it is easy to see why availability is a poor criterion for setting failure-finding intervals unless it is based on a robust quantitative model. If we pluck a required availability level out of the air, then we make the assumption that the resulting multiple failure rate is tolerable. But the mean time between multiple failures achieved depends on both the protective device availability and on the demand rate; because the multiple failure rate is what is ultimately important to us, we always need to take into account the demand rate.

For multiple failures that have safety or environmental consequences, use the required mean time between multiple failures to determine the device availability that is necessary

Example

A crane's overhoist protection switch is designed to stop upward movement of the load if it goes beyond a set position. If this switch were to fail when required, the crane would be damaged and its load could drop 20m to the floor below, possibly injuring or killing several workers.

The manufacturer's data suggest that the minimum mean time between failures of the switch (failure to operate when required) is 150 years. An overhoist condition that activates the switch occurs about once every five years.

After discussion, the analysis group agrees that the multiple failure should occur no more often than once every million years.

The required average availability is therefore

$$A = 1 - \frac{M_{dem}}{M_{mf}} = 1 - \frac{5}{1000000} = 99.9995\%$$

Using the formula discussed later in this book, the group concludes that the switch should be tested twice per day.

…and the data?

Now we have a way to calculate the required device availability, but at the cost of needing two numbers rather than one:

- The rate of demands on the protective device (or the mean time between demands)
- The minimum tolerated mean time between multiple failures

Demand rates can vary over a huge range. Some protective systems are activated several times per day, while others may never be used over a period of decades. It is obviously relatively easy to find the demand rate for systems that are activated frequently, but data for rare demands may need research or may have to be estimated.

Specifying the shortest tolerated mean time between multiple failures can be far more challenging. How do we decide whether we should tolerate one failure per year, per century, per millennium, or per million years?

The tolerated multiple failure rate depends on a number of factors, including:

- The effects of the multiple failure
- The number of possible serious failures for which the organisation is responsible
- Who would be exposed to the multiple failure
- Constraints imposed by law and statutory bodies

The issue of tolerable risk is the subject of the next chapter.

5.4 Economic Basis

The previous section considered how the required protective device availability can be calculated if we know the demand rate on the system and if we specify a minimum tolerated mean time between multiple failures. The issue of how to determine the level of risk that can be tolerated was left for a later chapter.

The concept of "tolerated risk" can be applied to a range of failures that have safety effects.

Boiler and relief valves

Boiler pressure is limited by two relief valves. The required relief valve availability is determined by how often the boiler pressure exceeds a safe level and the tolerated mean time between unrelieved pressure excursions that could result in a boiler explosion.

Turbine overspeed system

A turbine overspeed system should shut down the turbine if its speed exceeds a safe level. The overspeed system availability is determined by how often overspeed events occur and the tolerated mean time between undetected overspeed events which may lead to serious damage and possible injury.

It can also be applied to failures that have environmental consequences.

Tank ultimate level switch

A tank ultimate level switch should shut down the supply pump and the upstream process if the tank level exceeds 30cm below the overflow. Overflowing effluent from the tank could lead to a reportable environmental incident. The level switch availability is determined by the rate of demands on the switch and the minimum tolerated mean time between environmental incidents.

Now consider applying the same technique to this example.

Pump low supply pressure switch

A low pressure switch is intended to shut down a pump if the suction line pressure drops below a set level. If it fails to trip when it is required, the pump could be damaged with a potential cost of about $1500, and about two hours' production would be lost, with a value of about $3500.

> The low pressure switch availability is determined by the rate of demands (how often the suction line pressure is low) and the minimum tolerated mean time between undetected low pressure events.

Although it is difficult to answer the question,

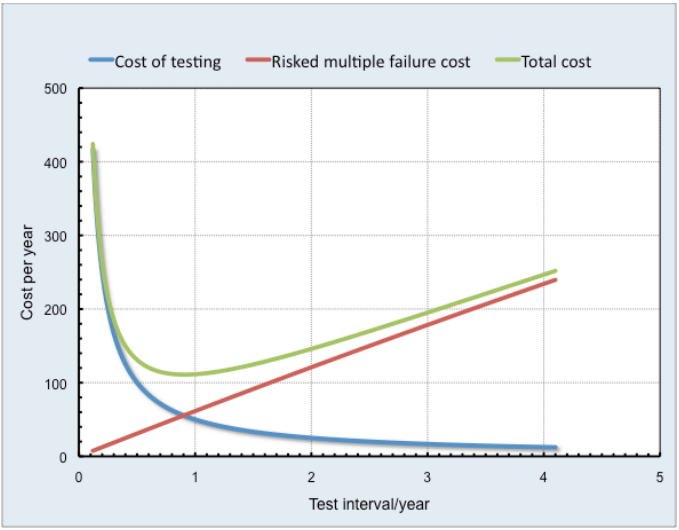
"How often are we willing to allow boiler explosions to occur?"

it is at least possible, perhaps after some discussion, to define an intolerable range of risk. However, if we now ask the question,

"How often are we willing to experience undetected pump low pressure events?"

it is not at all clear where the range of "tolerable risk" lies. On one hand, it is obvious that we should ensure that there is some level of protection against these events, because the potential economic costs are not insignificant. On the other hand, if the risk of a multiple failure were reduced to a very low level, the organisation would spend far too much on frequent testing of the pressure switch. Although we know that the extreme limits (high availability with too much testing or low availability with too little testing) are both undesirable, it is not possible to be sure where the right availability level is to be found.

This example suggests a different way to deal with multiple failures that have only economic (monetary) consequences. If testing infrequently results in unacceptable damage and downtime penalties, but the cost of very frequent tests is too high, then presumably there is a testing interval that results in the a lower expenditure than the two extremes. This is a balance between testing costs (which increase directly with testing frequency) and the risked costs of the multiple failure, which increase with lower protective device availability.



The graph above shows the relationship between testing costs and risked downtime costs as the failure-finding interval is increased. The total cost—the cost of testing plus the risked cost of downtime—has a broad minimum, in this case at a test interval of just under one year. This represents the lowest cost to the organisation, and is therefore the optimum testing interval.

To calculate the optimum failure-finding interval of a simple protective system we need four pieces of information.

- The mean time between failures of the protective device
- The mean time between demands on the protective device
- The cost of a single failure-finding task
- The cost of a single multiple failure event

Details of the calculation are given in a later chapter.

5.5 Key Points and Review

Three criteria can be used to set failure-finding intervals for protective devices.

Availability	The average availability required from the protective device
Mean time between multiple failures	The minimum tolerable mean time between multiple failures. This is usually applied to failures that have safety or environmental consequences
Lowest overall cost	The interval selected minimises the overall cost by balancing the cost of testing the protective device against the risked cost of damage and downtime if the multiple failure occurs

Availability is generally a poor criterion for setting failure-finding intervals unless there is a pre-existing detailed design or some other robust justification for selecting a specific minimum level of availability.

6 Tolerable Risk

6.1 Introduction

The management of systems that protect us against major incidents depends fundamentally on one number: the tolerable level of risk. This chapter is about how to determine what risk is tolerable, and who should make the decision.

It is difficult to reach a consensus on risk without complete openness and honesty, and in that spirit I am going to tell you now how you may feel after reading this section: dissatisfied. This isn't a set of rules that enable you to get directly to a single, right answer. Instead, each section is intended to help you to decide which factors are important, which can be given less weighting, and to point to techniques that help your organisation develop defensible risk requirements.

When I get frustrated with the difficulty of navigating through all the questions involved, I try to look at it another way. We are working in a unique area where some of the most important industrial decisions are made, affecting the lives of people we know and of millions that we don't know. We are trying to solve a problem that brings together engineering, mathematics, psychology, ethics, economics, business management and the law. Difficult? Yes, it is, but you won't ever be bored.

6.2 How dangerous can we be?

Management of protective systems brings with it a troubling question, one that very few people really want to answer.

How often are we willing to allow the ultimate multiple failure happen?

To put the question more directly:

> ***How often are we willing to injure and kill our employees and members of the public because of our own activities?***

If the mathematics of risk looks daunting, then I have some bad news. Setting levels of tolerable failure, and applying them consistently, is far more difficult.

Answering the question is difficult. Researchers, national organisations and employers have struggled with risk for nearly a century. I can't give you a neat flow chart that leads to a single number; but we can try to see what works and what doesn't, who needs to make the decision, and finally help to build a strategy that is better than pretending that the problem doesn't exist.

Some industries and some organisations make answering this question a core part of their risk management policy. Most don't.

This chapter is about finding our way through this problem, because otherwise it doesn't matter how well constructed and detailed our mathematical models are. Without an answer, management of protective devices is at best based on guesswork.

In chapter 5 we found that hidden failures and protective devices can be managed to achieve one of three targets:

- The average availability of the protective device
- A maximum tolerable rate of multiple failures
- An optimum balance between the cost of maintaining the protective device and the risked cost of multiple failures

The first option, managing the availability of a protective device, is relatively simple. The challenge here is to determine *why* a specific availability level is required. What are the reasons that availability should be 95%, or 99%, or 99.999%?

The third option, where the multiple failure has no safety or environmental consequences, enables an optimum failure-finding interval to be calculated fairly easily from reliability data, the cost of a multiple failure and the cost of carrying out a simple failure-finding task.

That leaves the second target: the maximum tolerable rate of multiple failures. In other words, how often are we prepared to allow the ultimate failure to happen? It is the tough question, so here goes.

6.3 Zero Risk?

The immediate reaction of most employees and of the general public is that no risk of serious injury or death is acceptable, and that everything must be done to reduce exposure. There must be no risk to me, my children, my community or my co-workers.

While this view is understandable, the harsh reality that is almost everything we do involves risk in some way. Driving (WHO, 2018), swimming (Chase *et al.*, 2008), flying (Ranter, 2017), playing football (Gouttebarge, 2014), using electrical equipment (Taylor *et al.*, 2002), drinking alcohol (IARC, 2012), skiing and snowboarding (Davidson and Laliotis, 1996), eating raw meat and even remaining unmarried (Harvard Medical School, 2010): nothing that we do is without at least some risk. Industrial activity is no different, so how do we decide whether the risk that it contributes is acceptable?

This chapter focuses on two core issues. The first is who should be involved in making the decision, and the second is how to determine a tolerable risk level.

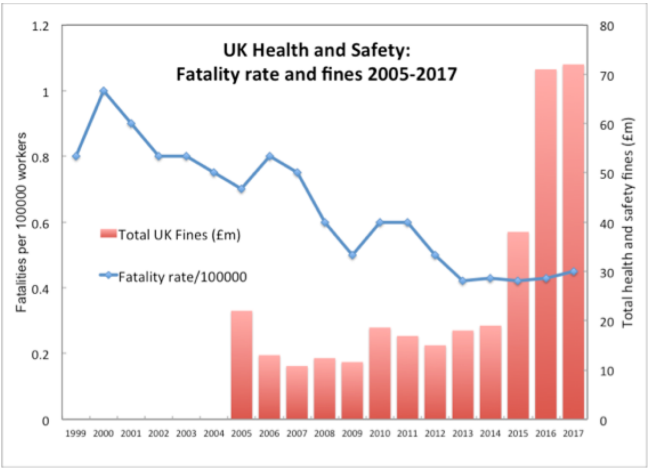
6.4 Who should decide?

Possible victims

Serious hazards affect the potential victims most directly: operators, technicians and maintainers, but also non-technical and administrative staff. This group feels a number of competing pressures. While they have a clear personal motivation to reduce the risk as far as possible, they are also keenly aware that their employment depends on continued operation of the plant or process.

Managers and owners

Managers have a responsibility to manage the risk to employees while at the same time containing costs and providing a reasonable return to the organisation's ultimate owners. They may be responsible for overseeing the safety of many processes, and need to ensure that money is deployed where it will provide the greatest risk reduction.



Managers and owners in modern organisations have a specific reason to take their responsibilities seriously. In response to society's increasing pressure for responsible operation, managers of dangerous assets have stripped of their ability to hide behind limited liability corporate structures. Negligence can bring them into a law court.

The public

Almost all industrial facilities need to consider their role as "good neighbours" to those who live around the plant and who could potentially be involved if a major incident occurred.

Society

Society expresses its expectations for safety and environmental risk through laws and regulations. While laws set up a general framework, much of the detailed responsibility for monitoring, surveillance and expert advice is delegated to bodies such as OSHA and the UK HSE.

The role of statutory bodies can vary. For example, some bodies prescribe specific maintenance tasks and intervals for common equipment such as lifting gear. More often the requirements are far less definite, perhaps referring to "industry norms" or "best practice".

The one figure that we would like to be given almost never appears in laws or regulations: a definitive maximum tolerable risk[2]. Government bodies may have overall targets for hundreds or thousands of different risks—for example, fatalities in the construction industry—but they are used for monitoring and to assess the effectiveness of regulation and changes in working practice. It is more common to impose requirements on the asset owner or operator to develop a safety case, to compile risk analysis and other documents that list and quantify possible hazards and justify the operator's risk management measures.

Even if regulations have little direct information to give us on acceptable risk, it does raise an important issue: modern industrial processes may only be operated if they conform to society's expectations. Therefore we have to be certain that, whatever the engineering or mathematics may say, our recommendations for managing risk conform to current and future legislation.

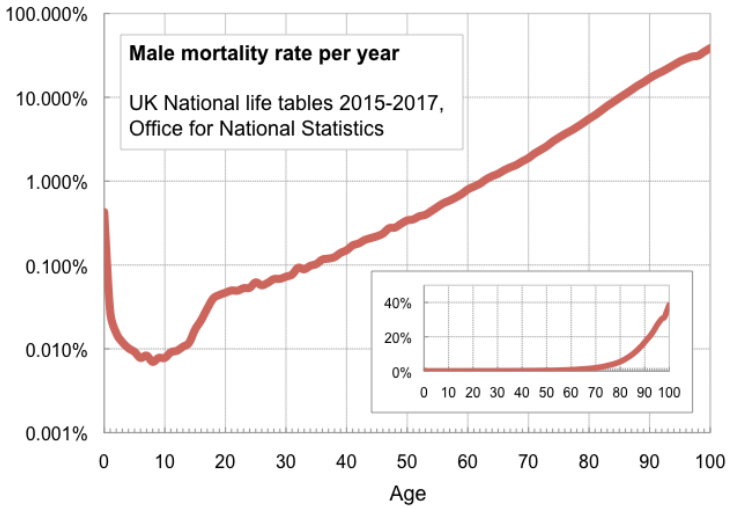
6.5 A Baseline

If zero risk is an impossibility, if death and injury don't go away even if we stay at home, drink water and eat lentils, is there a way to get some idea of what might be within the tolerable range?

Pick someone at random and ask what their acceptable level of risk is. The chances are they won't know. That isn't a surprise; I cannot think of anyone who uses some absolute standard to decide what he or she will and won't do. But I might stand a chance of getting a more definite answer if I ask about a specific activity: think about the risk of going to watch a sports match, flying on a scheduled airline, or climbing K2. The answers will differ from person to person, but I am far more likely to get a "yes", "no" or "maybe".

Although we don't have a built-in idea of absolute risk, we do understand it in a relative way.

[2] Maximum tolerated risk is not often stated explicitly by statutory bodies, but there are some exceptions. For example, the UK Health and Safety Executive provides advice on residential land planning applications that involve building close to known hazards. In this case it applies specific risk contours to the areas around the hazard. See *UKHSE's Land Use Planning Methodology* (no date)



In normal circumstances life is fairly safe. Exactly how safe is something that most governments measure by recording birth and death statistics to create something like the graph above, taken here from official UK life tables. It shows the probability (strictly the *rate*) of death in any year taking into account almost every male in the UK. The data includes all causes: disease, accidents, old age, industrial incidents and everything else.

Mortality rate is sometimes used by reliability engineers as an example of the "bath-tub" failure curve because being old and being young are both more risky than somewhere between the two. In reality that picture is inaccurate; the curve is far more like a wear-out "Pattern B" trend (see the linear curve inset above). The detailed sort-of-bathtub picture only appears if you plot the rate on a logarithmic axis (main chart). Being very young is risky, but the overall (and slightly depressing) trend is of steadily increasing mortality after your eighth birthday.

This relationship between age and risk even has a name: *The Gompertz Law of Human Mortality*, after the actuary who commented on it in 1825. The law says that roughly speaking, whatever your chance of dying this year, it will be twice as high in eight years' time.

For an individual at the start or in the middle of working life, the chance of death in any year is somewhere around 1 in 1000. Although it isn't a number that is at the forefront of anyone's mind, it's one that fits with experience: if I'm 30 years old, deaths among my friends and acquaintances are rare, but not totally unknown. This gives us an anchor for decision-making: a risk of around 100 per 100,000 per year for the general population. If I'm considering an office or retail job (occupation-related mortality about 2 per 100000, US Bureau of Labor Statistics, 2012), the additional risk probably doesn't play a big part in my decision. If I'm considering becoming a truck driver (24 per 100,000) it might; and if someone tries to persuade me to move into offshore fishing (120 per 100,000), more than doubling my basic mortality rate, it could be my main concern.

Risk could be my main concern, but it might not matter so much if the rewards compensate in some way for the risk. This second part of the equation—evaluating the benefit of the increased risk—is what makes decisions so controversial. Even if we agree on the magnitude of risk, we all evaluate the benefits to ourselves, our families and our community in different ways.

6.6 Comparing Risks: Voluntary Hazards

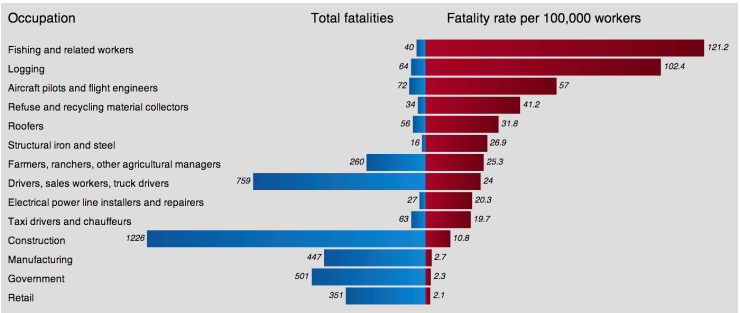
Sometimes considering everyday risks—activities undertaken voluntarily and usually without payment—can help decision-makers to focus on workplace risks.

Injury and mortality rates for a variety of sports are easy to find online. The table below summarises approximate mortality rates for a few of them.

Activity	Annual Mortality Risk	Source
Cycling	1 in 90,000	Turk *et al*. (2008)
Swimming	1 in 56,000	Turk *et al*. (2008)
Soccer	1 in 100,000	Turk *et al*. (2008)
American football	1 in 182,000	Cantu et al. (2003)
Canoeing	1 in 750,000 outings	UK HSE
Scuba diving	1 in 200,000 dives	UK HSE
Travel by car	1 in 6,700	UK ONS
Travel by motorcycle (2018)	1 in 2800	UK Police Federation
Smoking (adult life)	Roughly doubles mortality at most ages	Sakata *et al*. (2012)
Hang gliding	1 in 120,000 flights	UK HSE

6.7 Context is Everything

If you still think that there is a single standard for risk at work or that there should be one, take a look at the chart of US occupational fatality rates below. The right-hand red bars show the fatality rate per year per 100,000 workers, and the blue bars the total number of fatalities per year (US Bureau of Labor Statistics, 2012).



US Occupational fatalities and fatality rates per 100,000 individuals

The difference between mortality is striking: for example, there is a factor of about 50 between the rates in retail and fishing. Any pretence we may have that consistent standards apply to all workers is obviously wrong. Interestingly, although there may be some link between the physical risks that people take and their salary, risk in the agricultural sector shows that the relationship is not a simple one.

In fact, even within the same industry, risk levels can vary massively between jobs; for example, offshore industry drilling personnel are usually exposed to far higher levels of risk than catering and laundry staff on the same platform.

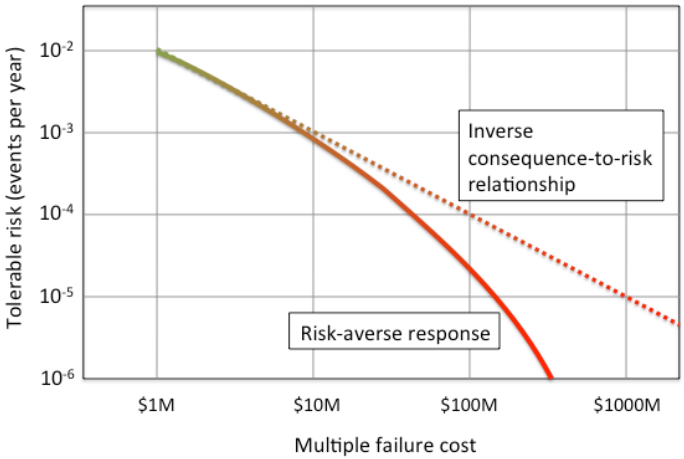
Could we apply the same standard of risk to retail and shipping, or to taxi drivers and loggers? On one hand, the chart suggests not. To someone outside the industry, it can be difficult to envisage a way to achieve substantial risk reduction without simply abandoning that activity. Perhaps—just perhaps—we can imagine a world without offshore fishing, but probably not a world without roofers or one that is limited to single-storey construction.

6.8 Magnitude and Type of Consequences

The most obvious factor that determines tolerable risk is what happens when the multiple failure occurs. In other words, the magnitude of consequences directly determines the level of risk that is tolerable.

Even so, the relationship between tolerability and the magnitude of consequences is not what you might expect. If a company accepts a one-in-a-hundred risk of a failure that costs $1m, what standard should apply to more expensive failures? Logic says that it should accept a chance of 1 in 1000 years for $10m, and 1 in 100000 years for a catastrophic $1bn event.

That isn't what actually happens.



The inverse consequence-to-risk relationship is a straight line. In reality, the response becomes risk-averse at a value that depends on the resources of the individual or organisation exposed to the risk

Individuals and organisations become more risk-averse as the magnitude of consequences increases. The relationship between an event and the tolerable level of risk is almost never linear. It becomes more *risk-averse* as the consequences become more severe.

Let me introduce you to Jim. One morning he sets off for work with a ten dollar bill in his back pocket. In the evening it's gone. Perhaps he dropped it when he pulled out his phone, or someone might have grabbed it while he was on a crowded train. He kicks himself briefly, opens a beer and carries on with his life. When the same thing happens to a $20 bill a few weeks later, he's twice as maddened, opens two beers and sits in front of the television. What he's not doing is calling his insurance company to arrange cover.

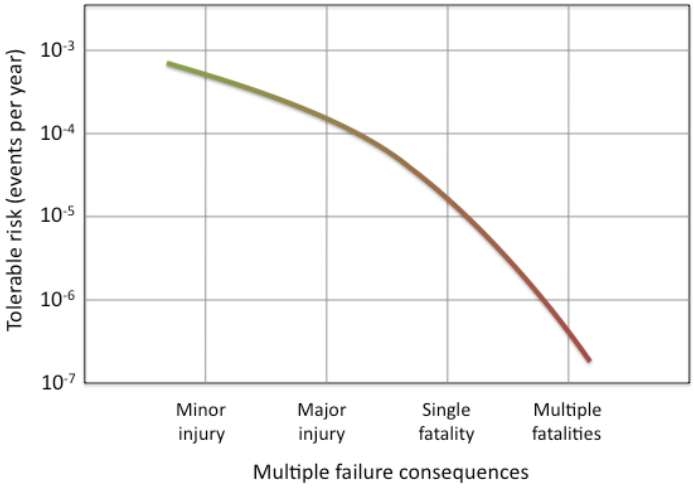
A few days later Jim gets his annual renewal notice for property insurance. Jim's home has a nice outlook surrounded by trees and it is worth about $1M. The chance of losing a home to fire varies depending on where you live, but in Jim's area it is around 1 in 10000 per year. His risked loss per year is $100 ($1M/10,000), which is does not sound any more annoying than losing some money from his back pocket. Jim didn't consider insuring his ten dollar bills, but he's made certain to insure the house against fire.

Why?

Losing $100 in cash every year would be embarrassing, but losing a $1M home would be different. Jim borrowed the money to buy the house, and he would be losing money he doesn't have. There may be a really small chance of the fire happening, but if it did, it would be a disaster. So he's happy to pay the insurance company for cover, even though they charge him far more than $100 per year. A small chance of a $1M loss is well past the point where he becomes risk-averse.

The same principle applies to safety-related issues. An organisation may be willing to accept that an event causing a single fatality could occur every million years. When the same organisation analyses one event that could cause ten fatalities at once, it is very unlikely to tolerate one event per ten million years.

Partly this is because of an innate response to severe events: a single fatality seldom reaches national news media, while an incident that kills ten individuals probably would; a hundred deaths would be considered a major national disaster. Beyond this subjective response, an organisation that is responsible for a single fatality would face a detailed inquiry, but multiple fatalities attract the attention of statutory health and safety bodies and insurers, with the possibility that those responsible may have to suspend or terminate operations until a detailed inquiry has been concluded.



A risk-averse relationship between the consequences of a multiple failure and tolerable risk. ***The vertical scale of this chart is for illustration only and should not be applied to a real-life analysis.***

The consequences of failure are not always immediate. For example, exposure to radioactive materials is known to increase the risk of individuals developing cancer. It is impossible to determine when that will happen, the threshold exposure level where risk increases, or even whether an individual will be affected at all. The development of effects may be even more distant if a hazardous material can give rise to genetic defects in an individual's unborn children. Individuals and regulatory bodies can be massively risk-intolerant if the *possible*, *imaginable* consequences of an incident are very severe, and particularly if they are also uncertain.

6.9 Personal Control

Given a choice of a long road trip—say, from New York to Los Angeles—by road or by taking a scheduled flight, which one would you choose?

Looking at the decision purely from a risk standpoint, the numbers are like this.

> The distance from New York to Los Angeles is about 3900 km or 2450 miles. The average mortality rate for scheduled passengers in the USA is 0.05 per billion kilometres, giving a chance of death on this trip of 1 in 5,000,000.
>
> I'm going to use the same distance for the road trip, although in reality it will probably be a few hundred miles further. The mortality rate for car drivers in the USA is 3.1 deaths per billion kilometres travelled. Over a distance of 3900 km, that makes a risk of about 1 in 80,000 for the trip.

Very few people choose to drive for days rather than take a five- or six-hour flight. But if you ask them how they feel about the risk, some will say that they feel more exposed to danger in the air than on the road. Perhaps that feeling comes from everyday familiarity with driving and from envisioning the scale of the consequences if something did go wrong in the air.

Part of what distorts our risk perception is something else: humans like to be in control, and we intensely dislike handing over control to other people or to machines.

The urge to be "in control" has some interesting effects on our evaluation of our own abilities. For example, it has been shown repeatedly that individuals consistently overrate their own driving ability.

> In one survey by Svenson (1981), 69% of a Swedish sample placed themselves in the top 50% of drivers. They were impressively beaten by a US sample, where 93% rated themselves in the top half of drivers.

Being in control means that we also believe that we are safer.

Control has a direct bearing on tolerable risk. In general we are more willing to accept a risk if we are in control of it, perhaps because we form part of a team that operates, maintains or manages an asset. The converse is also true: standards of tolerable risk are generally *more* strict when considering groups who have no control, such as those living in the immediate area close to a hazardous process.

6.10 Degree of Exposure

The role of exposure is obvious: not everyone is exposed to every hazard every day for twenty-four hours per day. Obvious, but sometimes easily forgotten.

If the equipment is operational all the time, but there are three shifts of eight hours every day, one individual is exposed to the hazard for only one third of the time. So an individual's tolerated risk of, say, 1 in 1,000,000 years could be achieved by a failure management programme that ensures an equipment failure rate of 1 in 330,000 years. On the other side, some people are exposed more or less continuously to hazards. This could include, for example, those who live close to a hazardous site.

> An extreme example of exposure is the difference between commercial aviation passengers and pilots. A passenger taking a flight every month would be exposed to an additional risk of death of around 1 in 1,000,000 years. The risk for a full-time pilot working on the same aircraft is around 1 in 16,000 years, depending on distance flown and shift patterns.

On the other hand, members of the community who live around an industrial facility may be exposed to the hazard for substantial parts of most days.

6.11 Levels of Risk

So far this section has discussed single, isolated events, looking at the factors that could determine a tolerable level of risk. But we know that working life can expose one individual such as an operator or maintainer to dozens of high consequence hazards.

Suppose that Alice decides that she will tolerate exposure to work-related fatality at a rate of no more than one in a million per year. If she is exposed to twenty life-threatening failure modes, then the risk of each one of those threats needs to be reduced substantially to achieve her overall target. If the risk is spread equally between them (that's an assumption: there is no rule that says that it must be), the rate of failure of each source must be less than one in twenty million years.

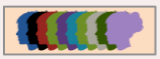
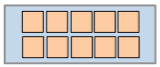
Alice is part of a group ten engineers who are each responsible for similar equipment with 20 failure modes. Her nine colleagues, who are exposed to the same types of hazard as Alice, have similar safety standards and they individually come to the same conclusion: that they would be satisfied if the rate of each failure were one in twenty million years.

Suppose that the company applies this one standard consistently across the organisation: an individual engineer is exposed to a fatality risk of no more than one in a million years. Remember that this is a simplified example; there could easily be several thousand potentially lethal failure modes on an industrial site, but in the real world they are likely to have very different failure rates and consequences, while employees will be exposed to them in different ways. To make the calculation clearer in this example, our hypothetical company consists of identical failure modes, engineers and groups of engineers all the way from Alice's office up to the global corporate level.

First there is Alice's group supervisor. He manages 10 engineers like Alice, and each of the engineers is exposed to 20 potentially lethal failure modes. If each of the engineers has a risk of 1 in 1,000,000 of fatality from these failure modes, then the fatality rate *for his whole group* is 10 times higher than for Alice, or 1 in 100,000 years.

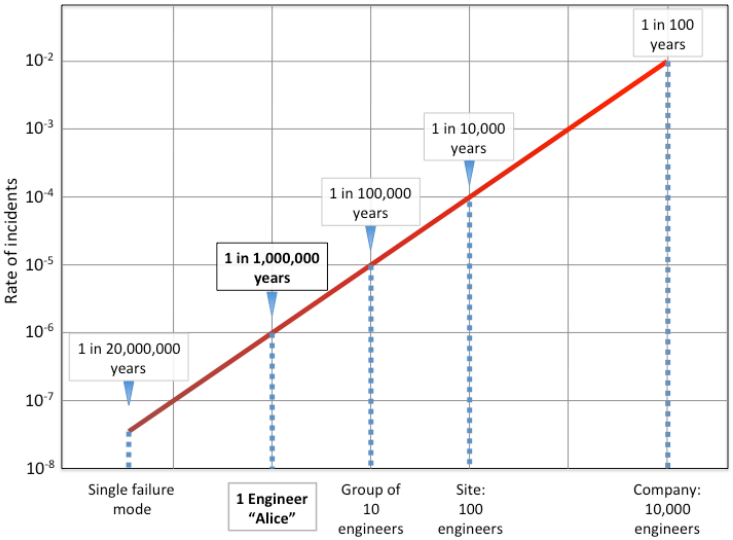
The site is made up of ten similar areas. Each of the areas has a group of ten engineers like Alice's, and each engineer is exposed to 20 similar failure modes.

If the other 99 engineers on the site have the same standard as Alice, on average there will be one death per 10,000 years somewhere on the site.

Level	Consists of…		Mean time between fatalities
Single failure	One failure mode		20,000,000 years
Alice	Exposure to 20 failures		1,000,000 years
Alice's group	10 engineers exposed to 20 failures each		100,000 years
Site	10 identical sets of equipment and engineers		10,000 years
Company	100 identical sites around the world		100 years

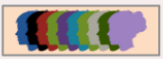
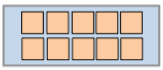
"Woman" icon by Peleg Red from the Noun Project

Notice what has happened as we move from Alice's personal viewpoint up to the site level. Her own standard of 1 death per million years sounds remote. It makes a negligible contribution to her overall personal risk, which might be 1 in 1000 per year if she is around 30 years old. Driving to work is probably far more dangerous. As we move from the individual to the group and then to the site level using the same standard of risk, the chance of a death shifts from being almost unthinkable (1 in 1,000,000 per year) to a remote possibility (1 in 100,000) to rare but possible (1 in 10,000 per year).



The company owns 100 similar sites around the world. Applying the same standard again means that someone, somewhere within the company may be killed once in 100 years. What started with Alice as a virtually unthinkable one-in-a-million year chance has now become a real possibility over a working lifetime. While Alice probably has far more important things to do than worry about her own risk, the company needs to put in place measures for managing the very real possibility that someone, somewhere may become a victim in any year. At this level, "management" is likely to be more than just technical engineering. At very least, the organisation will want to ensure that it has fully documented the hazards involved, that it has approved the design and maintenance decisions that achieve a 1 in 1,000,000 year standard, and that its insurance is adequate for the day it is needed.

Let's now look at the situation differently. What would happen if, instead of risking a fatality once in 100 years, the company's management wanted to reduce the risk of a fatality anywhere in the organisation to less than 1 in 1,000,000 years?

Level	Consists of...		Target mean time between fatalities
Single failure	One failure mode		200,000,000,000 years
Alice	Exposure to 20 failures		10,000,000,000 years
Alice's group	10 engineers exposed to 20 failures each		1,000,000,000 years
Site	10 identical sets of equipment and engineers		100,000,000 years
Company	100 identical sites around the world		1,000,000 years

Working from the corporate level downwards, the requirement for one fatality per million years at the highest level translates into one in 100,000,000 years at each of the 100 sites. At the group level (10 groups per site), that becomes 1 in 1,000,000,000 years, and 1 in 10,000,000,000 years for each of the 10 engineers in the group. Each engineer is exposed to 20 similar failure modes, and each failure mode should occur no more than once in 200,000,000,000 years on average.

Once in two hundred billion years is a difficult standard to meet for a number of reasons. First, and most obvious, is that the engineering involved would be complex and expensive. The second problem is uncertainty: when the tolerated risk becomes very small, it is progressively more difficult to be certain of meeting the standard. Highly unlikely events, obscure forms of human error, and (probably most difficult to analyse) common mode and common cause failures can cut orders of magnitude from theoretical risk levels. There are some exceptions, but in general, such low risk levels are very difficult to achieve.

6.12 Other Factors

Two other factors deserve to be given space in a discussion of tolerable risk. Both of them can be useful tools, but they can also distract from the primary aim of developing sound risk targets. Both techniques are described in far more detail in a later chapter, *Other Topics*.

ALARP

The concept of a risk that is ALARP (*As Low As Reasonably Practicable*, or sometimes *As Low As Reasonably Possible*) was the outcome of a landmark UK court case in 1949 that centred on whether an employer's responsibility was to *eliminate* every possible hazard or to do *everything practicable* to remove or mitigate dangers.

The judgement came down on the side of doing everything practicable, and the associated risk standard became known as *ALARP*, *As Low as Reasonably Practicable*.

Where its principles are applied with careful thought, the court's judgement still makes perfect sense. Unfortunately, ALARP can also be an excuse for inaction and acceptance of the *status quo*.

Criticality or Frequency/Severity Tables

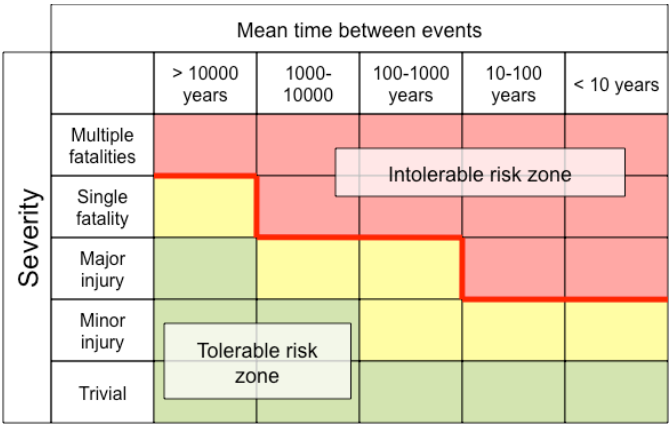
Criticality tables are intended to help analysis groups to summarise the overall risk associated in a failure as a single code. One dimension in the table represents the severity of failure consequences, the other its frequency of occurrence. The cell where the row and column meet contains a code that corresponds to the risk assigned to that failure mode.

		PROBABILITY OF OCCURRENCE				
		(A) FREQUENT $R_{POC} = 1$ $FR_{POC} = 2$	(B) PROBABLE $R_{POC} = 2$ $FR_{POC} = 3$	C) OCCASIONAL $R_{POC} = 3$ $FR_{POC} = 4$	(D) REMOTE $R_{POC} = 4$ $FR_{POC} = 5$	(E) EXTREMELY UNLIKELY $R_{POC} = 5$ $FR_{POC} = 6$
HAZARD SEVERITY CATEGORY	(I) CATASTROPHIC $R_{HS} = 1$ $FR_{HS} = 1$	1 (risk=4)	2 (risk=6)	3 (risk=8)	4 (risk=10)	5 (risk=12)
	(II) CRITICAL $R_{HS} = 2$ $FR_{HS} = 4$	3 (risk=8)	5 (risk=12)	6 (risk=16)	7 (risk=20)	8 (risk=24)
	(III) MARGINAL $R_{HS} = 3$ $FR_{HS} = 8$	6 (risk=16)	8 (risk=24)	9 (risk=32)	10 (risk=40)	11 (risk=48)
	(IV) NEGLIGIBLE $R_{HS} = 4$ $FR_{HS} = 16$	9 (risk=32)	11 (risk=48)	12 (risk=64)	13 (risk=80)	14 (risk=96)

UK Ministry of Defence 00-45 RCM Standard criticality table

Applications of criticality tables include:

- Reporting the number and proportion of high probability/high consequence failures in a system
- Identifying failures that pose an intolerable risk by drawing a boundary line on the chart
- Demonstrating that existing and proposed failure management policies reduce risks to tolerable levels



These tables can cause problems when they are used for serious risk analysis. A full discussion of these issues has been postponed until a later chapter. In summary:

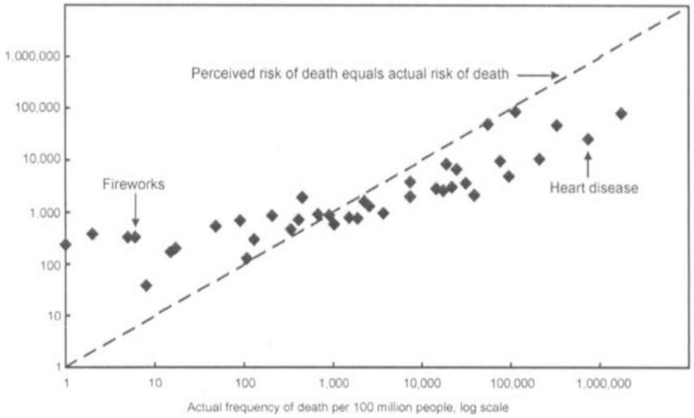
- Each row and column in the table represents a range of frequencies, injuries or financial costs.
- By allocating a single code to each failure, detailed information on frequencies and consequences is lost, and as a result reporting becomes less precise than it could be
- A more precise way to compare financial consequences already exists: it is called "money"
- Criticality codes hide uncertainty by encouraging analysts to pick a single cell, when in reality the consequences, frequency or both may be unknown within wide limits
- Time and attention is taken away from the failure analysis process as the review group tries to assign the "correct" criticality code to each failure
- Tables are context-specific: the frequencies and severity categories that apply in one context do not necessarily apply to another. For example, it would be unrealistic to expect the matrix for a food production plant to apply unchanged to an offshore oil platform or a pharmaceutical plant.

6.13 Human Attitudes to Risk

People are badly calibrated

How good are people at estimating the chance of an event occurring? The answer to this question is critical for us, because so many of the hazards that we need to consider are so rare.

A paper by Lichtenstein (1978) demonstrates that humans have a specific problem when they are asked to estimate risk: they overestimate the frequency of rare threats, and they underestimate the frequency of common events. By asking subjects to compare the likelihood of death due to, say, a firework accident with the chance of being involved in a fatal motor crash, they were able to tease out the way in which our perception deviates from reality when something is almost certain to happen and when it almost never happens.



Our instincts that work well for a 10% chance start to work badly when the chance is 0.001% or 99.999%.

There is more bad news for our faith in human calibration.

- People exaggerate the chance of extreme, spectacular risks such as terrorism and earthquakes, but they downplay everyday risks including tripping and driving
- As we have already seen, we tend to underestimate risks in situations where we have control, but overestimate them where we have little or no control

Increasing risk is unacceptable

Reducing risk costs money, and one of the frustrations for any manufacturer is knowing what risk reduction is worth to a consumer. While the risks involved in consumer goods are at a very different level from those in a manufacturing or process plant, a study by Viscusi (1987) has some interesting insights into the psychology of risk-based decision making.

> Subjects in the study were asked to consider two products, a toilet cleaner and an insecticide. Using either product was said to be associated with a small chance of harm which would require medical attention. The chance of these effects was stated as 15/1000, 10/10000 or 5/10000 per product package used. Subjects were told the current price of the product, and asked how much they would pay in addition for a product that reduced the risk to zero.
>
> As expected, almost everyone was willing to pay more for risk reduction, but the amount suggested was relatively small, typically between $1 and $4 for a $10 product, depending on the initial level of risk. The researchers then posed another question: how much price reduction would the participant want in order to accept a risk *increase* of 5/10000? Now the responses were very different. Everyone thought that the product would be too dangerous to buy at any price. When the researchers proposed a smaller risk increase of only 1/10000, between 60% and 75% of participants still refused to consider buying the product. Those who did respond wanted an average price reduction of about $5.50 for a risk increase of 1/10000.

The general lesson is that human risk response is asymmetrical: we expect to pay more (but not too much more) for risk reduction, but even a small risk increase is unacceptable. The impact on decision-making is important. When an expectation of risk has been established, it is far more difficult to relax the standard than to tighten it, even if the resources used to manage the hazard would be better employed elsewhere.

6.14 Key Points and Review

Deciding what level of risk is tolerable is not one person's responsibility.

The decision needs to be made by those who have some stake, including:

- Potential victims
- Employees
- Business managers and owners
- The general public

All applicable safety and environmental regulations have to be carefully considered.

Acceptable risk is influenced by the benefit that the process or activity brings to society and to individuals. A risk that is completely intolerable in many situations may be acceptable in others.

Perception of risk for a site or at the whole company level is very different from an individual's view. Hazards that are vanishingly rare for one person may happen frequently somewhere in a large organisation.

Individuals and organisations can be far more risk-averse when dealing with high-consequence events than might be expected from an inverse relationship between tolerable risk and consequences.

Event consequences and frequency for rare events are almost always uncertain. Be prepared to consider the full range of possibilities.

7 Writing Failure-Finding Tasks

7.1 Introduction

Once hidden failures have been identified and failure-finding task intervals calculated, the tasks themselves need to be written down in a form that can be carried out reliably by a technician, so that he or she will always do the right work on the right equipment at the right time. If the tasks are unclear, ambiguous or confusing, the time spent analysing hidden failures is wasted.

This section discusses some of the issues that determine whether translating task requirements into words and diagrams succeeds or fails. It is a complex area that needs to take into account the equipment and the engineers' experience, in addition to a wide range of other factors. A single chapter in this guide cannot cover the subject in any depth, but many resources are available in book form and online. An excellent high-level guide to writing clear tasks and warnings is published by the UK Health and Safety Executive (UKHSE, 1999).

7.2 Human Issues

Most failure-finding tasks are carried out by humans, and humans are fallible even when they are given perfect instructions. They become tired and distracted. Their sleep patterns, personal lives, and even the time of their last meal influence their attention to detail.

Many of the factors that influence the quality of failure-finding—ultimately the ability of people to follow written instructions reliably and repeatably—lie in the areas of psychology and industrial human factor engineering. This is a vast and growing field that is well beyond the scope of this book. However, a good practical summary of some of the issues involved can be found in Alan Hobbs' 2008 report for the ATSB (Hobbs, 2008). His description of the pressures on maintenance personnel is strikingly appropriate.

"From a human factors perspective, maintenance personnel have more in common with doctors than with pilots. We know from medicine that iatrogenic, or doctor-caused, injury can be a significant threat to patient health. Medical errors include surgical instruments sewn up inside patients, disorders being misdiagnosed, and very occasionally, surgeons operating on the wrong limb. Most aircraft maintenance personnel will be familiar with these types of errors.

"Opening up a healthy patient at regular intervals to check that organs are functioning normally would not be an appropriate strategy in health care, yet preventative maintenance in aviation often requires us to disassemble and inspect normally functioning systems, with the attendant risk of error."

7.3 The Curse of High Reliability

When your organisation sets up protective device maintenance, it assumes that the testing tasks will be carried out. But a serious issue affects protective devices: most of them are very, very reliable.

Why is high reliability a problem?

If a technician carries out a failure-finding task every month on a system with 99% availability, the task will detect a hidden failure on average about once every eight years. With 99.9% device availability, the tester would probably never report a problem in his or her entire career.

Here is the core of the problem. We demand high availability from protective systems, so only a very small proportion of tests ever find the system in a failed state. Almost all the test results are predictable. As a result, failure-finding tasks are sometimes ignored, or signed off as complete when the test has been missed or only partially completed. Perhaps understandably, incomplete or missed failure-finding tasks are frequently those that are difficult or uncomfortable to carry out, such as those in locations that are difficult to access, where the area is particularly hot, cold or wet, or where the technician has to wear awkward protective equipment.

Missing failure-finding tasks can quickly become endemic in a maintenance organisation because hidden failures by definition have no immediate consequences. Only the risk of a multiple failure is increased, with the actual increase depending on the protective system's configuration.

Failing to check a single oil pressure sensor 50% of the time could increase the rate of undetected lubrication incidents by a factor of two; not checking a pair of relief valves might raise the risk of overpressure events by a factor of four. If failure-finding intervals are not respected, the organisation's risk management is eroded but no one is aware of what is happening.

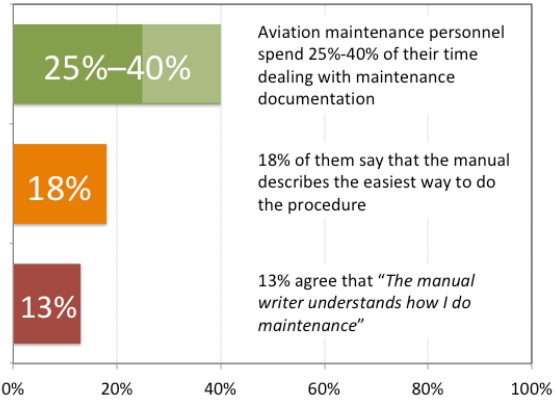
Poor maintenance completion rates are often an issue of company or department culture, but there are some steps that can be taken to improve failure-finding performance.

Monitor backlog

Failure-finding tasks that appear regularly in the backlog of late tasks should be investigated and appropriate measures taken to ensure that they are carried out on time.

Involve, inform and educate

In a survey among aircraft engineers cited by ATSB report AR-2008-055 (Hobbs, 2008), only 13 per cent agreed with the statement "*The manual writer understands how I do maintenance*". Individuals are more likely to follow procedures if they have been involved in writing them, or have been consulted before implementation.



While reliability analysts may understand that finding a hidden failure is expected to be a rare event, the technicians, operators and maintainers who actually test safety devices may not.

It can also help if those involved in the maintenance understand the direct link between failure-finding and the ultimate risk of multiple failures, and it can help them accept that "not failed" is an expected condition, not a reason for skipping the test.

Trust but verify

We need to manage risk effectively and that risk depends on the individuals who carry out failure-finding, so we have to trust our technicians, operators and maintainers. Imposing additional checks during or after failure-finding may seem intrusive and could quickly demotivate staff, making them feel untrusted and undervalued. Where additional checks are needed, they have to be based on a clear need for high availability that everyone involved understands.

- Independent checks are not a substitute for clear maintenance tasks. Start by reviewing failure-finding task descriptions. Ensure that the steps are clear, in a logical order, and that the conditions for failure are well understood.
- Prioritise hidden failure modes where the consequences of a multiple failure are particularly severe. Review these failure modes and estimate honestly the impact of failing to carry out a failure-finding task, or of leaving the protective system disabled after the task has been carried out.
- Consider adding appropriate independent checks and additional sign-offs for these critical tasks.
- Ensure that everyone involved in critical failure-finding tasks understands the seriousness of the multiple failures that the tasks are intended to manage.

7.4 Checking and Multiple Sign-Offs

Task sign-offs are a formal way to provide an audit trail demonstrating that critical tasks have been carried out correctly. The section above suggested that additional checks and sign-offs could be required for complex or critical tasks. In addition to providing an independent audit of the work, cross-checking can help to eliminate errors made because of fatigue or distraction. But will these measures guarantee that failure-finding is carried out exactly how and when it should be?

In the real world, unfortunately not.

First consider what the tester and the witness are being asked to confirm.

- The test has been carried out
- All steps of the test has been carried out correctly
- Any isolations and overrides have been removed, and the protective system is active and working at the end of the test

It is difficult to determine exactly how likely it is that a technician will fail to complete a task. In any case, the chance that the tester completes the task exactly as specified depends on dozens of variables: the complexity of the task, the engineer's experience and possible distractions as well as other factors. We do know that the chance is not 100%, and a review of figures (Smith, 2017) for simple tasks suggests that the chance of leaving at least part of the task uncompleted, but signing off the work anyway, could be at least 1%. Remember that this is a wild guess; the figure for a very complex task could be much higher.

If the chance of signing off incomplete work is 1 in 100, and we ask the technician's supervisor to check and countersign, how often will the supervisor not notice any issues and sign off incorrect work as if it has been done? Is that chance also 1 in 100, making the overall probability that the task is incomplete

$$1/100 \times 1/100 = 1/10000?$$

Again it is difficult to get firm answers, but it seems unlikely. The supervisor *expects* the work to be correct. In practice, the chance of a problem being found is much closer to 99% than 90%, so the overall chance of failing to carry out the task correctly could be

$$1/100 \times 1/10 = 1/1000$$

That is an improvement on single sign-off, but still a risk to be considered if the consequences of a multiple failure are severe. Specifically, if the protective device's failure-finding interval is attempting to achieve a device availability of more than about 99.9%, it is worth seriously considering the impact of incomplete tasks on the availability that is achieved in practice.

7.5 Conflicting Information

Conflicts cause confusion, and confusion leads to errors. If your failure-finding task is different from the manufacturer's description of the same maintenance, or if the engineer carrying it out has done it differently in the past, the outcome could be a lottery.

- Check for inconsistencies between the maintenance task you are writing and other information sources that the maintainer can access such as manufacturer's manuals and operating procedures
- Review any similar tasks in other areas of your organisation

- Consider the technician's experience and whether this task may be different from other, similar tasks that he or she carries out
- Compare both the steps and any parameters such as temperature, pressure, flow rate and so on that are specified in the task
- Try to resolve any conflicts that you find
- If not all the conflicts can be fixed (perhaps the new failure-finding task is deliberately different from the manufacturer's recommendation), make it absolutely clear which version of the task is to be carried out

7.6 Invasive Tasks

One of the reasons why RCM preferentially selects condition-based maintenance is that monitoring is generally less costly, less disruptive, and less invasive than fixed interval replacement and overhaul. Highly invasive maintenance is a common source of equipment unreliability.

Some failure-finding tasks cannot be carried out without a degree of invasive maintenance, including

- *Isolations* that prevent normal inputs from reaching the protective device
- *Removal* of components for specialist testing
- *Overrides* inhibiting normal actions when the protective device is triggered
- General *dismantling* and *disassembly* of protective systems to allow components to be tested

How could operators fail to recognise that a protective system is under test? It may be difficult to believe that it could happen, particularly in a small plant, but that is exactly what gave rise to the Piper Alpha oil platform incident. In that case a relief valve was not disabled; it was completely missing. The tag-out process failed to notify the operators, who started up a compressor without knowing that the relief valve was not there.

Where possible, non-invasive tasks should be preferred to invasive tasks. Where they are necessary, the following rules may help to reduce the risk of leaving a protective device in a failed state.

- The task should indicate how anyone operating or using the equipment is notified at the start and end of the test
- Where appropriate, tags should clearly indicate what systems are under test and special restrictions that are in place
- Operators and maintainers should understand how the system will behave while the protective system is in its "test" state, including any indications on the operator's console

- For every isolation or override made before testing starts there must be a balancing task step to undo the isolation or override
- Check that instructions for reversing isolations and overrides are clear and that they fully undo their effects
- If you use a flowchart to guide the engineer through a failure-finding task, ensure that isolations and overrides will be removed whatever path is followed through the diagram
- Consider adding one or more task steps that explicitly check that overrides and isolations have been removed

7.7 Stress and Wear Caused by the Task

Scheduled tests usually happen far more often than real demands on protective systems. If the system is stressed in some way by testing, then frequent failure-finding could itself be a source of problems, particularly if any of the associated failure modes are age-related. Examples include the following.

- Wear of moving parts in a fire pump[3] or emergency generator
- Fatigue caused by pressure or temperature cycling during a test
- Corrosion or erosion caused by exposure to product
- Damage to circuit boards caused by vibration during tests
- Wear of high voltage contacts caused by breaking or making a circuit under load

It is sometimes impossible to avoid stressing the protective system while carrying out a realistic failure-finding task. In these cases it is important to identify any failure modes that could be testing-related, and to ensure that the correct scheduled maintenance tasks are in place to detect or prevent failure.

[3] It is common for diesel-driven fire pumps to operate in two different modes depending on whether they are being tested or responding to a real fire. In "testing" mode, the engine shuts down as expected if it might be damaged by high temperatures or lack of lubrication. In "run" mode, when it is responding to a real fire, the engine would ignores most shutdown signals and continues to pump water until it seizes.

7.8 Failure-Finding: Writing the Task

Balance completeness, safety and practicality

Failure-finding means checking a protective device or system to ensure that it could operate correctly if abnormal conditions occurred. The most effective failure-finding tasks simulate failure of the protected system, so that the scheduled task tests the whole protective system: its sensors, any signal processing or control unit, and the final actuators or annunciators.

Testing the entire protective system under realistic conditions is an ideal that sometimes cannot be achieved because of practical considerations. As an example, consider writing a task to test a stand-alone local fire alarm system consisting of on a smoke detector, control unit and annunciator.[4]

Proposed Task	Completeness	Practicality
Push the "test" button on the control unit	Tests the annunciator and part of the control system. Does not test the sensor or parts of the control system input.	Easy to do. No risk of starting a fire. The only equipment required is a short ladder or testing prod.
Use a non-flammable smoke aerosol to simulate smoke	Tests the whole system	Relatively easy. Requires smoke aerosol. No risk of starting a fire.
Hold a piece of smoking paper under the fire sensor	Tests the whole system	Relatively easy. Requires simple tools. Slight risk of injury and could start a fire.

The first task is by far the easiest. The technician pushes a button and notes whether the unit's annunciator sounds. The task demonstrates that the annunciator works, but it tests only part of the control circuit (the test button is often wired separately to the microcontroller) and it does not test the detector and input circuit at all.

The third task is the most realistic: a small fire under the sensor provides a near-perfect test of its ability to detect smoke (although the quantity and quality of the smoke are not well controlled). The practical disadvantages are the risk of injury and the very real possibility of starting a fire, or at least causing localised damage from burning embers.

[4] Even the simplest domestic smoke detectors can have a limited self-test capability. For detectors based on scattering of light from smoke particles, the control unit pulses the light source at higher power than usual and monitors the signal from the photodiode detector. Complete absence of a signal indicates that the detector is not working.

The second task is a compromise. The whole protective system is tested with a hazard-free aerosol spray that mimics fire. Of course, it is exactly because of the limitations of the first task and the practical risks of the third that a market for these aerosol sprays exists.

It is not always possible to find a task that is a perfect compromise between only partial testing of the protective system and the risk of causing the multiple failure or other damage. If the whole protective system cannot be checked, or if the task exposes staff or equipment to excessive risk, refer the task to the responsible manager for urgent review.

Level of detail

- The level of detail depends on the skill and familiarity of the person carrying out the task. There needs to be enough detail to describe unambiguously and completely what has to be done.
- Be specific. Do not use words such as "Test" or "Check" in isolation. Say what is being checked or tested, and what the engineer needs to look for to get a definitive positive or negative result.
- Describe the equipment. Provide equipment and tag numbers wherever possible. Include the locations of gauges, switches and other relevant instrumentation.
- Define what constitutes functional failure, for example,

  "The relief valve must be fully seated and should not pass product at pressures below 180 kPa, and it must be fully open at 205 kPa"
- Unless it is absolutely certain that a multiple failure could not occur during the test, describe in detail how to put in place additional protection during the test.

  For the entire duration of the test, one engineer must be stationed at the local turbine speed meter with immediate access to the manual turbine shutdown control. This engineer must monitor the turbine speed continuously without distraction and use the local stop control to shut down the turbine if its speed exceeds 4640 RPM.

Before the test

The preamble should state clearly:

- Whether the protected system should be running normally, shut down, or in a specific "testing" state before the test starts
- How to ensure that operators and other staff are aware that testing is in progress
- Precautions that need to be taken before the testing is carried out, such as making safe equipment with high voltages, high pressures, high temperatures and other hazards
- Whether it is necessary to disable the protective system, and exactly which components of the system need to be disabled or overridden
- Any precautions that are necessary to ensure that the multiple failure does not occur as a result of the test if the protective system does not operate

Safety precautions

- All safety precautions should be written explicitly, clearly, and in order
- Include as much detail in the task as possible
- If you have to refer to external warnings and safety information, ensure that it will be easily available to the technician *when he or she is carrying out the task*

System conditions

- State whether the system should be running normally, shut down, or adjusted to a specific testing state
- Describe in detail what parts of the protective system must be disabled and what inputs need to be disconnected or overridden
- Write clear instructions for any interference with the normal operation of the protective system that is required during testing. This includes work such as disabling the protective system's inputs or outputs in some way
- The instructions to re-enable component parts, for reconnection of inputs and removal of overrides must match one-to-one with the pre-test instructions. Check them against the instructions that were followed to prepare the system for testing to ensure that every component that was disabled is re-enabled before the task is completed
- If its is possible, tell the maintainer how to obtain positive proof that the system is working normally when the task is complete

Supporting information

- Include any tools and spare parts in the task. Ensure that someone reading the task for the first time would arrive to do the work with all essential tools and materials.
- Try to include as much information in the task as possible.
- If you need to refer to other sources such as manufacturer's maintenance guides or operating manuals, ensure that the information will be readily available to someone carrying out the failure-finding task.

After the test

Write clear instructions for restoring the protected and protective systems when the test is complete and for ensuring that the protective system is fully operational.

Consider particularly the following.

- Describe in detail how to remove any inhibits or overrides that were put in place before the test
- Document how the protected system should be returned to normal operation, and how to make everyone aware that the system is no longer in "test" mode
- Describe as clearly as possible how the testing crew should double-check that the protective system is fully operational again, with no components disabled, disconnected or overridden
- State clearly how the test results should be signed off, and any requirements for a second signature from a witness or supervisor

Remedial instructions

- Remedial work needs to be carried out if failure-finding discovers a hidden failure.
- Simple remedial instructions may be included in the task itself, particularly if the work can be done out by the person who carried out the test. Including remedial information has the advantage that the work is immediately available, and the maintenance crew does not have to look up information in a manual or another task.
- Longer remedial instructions could make the task long and unnecessarily complex. Consider providing a clear reference to the external document or procedure, and ensure that an engineer in the field has easy access to the documentation.
- Where separate remedial work is needed, ensure that any procedures for notifying staff about the protective device failure and possibly shutting down the affected process are either within the failure-finding task description or referenced from it. This is particularly important if the protected system will be left in service without a functioning protective device, because operators have to make decisions based on performance data and they need to know when systems will not give them the protection or feedback that they expect.

Write steps in order

Always write steps in the order they will be carried out.

When a single step includes more than one action, try to make the actions flow from the beginning to the end of the sentence. For example, to open valves A and B in order, write

- ✓ Open valve A then open valve B

but not

- ✖ Open valve B after opening valve A

and also not

- ✖ Do not open valve B until valve A is open

Simple steps

Try to describe each step individually rather than combining several steps into a list. This isn't a work of literature; it is intended to be a clear statement of what needs to be done. Make it easy for the maintenance crew to remember how far they are through the list of steps, and do not overload them with too much information in any one step.

High Pressure Cut-Out Testing Procedure (too condensed)

1 With the plant running normally at 23-27 bar, adjust the pressure control valve to increase delivery pressure SLOWLY while monitoring pressure on the discharge pressure gauge. If the compressor does not trip at 30 bar, stop the compressor manually. Do not allow the pressure to exceed 35 bar.

2 If the compressor fails to trip, stop the compressor before 35 bar and notify the operators. Report failure and take the unit offline for repair.

3 Reset the pressure control valve.

High Pressure Cut-Out Testing Procedure (better)

IMPORTANT: The maximum safe system pressure is 40 bar.

IMPORTANT: Pressures above 40 bar may damage the system and could cause serious injury.

1 Ensure that the plant is running normally.

2 Check that the system pressure is within normal limits, between 23-27 bar.

3 Note or mark the position of the pressure control valve during normal operation

4 Adjust the pressure control valve and SLOWLY increase delivery pressure towards 30 bar.

5 If the compressor does not trip at 30 bar:

5a Stop the compressor manually before the pressure reaches 35 bar

5b Notify operators that the test has failed

5c Request that the unit is taken offline until the repair has been completed

5d Note test failure in the maintenance feedback form and schedule remedial work to repair the pressure switch

6 Set the pressure control valve to its pre-test position noted in step 3 above.

7.9 Key Points and Review

Before writing the task, consider the following issues.

- How can the task description be made as clear as possible?
- How often will a typical technician encounter a failed device? Will the failure be recognised? What additional information does the task need to provide if the failure is rare?
- For complex tasks, those with isolations and overrides, and protective devices with very high availability requirements, consider whether a supervisor or independent witness should validate the test
- Could existing tasks in a manufacturer's manual or on similar equipment conflict with this task description? If so, try to resolve any conflicts and make clear which version of the task is to be carried out
- Are some of the task steps invasive, or could the task stress or wear the protective system? Is it possible to design a task that is equally effective but less invasive?

When writing the task:

- As far as is practicable, ensure that the task checks the whole system
- Tailor the level of detail to the experience of those carrying out the task.
- Write clearly
- Unbundle long paragraphs into multiple steps where possible
- Write the task steps in order
- Include any safety precautions and preparatory work
- Include supporting information such as tools and spare parts required
- Clearly explain the required system state when the test is carried out (running, standby, "test" mode and so on)
- Any isolations, overrides or similar requirements should be written clearly
- Removal of isolations and overrides should match one-for-one to the pre-test instructions
- Detail any post-test actions required
- Describe what should be done if the protective device has failed. Include remedial instructions or refer to up-to-date remedial tasks in the ERP, MMS, manufacturer's manuals or other source

Section 2

Failure-Finding Task Intervals for Simple Systems

8 Availability

8.1 Introduction

In an earlier chapter we demonstrated that the availability of a protective system is not fixed by its design; in most cases it can be increased by testing the device to check that it is working, and fixing it if it is not. In general, more frequent checks lead to higher device availability, and less frequent checks deliver lower availability. This chapter develops the relationship between failure-finding task interval and the availability of a protective system.

⇒ *Remember that availability alone should not be used to set failure-finding intervals unless the target level has been derived from a robust, quantitative risk model.*

8.2 Availability and Failure Rate

The availability of a protective device is key to setting up a maintenance policy, but how do we calculate the availability of a real device?

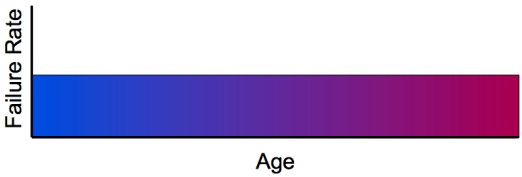
First, what factors contribute to the a protective device such as a low oil pressure sensor being unable to do its job? Among others, we may need to consider the following.

- Failure of the component during its working life
- Installation of a non-functioning component
- Disablement of the switch during a planned test, after which the switch was not reconnected
- Failure of external services such as power, networks and data buses

We have already seen that checking the device at fixed intervals enables us to influence the level of availability achieved, although these checks will have no direct effect on the other causes of unavailability listed above.

Common sense tells us that the availability of a device is directly related to its failure rate. If System A uses a switch whose mean time between failures is 100 years, and System B uses a switch whose mean time between failures is 10 years, then *if they are subject to the same maintenance policy*, we should expect that the switch in System A would demonstrate a far higher availability than that in System B. The less reliable a device is, the lower its availability is expected to be.

In this section we will deal with devices that fail at random. Random failure means that the chance of a failure occurring does not depend on the previous history of the device: not on the age of the device, the time of year, the phases of the moon, the number of starts or stops, or anything else. The chance of a working device failing on any chosen day is exactly the same as that of it failing on any other day.



Age-independent random failure: the probability of failure is independent of age

To make the calculation more concrete, the following example determines the availability of a high temperature trip system which has a 10% chance of failure in any one year[5].

The calculation begins when the device is newly installed or has just been checked. In this section we make the assumption that the trip is fully functional at time zero.

[5] It is a slightly odd fact that a failure rate of "10% per year" does not mean that exactly 90% of the devices are working at the end of the first year, or for that matter that 81% (0.9 x 0.9) are working at the end of year two. This is because the technical failure rate is applied continuously, not just at the end of the first year. The formula actually used is $R(t) = \exp(-\lambda t)$, where λ is the failure rate (0.1 or 10%) and t is measured in years. The actual proportion surviving to the end of the first year is 90.48%, dropping to 81.87% at the end of year 2. The difference is similar to that of a bank quoting 10% interest on your account; if interest is added once per year, you have $110 for every $100 in the account. If the bank adds interest every month, you get interest on the interest that has already been added, and at the end of the year you have $110.47 for every $100 invested. The more often interest is added, the more you get, until if the bank adds interest continuously, you have $110.52 for every $100 after one year.

If the trip is fully functional at time zero, what is the chance that it is still functional at the end of the first year? If the chance of failure is 10% in any year, the probability that it is still functional is

100%–10% =90%

What is the probability that the trip is still functional at the end of the second year, assuming that it is not checked or replaced during that period?

The chance that the device is still functional at the end of year 2 is given by:

(Chance that the device is working at the end of year 1)
x
(chance that the device does not fail during year 2)

In this case, the chance that the device is still functional after two years is

90% x 90% = 81%

Similarly, the chance that it is functional after three years is

(Probability working at end of year 2)
x
(chance of non-failure in year 3),

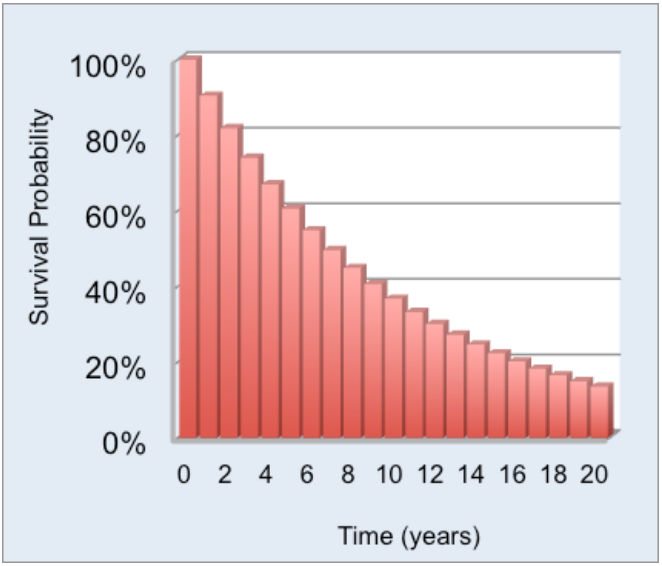
or

90% x 90% x 90% = 72.9%

The table below summarises the availability at the end of each of the first ten years.

End of year	Probability that trip functions
Start	100%
1	90%
2	81%
3	72.9%
4	65.6%
5	59.0%
6	53.1%
7	47.8%
8	43.0%
9	38.7%
10	34.9%

Note that the probability that the device is functional at the end of the third year is almost 73%, not 70%. Although there is a 10% chance of failure per year, this is a *conditional probability*: there is a 10% chance of failure of a device that is working at the start of the year. Since there is a 90% chance that the trip is working at the start of the year, the chance of failure during the second year is 90% x 10% = 9%.



Survival probability to the start of each year for a trip device that has a random failure pattern with a mean time between failures of 10 years

Sometimes it can be helpful to look at this in a different way. Imagine that there are 100 trips at time zero. If every trip is replaced when it fails, there are always 100 trips operational, and about 10 fail every year. However, if trips are not replaced when they fail, then about 90 remain after year 1; therefore the number of failures during year 2 is lower than during the first year because there are fewer working trips which can fail. As the number of working trips diminishes over time, the number of failures decreases as well. The number of failures goes down although the rate of failure per trip stays the same.

Since failure of the trip system is hidden, its availability is given by the probability that the trip is functional at the time of a demand. The availability is 90% after one year; after two years, 81%; after three, 73% and so on. Therefore the graph above shows the relationship between failure rate of the protective system and its availability.

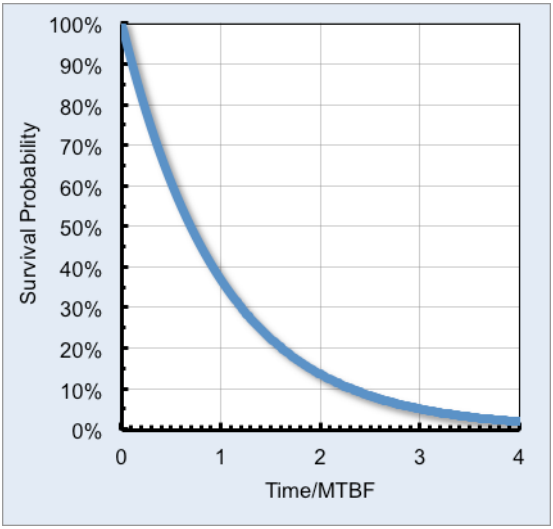
The graph shown above is an approximation. The true relationship between the failure rate of the trip and its availability is

$$A = e^{-t/M_{dev}}$$

Where

A	is the availability of the protective system
t	is the time since the device was installed or tested
M_{dev}	is the mean time between failures of the protective device
e	is the number 2.7182818..., the base of natural logarithms

The exact survival curve is shown below.



Curve showing the probability of survival to a given time divided by the device's mean time between failures

The survival curve shows the probability that a device functions after a specified time, expressed as a proportion of the device's mean time between failures. Its value is 1 (100%) initially and it decays towards zero, although in theory it never actually reaches it.

One feature of the survival curve is important to the derivation of most of the formulae used in this book. The first part of the curve, up to a time that is around 5% of the mean time between failures, is very nearly a straight line. At 5% of the mean time between failures, the difference between the straight line and the curve is just over 0.1%; at 10%, the difference is under 0.5%. Most of the formulae assume that the relationship between availability and time is a straight line. In order for the formulae to be valid, the following condition must apply.

The failure-finding interval must be less than about 5% of the mean time between failures of the protective device.

The equation of the first part of the survival curve is derived in section A.3. For times up to about 5% of the mean time between failures of the protective device, its availability is given by the following formula.

$$A = 1 - t/M_{dev}$$

8.3 Minimum Availability Calculations

Availability is the simplest criterion used to set failure-finding task intervals. A target availability of the protective system is chosen, then the failure-finding interval is calculated to achieve at least that level of availability.

If we want to achieve a given minimum availability, we already have all the tools needed to calculate the failure-finding interval for a real device. We choose the availability required and rearrange the formula

$$A = 1 - t/M_{dev}$$

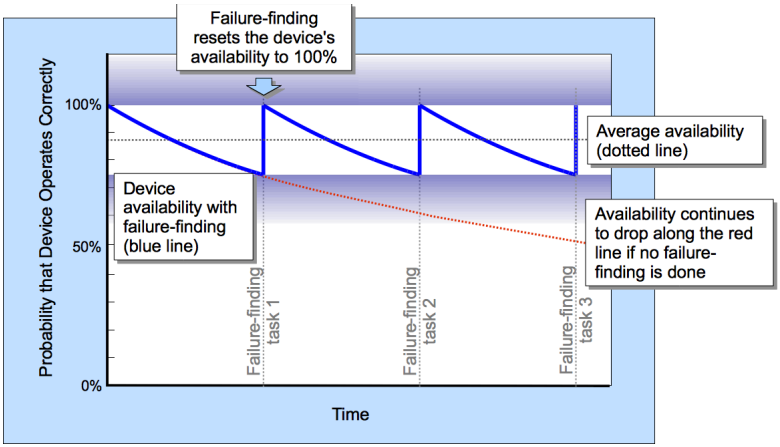
so that we can calculate the failure-finding interval from the device's mean time between failures and the required minimum availability.

$$T_{ff} = (1 - A)M_{dev}$$

Don't start celebrating just yet, though: this is not the calculation that is normally used. For reasons that will become more obvious later when we calculate the rate of multiple failures using availability and demand information, the target is usually the average rather then the minimum availability.

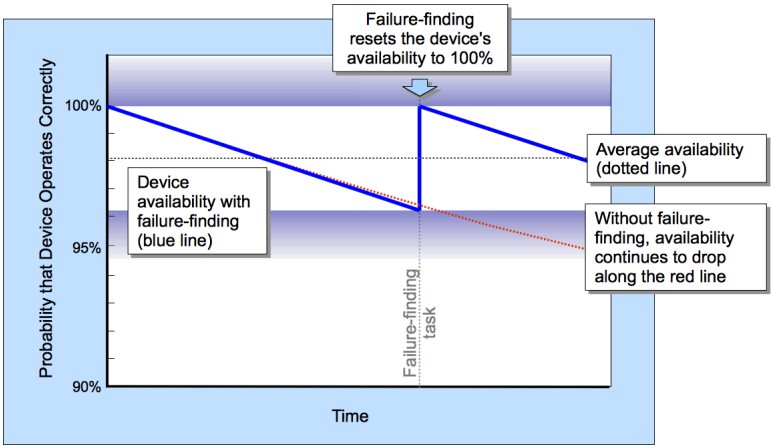
8.4 Availability-based Calculations: Average Availability

The formulae introduced in the previous section calculate the availability of a device at an instant in time. Availability starts at 100% and declines until we test the device, and if necessary repair it. The availability target used in this section is not the instantaneous availability, but the *average* availability of the device assessed over a period of time. The graph below shows how the device's availability changes over the course of time between failure-finding tasks. The dotted line marks the device's average availability.



Availability of an ideal protective device with failure-finding (blue) and without failure-finding (red)

A mathematical derivation of the average availability is shown in section A.4, but if we assume that the graph is a straight line, it is easy to derive the average value visually simply by looking at the graph of availability between two failure-finding tasks.



Availability is approximated by a straight line. Note that the left hand axis has been expanded.

The availability of the device immediately after failure-finding (and, if necessary, a repair) is assumed to be 100%. Provided that its availability stays above about 95%, its availability drops approximately according to the equation

$$A = 1 - t/M_{dev}$$

so that when it reaches the failure-finding task interval T_{ff}, its availability is

$$A = 1 - T_{ff}/M_{dev}$$

The average availability over the period (shown by the dotted line) is

$$\bar{A} = 1 - \frac{T_{ff}}{2M_{dev}}$$

Therefore if the target average availability of the protective device is A, then the failure-finding task interval needed to achieve it is

$$T_{ff} = 2M_{dev}(1 - \bar{A})$$

8.5 General conditions

The following conditions apply to the calculations in this chapter.

1 The calculation applies only to one failure mode of a single, simple protective device. It does not apply to multiple failure modes, or to protective systems that consist of several simple devices, such as a pair of pressure relief valves, redundant backup generators, or a 2-of-3 voting system.

2 Failures of the protective device must be random; there must be no relationship between the chance that a device has failed and its age or the time since it was last maintained.

3 The calculation assumes that the failure-finding interval is small compared with the device's mean time between failures. Typically the failure-finding interval should be less than about 5% of the mean time between failures of the protective device.

4 Unavailability of the protective device due to scheduled or unscheduled maintenance is not included in the calculations, and may need to be taken into account in calculating the overall unavailability of the device.

Availability is the simplest criterion that can be used to calculate failure-finding intervals; but this leads immediately to the question of how to determine the availability required. That issue is dealt with in the following chapter.

8.6 Examples

Fan vane switch

A fan is used to dilute boiler flue gases by mixing them with air before they are dispersed at low level. Local regulations state that the CO_2 content of the discharged gas must be below 1%.

If the fan fails for some reason, or if the ducting is blocked, a vane switch shuts down the boiler to prevent discharges with a CO_2 content above the allowed limit.

The mean time between failures of the vane switch is estimated to be 10 years. The required average availability of the shutdown switch is 99.7%.

How often should the vane switch be tested?

The table below summarises the relevant numbers.

Term	Description	Value
M_{dev}	Mean time between failures of the vane switch	10 years
$\bar{A}$	Required average availability of the vane switch	99.7%
T_{ff}	Vane switch failure-finding interval	To be calculated

The failure-finding interval needed to achieve 99.7% availability is

$$T_{ff} = 2M_{dev}(1 - \bar{A})$$

or

$$T_{ff} = 2 \times 10 \times (1 - 0.997) \; = \; 0.06 \text{ years}$$

The required testing interval is 0.06 years, or about every three weeks.

Oil Pressure Switch

A diesel generator's engine contains a low oil pressure warning switch. If oil pressure drops below 1.5 bar, a red light illuminates on the control panel and the operator is expected to take action to prevent damage to the engine.

The manufacturer's data show that the switch's mean time between failures is 200 years. The required average availability is 99%.

How often should the switch be tested?

The table below summarises the relevant numbers.

Term	Description	Value
M_{dev}	Mean time between failures of the low oil pressure switch	200 years
$\bar{A}$	Required average availability of the low oil pressure switch	99%
T_{ff}	Oil pressure switch failure-finding interval	To be calculated

The failure-finding interval needed to achieve 99% availability is

$$T_{ff} = 2M_{dev}(1 - \bar{A})$$

or

$$T_{ff} = 2 \times 200 \times (1 - 0.99) = 4 \text{ years}$$

So the low oil pressure switch should be checked every four years.

⇒ *This calculation does not include the other components of the alarm system that might fail. Later chapters deal with more complex systems that include more than one component.*

Gas detector

A compartment in an offshore production facility contains a combustible gas detector that should raise an alarm if the gas concentration rises above 10% of the lower explosive limit (LEL).

Records show that the mean time between failures of the detector is about 20000 hours. A quantitative risk assessment implies that the required availability is 99.99%.

How often should the detector be tested?

The table below summarises the relevant numbers.

Term	Description	Value
M_{dev}	Mean time between failures of the gas detector	20000 hours
$\bar{A}$	Required average availability of the gas detector	99.99%
T_{ff}	Gas detector failure-finding interval	To be calculated

The failure-finding interval needed to achieve 99.99% availability is

$$T_{ff} = 2 \times \frac{20000}{8760} \times (1 - 0.9999) = 0.00046 \text{ years}$$

The detector should be checked every four hours to achieve 99.99% availability. This testing interval is unlikely to be acceptable: this is an indication that the device is incapable of delivering the availability required, and the system should be redesigned in some way to provide better reliability.

8.7 Time to Repair

The availability calculations used in this chapter do not take account of the time taken to repair the protective device. The primary reason is this: if the failure-finding task discovers that the device has failed, the operators will normally take measures to reduce the risk of a multiple failure until the repair is complete. If a pressure relief valve is found to be stuck closed, the associated process will probably be shut down until it is repaired; if a high process temperature alarm is found to be inactive, the operators might dedicate someone to watch local gauges until the alarm is available again.

The time to repair (expressed as the MTTR, or mean time to repair) has to be mentioned here because it does form part of some methodologies for managing hazardous systems, including SIL (Safety Integrity Levels). The modifications needed to take account of repair time are discussed later in the book.

8.8 Key Points and Review

Because failure of a protective device is hidden, we cannot be certain whether it will function correctly when a demand occurs.

The availability of a protective device is the probability that it will work at a specific time.

The average availability of a protective device depends on its reliability, as measured by its mean time between failures, and how frequently it is tested.

Under a number of assumptions, there is a direct relationship between the availability of a protective device, its mean time between failures and how frequently it is tested. Therefore it is possible to calculate how often a protective system needs to be tested to achieve the desired level of availability.

Although availability is the simplest criterion for determining failure-finding task intervals, its use is only justified if the availability chosen can be robustly defended.

9 Risk

9.1 Introduction

In the previous chapter, failure-finding intervals were set by determining the test frequency that results in the required average availability of the protective device .

It was emphasized that the availability requirement should be derived from a rigorous, robust model. There is no problem if the system has been subjected to a quantitative method such as Fault Tree Analysis (FTA), but for most industrial systems the required availability level simply does not exist. This chapter demonstrates how to derive the availability needed from two numbers:

- The mean time between demands on the protective system
- The minimum tolerable mean time between multiple failures

9.2 Getting to Availability

Although the ultimate objective of managing hidden failures is to reduce or eliminate the risk of multiple failures, the previous chapter's availability calculation does nothing to connect the failure-finding interval to the risk of a multiple failure. Since the whole point of maintaining the protective system is to reduce the chance of a multiple failure happening, it makes sense that the task interval should be based on how often we are willing to allow the multiple failure to occur.

The relationship between availability and multiple failure rate is very simple, but it involves one more parameter, as demonstrated by this example.

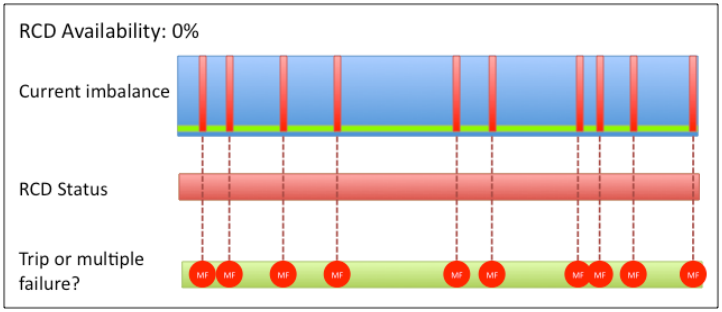
> A small 230 volt electrical installation is protected by a residual current detector (RCD) which is intended to cut off the power if current flows to earth, perhaps because of a fault or because someone has accidentally touched the live wire. The device works by comparing the current in the live wire with the return current, and tripping if the imbalance is more than a few milliamps.
>
> Records show that the RCD is tripped in normal use (not during testing) about once per year.

The multiple failure that the device protects against is that there is a fault and the power is not cut, leading to equipment damage, injury, or even death.

What is the relationship between the availability of the RCD and the rate of multiple failures?

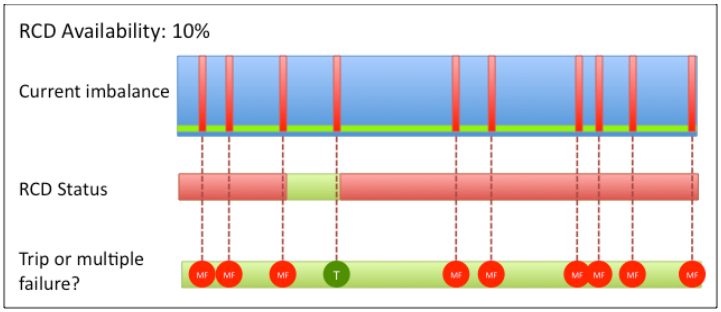
We know that the availability of the RCD depends on its reliability and on how often it is tested. Since we already know how to work out its average availability, we are going to treat the availability as a variable and work out the multiple failure rate.

First suppose that the RCD is never tested. Assuming that it worked when it was installed, its availability decays away over time and eventually should be close to zero. Ignoring (just for convenience) the very early part of its life, the RCD will always be in a failed state. A drawing of a typical history might look like this; demands on the RCD are assumed to occur at random.



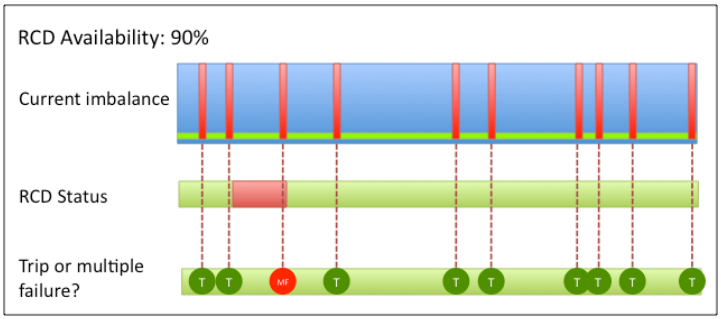
How often does the multiple failure occur? The simple answer is: every time that there is a demand on the RCD, because the device availability is zero. If a demand occurs on average once a year, then the multiple failure also occurs once a year.

Now suppose that the RCD availability is improved a little, so that on average it works 10% of the time. How could the same history look?



Because failures of the RCD and demands occur at random, it is possible that there could be 10 multiple failures, 9 or any other number. However, on average one in ten demands would lead to a trip, and nine out of ten demands would end in a multiple failure.

If we now increase the RCD's availability to 90%, perhaps by implementing some form of regular failure-finding task, then only one out of 10 demands (on average) would result in a multiple failure.



There is a simple relationship between the mean time between demands on the system, the average protective device availability, and the mean time between multiple failures:

$$M_{mf} = \frac{M_{dem}}{(1 - \bar{A})}$$

If the protective device availability is 100%, the mean time between multiple failures is infinite; in other words, the multiple failure never happens.

9.3 Risk-based Calculations

Now that we have a link between demand rate, the mean time between multiple failures and average device availability, we are in a position to work out the mean time between multiple failures that would be achieved if we checked a device at a specific failure-finding interval.

In the previous chapter we found that the average availability of a simple, single protective device that fails at random is

$$\overline{A} = 1 - \frac{T_{ff}}{2M_{dev}}$$

From the equation above, the mean time between multiple failures is

$$M_{mf} = \frac{M_{dem}}{(1 - \overline{A})}$$

Putting these two formulae together, the multiple failure rate for a given demand rate and failure-finding interval is

$$M_{mf} = \frac{2M_{dem}M_{dev}}{T_{ff}}$$

In chapter 5 we found that the objective of failure-finding for multiple failures that have safety or environmental consequences is to reduce the rate of multiple failures to a tolerable level. By rearranging the formula, the failure-finding interval needed to achieve a given mean time between multiple failures is

$$T_{ff} = \frac{2M_{dem}M_{dev}}{M_{mf}}$$

To summarise, the failure-finding interval for a risk-based system is determined by the following factors.

Symbol	Description
M_{dev}	Mean time between failures of each protective device
M_{dem}	Mean time between demands on the protective system
M_{mf}	The lowest tolerable mean time between multiple failures

Remember that use of this formula is subject to a number of caveats including the following.

- There is one protective device
- The protective device fails at random
- The protective device is guaranteed to be working immediately after installation
- The failure-finding task is always effective: the task discovers 100% of non-working devices and any devices that have to be repaired are fully functional immediately after the task has been carried out
- Demands on the protective device occur at random
- The failure-finding interval is less than about 5% of the protective device's mean time between failures
- The failure-finding interval is much less than the mean time between demands on the device
- The time taken to repair the protective device is insignificant, or other measures will be taken to prevent multiple failures if a scheduled task discovers a failed protective device

Most of these assumptions will be revisited and relaxed in later chapters, but bear in mind that you need to check your system, data and failure-finding interval carefully to make certain that these conditions are not broken.

9.4 Demand Rate

The failure-finding calculation is now based on the minimum tolerable mean time between multiple failures, but at the expense of adding a new item to data: the mean time between demands on the protective system.

The demand rate is how often the protective system has to operate because of abnormal conditions. If a pressure relief system on a boiler that has been installed for twenty years has been called on to relieve overpressure four times in that period (and if the demands occur at random), then the mean time between demands is five years.

Remember these points when working out the mean time between demands.

- Count any occasion on which the device has had to operate because of genuinely abnormal conditions that could have caused a multiple failure. The key question is, "How many times (or how often) has this device operated because of abnormal conditions." The question is *not*, "How many times has it failed to prevent a multiple failure?"

- Do not count deliberate demands on the system due to routine tests and maintenance
- The calculations assume that demands occur at random. If they are non-random (perhaps they tend to occur just after or just before major maintenance), then the formulae may not give correct answers.

You will probably need to talk to maintainers and operators to find the data that you need. To maximise the chances of obtaining the correct information, it is a good idea to phrase the question in terms that relate to the system under analysis. So, rather than asking

"How many times has a demand occurred on the fire alarm system?"

ask

"How many times has there been a fire in this building?"

If the protective system has been operational over a long period of time, try to be aware of any changes in operating context, such as increases in production rates or the introduction of different technologies, which might influence the demand rate. The demand rate used to set failure-finding intervals should be the anticipated future rate, which may be different from earlier experience.

Some system demand rates are easy to estimate. Domestic and industrial residual current detectors (RCDs), which protect users of mains power supplies, trip sufficiently frequently in normal use that most organisations can estimate the demand rate accurately. Most experienced drivers have at some time encountered a situation where their anti-lock brake system operated, and so they would be able to estimate how often they make a demand on the system.

Even where demands on one device are infrequent, there may be enough of them operating at any time to enable a realistic demand rate to be calculated: although an individual office building's fire alarm may detect be presented with a real fire only every few decades or so, there are plenty of aggregated statistics for different industry sectors, cities, regions and whole countries.

Even so, some demands are infrequent and just about unique to a specific organisation or process. The issue of finding demand rates in a variety of circumstances is discussed in more detail in chapter **Error! Reference source not found.**.

9.5 Multiple Failure Rate

The objective of the failure-finding task is to reduce the rate of multiple failures to a tolerable level, or equivalently to deliver a minimum tolerable mean time between multiple failures.

The failure-finding interval depends directly on this number, so it is vitally important that the rate of multiple failures is acceptable for all those who are likely to be affected by the hazard, including duty holders, senior management, company staff and members of the general public. Bear in mind that a single failure mode may represent only a small part of the organisation's risk, and that all significant failure modes need to be included in a complete risk management plan.

There is further discussion of tolerable failure rates, who should be involved in setting risk targets, and methods for determining individual multiple failure rates in chapters 6 and **Error! Reference source not found.**.

9.6 Examples

Oil Pipeline Low Pressure

A small lubricant pipeline runs close to an environmentally-sensitive area. A pressure switch is intended to shut down the oil pump if a significant leak occurs.

The mean time between failures of the pressure switch in this application is about two million hours. The low pressure switch has never been activated except during tests, but estimates suggest that it could be called on to operate about once every 10 years.

The minimum tolerated mean time between multiple failures (an undetected pipeline leak) is 10 000 years.

The required failure-finding interval is

$$T_{ff} = \frac{2M_{dem}M_{dev}}{M_{mf}}$$

or

$$T_{ff} = 2 \times 10 \times \frac{2000000}{8760} \times \frac{1}{10000} \text{ years} = 0.46 \text{ years}$$

The failure-finding task would have to be carried out every six months.

Standby Generator

A remote medical facility is subject to infrequent power outages that can last for at least several hours. When mains power is not available it relies on a single diesel generator which starts automatically when mains power is lost.

The group reviewing the standby power maintenance policy has decided to treat the generator, its engine, and the cut-in system as a single entity. The overall mean time between failures of similar systems at other installations is about 2 years. The mean time between demands on the system is about one year.

Although higher reliability would be desirable, the review group decided reluctantly that the chance of the generator being unable to produce power when required should be less than 1 in 1000 years.

The required failure-finding interval is

$$T_{ff} = \frac{2 \times 1 \times 2}{1000} \text{ years } = 1.5 \text{ days}$$

The failure-finding task would have to be carried out every day.

This interval is might not be acceptable for several reasons:

- It would probably be impractical to carry out the task at this interval
- Performing the task so frequently would place significant stress on the engine and so would contribute to wear and could result in lower reliability
- It indicates a gap between the desired reliability and what the equipment is capable of delivering

The most likely outcome is that the system would be redesigned to make it more reliable, perhaps by providing a second standby generator.

9.7 Time to Repair

The calculation method used in this chapter does not take into account device unavailability that arises during repair of the protective system. In most circumstances this is a reasonable assumption because maintainers and operators take care to reduce or eliminate the risk of a multiple failure during device repair. In most cases the affected system would be shut down, but sometimes the system operators could use alternative protection or closer system monitoring.

Time to repair should be included in the protective system downtime if no additional precautions are taken during the repair period. The modified formulae are derived later chapters.

9.8 Key Points and Review

Availability is not a useful criterion for determining failure-finding intervals unless it is supported by a robust model.

The required device availability can be calculated from two numbers:

- The demand rate on the protective device
- The minimum tolerated mean time between multiple failures

These lead to a simple relationship between the device reliability, demand rate and tolerated mean time between multiple failures.

10 Economic

10.1 Introduction

The previous chapter emphasized the importance of setting the right tolerable mean time between failures to drive the calculation of failure-finding intervals. Now consider this example.

> A pump provides water flow in a closed loop cooling system. If the pump breaks down, a standby pump starts automatically to take over the duty. If the standby pump failed to cut in when it was needed, the process would shut down because of low coolant flow within a few minutes. The time taken to repair one of the pumps would be about two hours, and production worth about $3000 would be lost.
>
> The mean time between failures of the duty pump is about two years, and the MTBF of the standby pump is about 5 years.
>
> How often should the standby pump be tested?

This information provides the mean time between failures of the protective device (5 years) and the demand rate (how often the duty pump fails: 2 year). The multiple failure effects are also known ($3000 loss). But how often is the organisation willing to tolerate a loss of $3000? Every year? Once a decade? How is it possible to define a tolerable level of risk without considering every other similar failure in the organisation?

10.2 Economic Calculations

The key to this problem is that the results of the multiple failure are only economic. The effects could be a minor hiccough, or they could represent weeks of production, but only money is involved. There are no safety or environmental effects.

Because the multiple failure effects are purely economic, we are free to strike a balance between two costs.

Cost of multiple failures	The risked cost per year due to multiple failures. The more often the failure-finding task is carried out, the lower these costs will be.
Cost of failure-finding	The cost per year of carrying out the failure-finding task, including labour, materials, and any downtime required to perform the task.

It is worth remarking at this point that these are two different types of cost. If the failure-finding task costs $50 every time it is carried out, and it needs to be done once per month, then the organisation will definitely spend $600 per year testing the standby pump. Multiple failures represent a *risked* cost. If the mean time between multiple failures is 100 years and each failure costs $3000, then the average cost of multiple failures is $3000/100 = $30 per year. This is very different from the cost of failure-finding because the organisation will not actually spend $30 per year. In most years it will spend nothing at all on multiple failures. In some years it will spend $3000 because of a single multiple failure; sometimes it might even face two or three multiple failures in a year. So while the cost of carrying out the task is a real, definite, fixed cost, the cost of multiple failures is a risked cost. This should be considered very carefully if the economic consequences of the multiple failure are severe.

10.3 Costs

The relationship between the failure-finding interval and the cost of failure-finding is simple; the cost per unit time is

$$\frac{C_{ff}}{T_{ff}}$$

where C_{ff} is the cost of carrying out a single failure-finding task. The risked cost of multiple failures is

$$\frac{C_{mf}}{M_{mf}}$$

where C_{mf} is the cost of a single multiple failure.

If the failure-finding interval is very short, the yearly cost of testing the standby pump is high but the cost of multiple failures is very small because the device availability is high. On the other hand if the failure-finding interval is long, the cost of tests is much lower but the cost of multiple failures is high. Somewhere between the two extremes is a point where the total cost to the business is at its lowest: this represents the optimum failure-finding interval.

The formulae that were developed in the last two chapters are all that are needed to work out the total cost of a specific task interval (failure-finding cost plus multiple failure cost). The mean time between multiple failures is

$$M_{mf} = \frac{2M_{dem}M_{dev}}{T_{ff}}$$

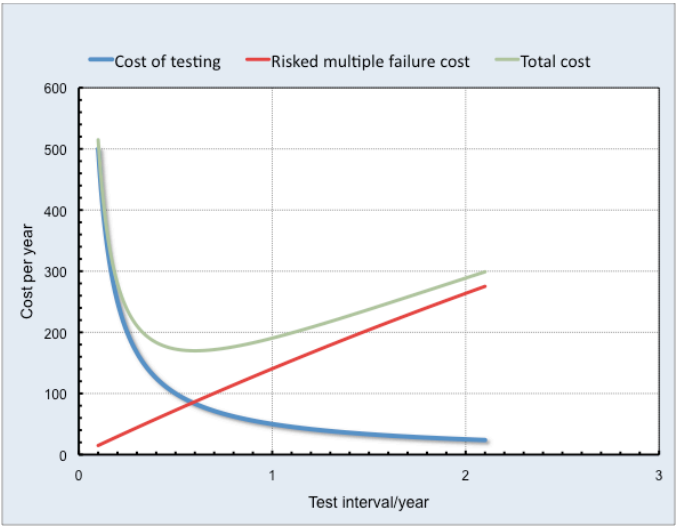
and the rate of expenditure on multiple failures per unit time is therefore

$$\frac{C_{mf} T_{ff}}{2\, M_{dem} M_{dev}}$$

The table and graph below show the cost of testing, the cost of multiple failures and the total cost per year for the duty/standby pump example above.

Test interval (years)	Average standby availability	M_{mf} (years)	Cost of testing /year	Multiple failure cost /year	Total cost /year
0.1	99.01%	201.34	$500	$15	$515
0.2	98.03%	101.34	$250	$30	$280
0.3	97.06%	68.01	$167	$44	$211
0.4	96.10%	51.34	$125	$58	$183
0.5	95.16%	41.34	$100	$73	$173
0.6	94.23%	34.68	$83	$87	$170
0.7	93.32%	29.92	$71	$100	$172
0.8	92.41%	26.35	$63	$114	$176
0.9	91.52%	23.58	$56	$127	$183
1	90.63%	21.36	$50	$140	$190
1.1	89.76%	19.54	$45	$154	$199
1.2	88.91%	18.03	$42	$166	$208
1.3	88.06%	16.75	$38	$179	$218
1.4	87.22%	15.65	$36	$192	$227
1.5	86.39%	14.70	$33	$204	$237
1.6	85.58%	13.87	$31	$216	$248
1.7	84.77%	13.13	$29	$228	$258
1.8	83.98%	12.48	$28	$240	$268
1.9	83.19%	11.90	$26	$252	$278
2	82.42%	11.38	$25	$264	$289
2.1	81.66%	10.90	$24	$275	$299
2.2	80.90%	10.47	$23	$286	$309
2.3	80.16%	10.08	$22	$298	$319
2.4	79.42%	9.72	$21	$309	$330
2.5	78.69%	9.39	$20	$320	$340

The minimum cost is at an interval about 0.6 years, or 7 months.



10.4 Optimisation

In the section above we found a point where the failure-finding interval minimises the overall cost to the business by plotting the total cost against the task interval. It is also possible to determine the best task interval by finding the minimum of the total cost formula.

The total cost consists of two components: the cost of performing the failure-finding task and the risked cost of multiple failures.

The cost of carrying out the failure-finding task is

$$\frac{C_{ff}}{T_{ff}}$$

and the risked cost of multiple failures is

$$\frac{C_{mf}}{M_{mf}} = \frac{C_{mf}(1-\bar{A})}{M_{dem}} = \frac{C_{mf}T_{ff}}{2\,M_{dem}M_{dev}}$$

so the total cost per unit time is

$$C_{total} = \frac{C_{ff}}{T_{ff}} + \frac{C_{mf}T_{ff}}{2\,M_{dev}M_{dem}}$$

The total cost can be minimised by using calculus; the failure-finding interval which minimises the total cost C_{total} is

$$T_{ff} = \sqrt{\frac{2\,C_{ff}M_{dev}M_{dem}}{C_{mf}}}$$

Where the following symbols are used.

Symbol	Description
M_{dev}	Mean time between failures of each protective device
M_{dem}	Mean time between demands on the protective system
C_{ff}	The cost of carrying out a single failure-finding task
C_{mf}	The cost of a single multiple failure

Substituting the values for the two pump example, and using $50 for the cost of a failure-finding task:

$$T_{ff} = \sqrt{\frac{2 \times 50 \times 5 \times 2}{3000}} = 0.577 \text{ years}$$

The task would probably be carried out every six months.

10.5 Assumptions

The usual assumptions apply to this calculation.

- There is one protective device
- The protective device fails at random
- The protective device is guaranteed to be working immediately after installation
- The failure-finding task is always effective: the task discovers 100% of non-working devices and any devices that have to be repaired are fully functional immediately after the task has been carried out
- Demands on the protective device occur at random
- The failure-finding interval is less than about 5% of the protective device's mean time between failures

- The failure-finding interval is much less than the mean time between demands on the device
- The time taken to repair the protective device is insignificant, or other measures will be taken to prevent multiple failures if a scheduled task discovers a failed protective device

10.6 Examples

Storage Tank Low Level Alarm

An ultrasonic system is used to monitor the level of solvent in a large storage tank in a polymer plant. It should raise an alarm in the control room if the tank level rises above 1.5m from the top of the tank or if it drops below 0.5m from the bottom. If an alarm sounds, the operators are usually able to adjust downstream usage or the supply rate to avoid a trip; in the worst case, they have time to initiate a "soft" shutdown. If the level continues to drop below 0.5m and no action is taken, a low level trip cuts off the delivery pump at 0.2m and the downstream process is shut down immediately.

Restarting the process after an unexpected trip takes several hours, and the total cost including lost production is likely to be about $10000. The low level alarm can be tested during normal operation because the technician can monitor a local level gauge to ensure that a trip does not occur; the total cost of carrying out the test is about $25.

Only four low level alarms have occurred in the past ten years. The manufacturer states that the alarm system's mean time between failures is about 50 years in this operating context.

How often should the alarm system be tested?

The following table summarises the information given in the problem.

Term	What it means	Value
M_{dev}	Mean time between failures of the alarm system (how often, on average, it would be unable to generate an alarm if a low tank level were to occur)	50 years
M_{dem}	How often on average we call on the alarm because of a low tank level	10/4 years
C_{ff}	How much it costs to check once that the alarm is operational	$25
C_{mf}	How much it would cost if the multiple failure occurred; *i.e.* that there was a low level but the alarm failed to sound	$10000
T_{ff}	How often we will test the low level alarm	

Using the economic failure-finding formula

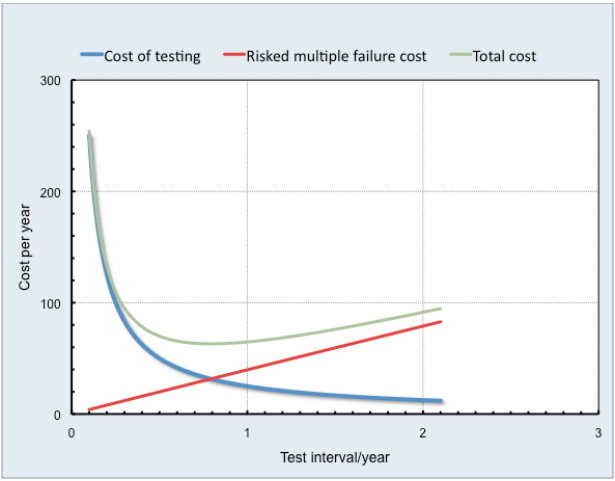
$$T_{ff} = \sqrt{\frac{2\,C_{ff}M_{dev}M_{dem}}{C_{mf}}}$$

the failure-finding interval required is

$$T_{ff} = \sqrt{2 \times 50 \times \frac{10}{4} \times 25 \times \frac{1}{10000}} = 0.79 \text{ years}$$

The task would probably be carried out every third quarter (9 months) if the maintenance scheduling system is sufficiently flexible.

What can be done if it is impractical to carry out the task at an interval of nine months? Fortunately the area of minimum cost is usually fairly broad, so there is reasonable scope for stretching or reducing the maintenance interval. The graph of total cost against failure-finding interval is shown below (in green).



The lowest total cost (at an interval of 0.79 years) is $63.08 per year. If it is impractical to carry out the task every nine months, then reducing the interval to six months would increase the total cost to $69.93 per year; stretching the interval to one year would increase the cost to $64.73, only 2.6% higher than the optimum.

Finally we need to check that the assumptions made in deriving the test interval are valid.

First, the failure-finding interval should be less than about 5% of the protective device's MTBF. With a failure-finding interval of 0.79 years, the interval is 0.79/50 = 1.6% of the alarm's mean time between failures.

Second, the test interval should be very much less than the mean time between demands. This is more marginal (0.79 versus 2.5 years), but it is unlikely to have a significant effect on the validity of the result (see chapter A, Mathematical Annex).

Gas Compressor Lubrication Oil

An auxiliary lube oil system provides lubrication for an inert gas compressor. If the lube oil system were to fail, the compressor's bearings would be seriously damaged before other sensors tripped the drive motor. The multiple failure is not expected to have any safety or environmental effects, but the total cost of replacing the bearings and production losses is high: it is estimated to be about $50,000.

The lube oil system has not failed since the compressor was installed two years ago, but experience with similar systems suggests a mean time between failures of about 12 years. The low pressure trip system has an MTBF of about 450000 hours in this application.

The failure-finding task is easy to carry out because the operators simply need to verify that a trip signal is sent when the system is on standby. The cost of carrying out this task is less than $10.

How often should the low pressure trip be tested?

The following table summarises the information given in the problem.

Term	What it means	Value
M_{dev}	Mean time between failures of the low pressure trip system (how often, on average, it would be unable to trip the motor if the lube oil pressure dropped)	450000 hours
M_{dem}	How often on average we call on the trip because of low lube oil pressure	12 years
C_{ff}	How much it costs to check once that the trip is operational	$10
C_{mf}	How much it would cost if the multiple failure occurred; i.e. that there was low oil pressure but the trip did not stop the drive motor	$50000
T_{ff}	How often we will test the low pressure trip system	

Using the economic failure-finding formula

$$T_{ff} = \sqrt{2 \times \frac{450000}{8760} \times 12 \times 10 \times \frac{1}{50000}} = 0.5 \text{ years}$$

The proposed failure-finding interval is about 1% of the pressure trip MTBF and much less than the demand rate.

If the cost of the multiple failure is fairly high, it is worth checking that the expected mean time between multiple failures is tolerable. This is particularly true if the failure could damage the organisation's reputation, perhaps by delaying product delivery to customers.

To calculate the mean time between multiple failures, we use the formula from the previous chapter:

$$M_{mf} = \frac{2M_{dem}M_{dev}}{T_{ff}}$$

For this example:

$$M_{mf} = \frac{2M_{dev}M_{dem}}{T_{ff}} = 2 \times \frac{450000}{8760} \times 12 \times \frac{1}{0.5} = 2466 \text{ years}$$

10.7 Key Points and Review

If the effects of the multiple failure are purely economic, it is possible to calculate an optimum failure-finding interval which balances the cost of carrying out the test against the risked cost of multiple failures:

$$T_{ff} = \sqrt{\frac{2\ C_{ff}M_{dev}M_{dem}}{C_{mf}}}$$

The formula may only be used if the multiple failure has no safety or environmental consequences.

The cost of carrying out the task is a real cost; the multiple failure cost is a *risked* cost that is equal to the cost of a single multiple failure divided by the mean time between multiple failures. There is always a risk that the multiple failure will occur. If it does, the organisation will bear the full cost of failure. If the financial consequences of the multiple failure are severe, ensure that the mean time between multiple failures is tolerable.

11 Parallel Systems

11.1 Introduction

Previous chapters have assumed that the protective device is a simple, single system whose reliability is characterised by a single mean time between failure figure.

It often happens that a single protective device cannot achieve a tolerable risk of a multiple failure, as in the hypothetical example below.

> A steam boiler is protected by a single pressure relief valve whose function is to vent excess steam if the boiler pressure exceeds 10 bar (1 MPa). If the boiler pressure exceeded this value and the pressure relief valve failed to operate, the boiler could explode and seriously injure or kill personnel in the vicinity.
>
> The best available information on the relief valve states that its mean time between failures in this operational context should be about 50 years. As far as anyone knows, the relief valve on this boiler has never had to lift during its two year life, but in the plant as a whole, similar valves are expected to experience about one demand every twenty years.
>
> A detailed analysis has led the management team to apply a standard of no more than one multiple failure every 4000000 years.
>
> The review team calculates that the required failure-finding interval for the relief valve is
>
> $$T_{ff} = \frac{2\, M_{dev}\, M_{dem}}{M_{mf}} = \frac{2 \times 50 \times 20}{4000000} \text{ years}$$
>
> The result is 0.0005 years, or a little over 4 hours.

Checking a relief valve every four hours is probably not a feasible maintenance task. The most obvious options are to find a more reliable relief valve or to increase the mean time between demands on the system, perhaps by improving the pressure control system. Even with these changes it seems unlikely that either of these could deliver the improvement that would be needed for failure-finding to take place at a reasonable interval[6].

[6] Although failure-finding is unlikely to be feasible in the example considered here, short failure-finding task intervals can sometimes be practical by using automatic or online test equipment.

If it is impossible to achieve the required level of risk with the existing equipment, what is to be done? A common sense redesign would replace the single relief valve with two (or possibly more) valves. Then if one valve fails to operate, the second is likely still to be operational. This is exactly how most boiler pressure relief systems are designed.

If we assume a yearly failure-finding interval for the relief valve system, what level of availability is achieved?

For a single valve whose mean time between failures is 50 years, a failure-finding task interval of a year leads to an average availability of

$$\bar{A} = 1 - T_{ff}/2M_{dev} = 1 - \frac{1}{2 \times 50} = 0.99 = 99\%$$

Simple reliability theory suggests that if a single system has an availability of 99% (unavailability of 1%), then the chance that *both* systems have failed when boiler overpressure occurs is

$$U = 1\% \times 1\% = 0.01\%$$

So there is approximately a one in 10000 chance that both systems have failed when required.

For reasons discussed below, the figure of one in 10000 is incorrect, but the principle is sound: increasing the number of parallel protective devices drastically reduces unavailability, provided that each device on its own is capable of preventing the multiple failure. In practice almost all real world boiler protection systems employ more than one relief valve in order to achieve an acceptable level of risk.

11.2 Availability

The following analysis assumes that all of the parallel protective devices are tested at the same time and that all the devices are identical. If this is the case, then the multiple failure can only happen if all the devices have failed. Assuming that all the devices are fully functional immediately after the failure-finding task, the chance that any one of them has failed at time t is (from Chapter 8)

$$U = 1 - e^{-t/M_{dev}}$$

If there are n identical, parallel devices, the chance that they have all failed at time t after the check is

$$U = \left(1 - e^{-t/M_{dev}}\right)^n$$

Section A.4 shows that the average unavailability over the failure-finding period T_{ff} is

$$\bar{U} = \frac{1}{n+1}\left(\frac{T_{ff}}{M_{dev}}\right)^n$$

and the failure-finding interval required to achieve a predetermined average availability A of the protective system is

$$T_{ff} = M_{dev}[(n+1)(1-\bar{A})]^{1/n}$$

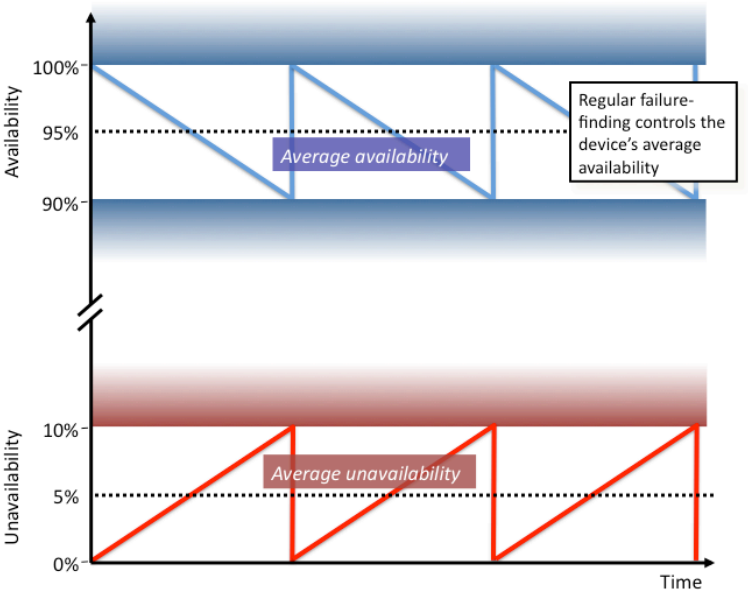
Where "To the power of 1/n" means "Take the nth root of the quantity inside the brackets."

For the example boiler discussed in section 11.1 and assuming a failure-finding interval of one year, the theoretical average availability achieved rises quickly as the number of parallel relief valves is increased.

Parallel relief valves	Availability achieved (task interval 1 year)
1	99%
2	99.987%
3	99.9998%
4	99.999997%

Why is the unavailability of n devices not just the unavailability of one device multiplied by itself n times, as would be expected from reliability theory?

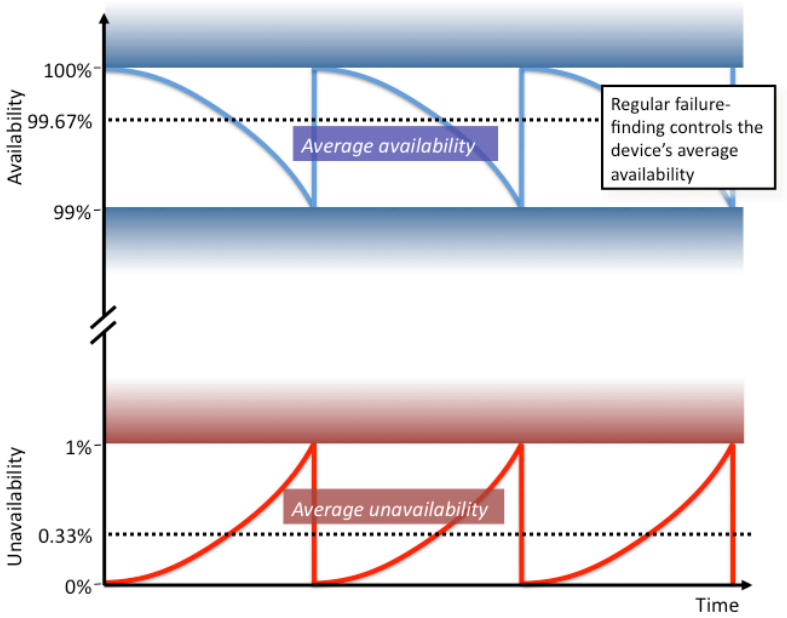
If the failure-finding interval is less than about 5% of the protective device's mean time between failures, the availability of a protective device varies linearly as shown in this graph. The availability is 100% immediately after failure-finding has taken place and it drops in a straight line until the next failure-finding task is carried out.



A failure-finding task is carried out on this device at an interval that is 10% of its mean time between failures. Its availability varies linearly between 0% and 10%, so its average over time is 5%.

Now suppose that two similar devices are connected in parallel. Either device can initiate the protective action. The failure-finding task is carried out on both devices at the same time.

The unavailability at any time is given by the product of the individual device unavailabilities, so now the device availability and unavailability look like this:



The unavailability at any time is the product of the two devices' individual unavailability levels, so the overall availability of the pair varies as the square of the time since testing. This explains why the average unavailability is now 0.33%, not 0.25% as you might have expected (5% x 5%). In qualitative terms, because the devices are tested together, the combination is very reliable immediately after the test, but the devices "grow old together" and the combination becomes much less reliable as time progresses.

11.3 Risk

Once the overall availability of the protective devices has been calculated, working out the mean time between multiple failures is easy: it is exactly the same as the method as we used for single devices.

If the mean time between demands on the system is M_{dem}, and the average unavailability of the protective system is U, then the mean time between multiple failures is

$$M_{mf} = \frac{M_{dem}}{\bar{U}}$$

So the mean time between multiple failures for a parallel protective system is

$$M_{mf} = (n+1)M_{dem}\left(\frac{M_{dev}}{T_{ff}}\right)^{n}$$

and the failure-finding task interval that achieves a required mean time between multiple failures M_{mf} is

$$T_{ff} = M_{dev}\left[\frac{(n+1)M_{dem}}{M_{mf}}\right]^{1/n}$$

Returning to the boiler example in section 11.1, let us calculate the failure-finding task interval again, but this time with two parallel relief valves ($n = 2$).

With $n = 2$, but all other data the same (M_{dem} = 20 years, M_{mf} = 4000000 years)

$$T_{ff} = M_{dev}\left[\frac{(n+1)M_{dem}}{M_{mf}}\right]^{1/n} = 50 \times \left[\frac{3 \times 20}{4000000}\right]^{1/2}$$

$$T_{ff} = 50 \times \sqrt{0.000015}\,\text{years}$$

or approximately 0.2 years.

11.4 Economic

A similar analysis gives the lowest cost failure-finding interval for a system that contains n parallel protective devices. Note that the cost of failure-finding, C_{ff}, is the cost of doing the whole failure-finding task, not just the cost of doing the task to one device.

The optimum failure-finding interval for n parallel devices (see section A.6) is

$$T_{ff} = \left[\frac{(n+1)M_{dem}\ C_{ff}\ M_{dev}^{n}}{n\ C_{mf}}\right]^{1/(n+1)}$$

11.5 Different Protective Devices in Parallel

Redundant protective systems usually consist of identical switches, relief valves or other protective devices in parallel. Occasionally a system is made from different devices, or components with different reliability characteristics. This may be a deliberate design decision to avoid common cause failures, or it may be the result of replacing obsolete components by non-identical, upgraded models. In either case it is possible that the mean time between failures of the devices may be different.

The analysis carried out in section 11.2 only needs a little modification in order to work. There are n individual protective devices in the system, but each has a different mean time between failures. If M_{dev1} is the mean time between failures of device 1, M_{dev2} is the MTBF of device 2, and so on up to n, then the unavailability of the system at time t is

$$U = \left(1 - e^{-t/M_{dev1}}\right)\left(1 - e^{-t/M_{dev2}}\right)\left(1 - e^{-t/M_{dev3}}\right)\ldots$$

Section A.4 shows that the average unavailability achieved is

$$\bar{U} = \frac{1}{n+1}\left(\frac{T_{ff}^{n}}{M_{dev1}\, M_{dev2}\, M_{dev3} \ldots M_{devn}}\right)$$

and the failure-finding interval that achieves availability A is

$$T_{ff} = \left[(n+1)(1-\bar{A})M_{dev1}\, M_{dev2}\, M_{dev3} \ldots M_{devn}\right]^{1/n}$$

The failure-finding interval to achieve a defined mean time between multiple failures M_{mf} is

$$T_{ff} = \left[\frac{(n+1)M_{dem}\, M_{dev1}\, M_{dev2}\, M_{dev3} \ldots M_{devn}}{M_{mf}}\right]^{1/n}$$

and the economic optimum failure-finding interval is

$$T_{ff} = \left[\frac{(n+1)M_{dem}\ C_{ff}\, M_{dev1}\, M_{dev2}\, M_{dev3} \ldots M_{devn}}{n\, C_{mf}}\right]^{1/(n+1)}$$

11.6 How Many Parallel Devices?

This section has shown that multiple redundant parallel protective devices can have significant advantages over single devices. These include

- Longer failure-finding task intervals can achieve the same level of risk or availability
- Failure-finding may be feasible for multiple devices when a single device would require checking too frequently to be feasible

If two devices are better than one, and three are better than two, why should we not use parallel devices in every protective system to achieve availability levels as close to 100% as we can?

There are many reasons why simply employing more parallel devices eventually fails to deliver the theoretical levels of availability and risk.

Protective System Design

While many protective system designs duplicate the sensors which detect abnormal conditions, many do not duplicate alarm and trip annunciators and actuators. Even when both sensors and actuators are duplicated, signal lines and controllers may not be. Duplication of components reduces unavailability of those components; eventually the overall system availability is dominated by those components which have not been duplicate or which cannot practically be duplicated.

Common Cause Failures

The analysis in this section has made several assumptions, but the central supposition is that failures of all the protective systems are *independent*. Over and above the requirement for failures to occur at random, it means that observing that one device has failed does not make it any more or less likely that one of the other devices has failed.

At first sight this seems a safe assumption. If one relief valve cannot open, why should its twin be in a failed state as well? In practice there can be many reasons: both may have been exposed to the same (abnormal) corrosive conditions; both valves may have been damaged by falling equipment; or the isolation valves before both valves may have been closed during maintenance and not re-opened.

Electronic equipment can be prone to a variety of common cause failures which may disable all devices simultaneously, or make them more likely to fail at or close to the same time. Some possible common causes include

- Loss of power
- Failure of signal and network lines
- Condensation or water leaks affecting all devices
- Overheating (or cooling, depending on environment) of all devices
- Vibration
- Structural damage, for example because of an explosion

Testing Procedures

Testing protective devices is a fundamental part of achieving acceptable levels of risk. In many cases the protective system, or part of it, must be disabled in order for the test to be carried out. There is a risk, often a significant one, that the personnel who carry out the test will forget to re-enable protective system when maintenance is complete. It is not only important that the testing task is carried out correctly; it is essential that the device is working afterwards.

It is difficult to obtain reliable estimates of the probability of leaving protective systems disabled. In any case, the chance depends on both the task being carried out and the design of the system. Many estimates of human error suggest that figures around 0.1%–1% may not be unusual if no special measures are taken.

If a protective system is expected to deliver 90% availability, system unavailability of 0.1%-1% due to the maintenance task may be acceptable. If the system is expected to achieve 99.9% or higher availability, then the failure-finding calculation becomes almost irrelevant because the system's unavailability is dominated by human error, not by failure of its components. Therefore, when very high availability levels are required, it is essential to be sure that the chance of the system being functional after the test is sufficiently close to 100%.

A specific problem arises with multiple redundant protective systems. For these systems, it is important to test each duplicated sensor, actuator or any other component individually. How feasible or easy this is depends on the system design, but it is not sufficient simply to replicate a demand on the system and check that there is a response.

> A tank is used to store a liquid intermediate component in a polymer fabrication facility. The pump which fills the tank is controlled by level meters in the tank itself. Because the product is potentially harmful, two trip switches at the top of the tank shut down the pump if the level reaches 95% of the overall tank volume.
>
> The mean time between failures of the trip switches is estimated to be about 100 years. Demands on the trip system are estimated to occur about once every five years, and the minimum tolerable mean time between multiple failures is 100000 years.
>
> The trip switches are checked during a yearly system shutdown and cleaning procedure. The cleaning fluid is pumped into the tank using the normal process pump, but the operators disable the normal level control. A technician watches the level gauges and ensures that the pump shuts down at 95% of the tank volume.

This is a reasonable testing procedure which for convenience uses the harmless cleaning fluid rather than the polymer precursor, and so it is likely to do little damage if the pump does not shut down in time. However, since the technician is only looking for the pump shutdown, he or she has no idea whether one trip switch operated or both. As we have already seen, if the technician assumes that both are working and the system continues to operate with only one trip switch functional, the availability of the trip system is very substantially reduced. What effect does this have on the mean time between multiple failures?

If both switches are checked and operational after each failure-finding task, the mean time between multiple failures is

$$M_{mf} = (n+1)M_{dem}\left(\frac{M_{dev}}{T_{ff}}\right)^n = 3 \times 5 \times \left(\frac{100}{1}\right)^2$$

or 150000 years.

If only one switch is operational, the expected mean time between multiple failures drops to

$$M_{mf} = (n+1)M_{dem}\left(\frac{M_{dev}}{T_{ff}}\right)^{n} = 2 \times 5 \times \left(\frac{100}{1}\right)^{1}$$

or 1000 years, a factor of 150 less than the original mean time between multiple failures, and one hundredth of the organisation's requirement.

Spurious Operation

The achievement of higher availability comes at a cost which cannot be avoided: an increase in the number of spurious trips due to mis-operation of one device. If any of the devices is able to process a demand on the protective system, then it follows that the number of spurious operations, where the protective system operates although conditions are normal, increases with the number of devices. If a single device trip system spuriously shuts down a compressor once per year, then implementing a two device trip system will result in one shutdown every six months. If the number of nuisance operations is a problem, other system designs such as 2-of-3 or 3-of-5 can deliver a compromise between high availability and low spurious operation rates (see the later section in this book for more detail).

11.7 Assumptions

The assumptions made in calculating the failure-finding interval of parallel devices are slightly different from those for a single device.

- Each protective device fails at random
- There are no common mode or common cause failures that will affect the devices
- Each protective device is guaranteed to be working immediately after installation
- The failure-finding task is always effective: the task discovers 100% of non-working devices and any devices that have to be repaired are fully functional immediately after the task has been carried out
- Demands on the protective devices occur at random
- The failure-finding interval is less than about 5% of each individual protective device's mean time between failures

- The failure-finding interval is much less than the mean time between demands on the devices
- The time taken to repair a broken protective device is insignificant, or other measures will be taken to prevent multiple failures if a scheduled task discovers a failed protective device

11.8 Examples

Standby Generators

A remote medical facility is subject to infrequent power outages that can last for at least several hours. When mains power is not available it relies on a single diesel generator which starts automatically when mains power is lost.

To reduce the risk of a power outage, it has been proposed that an additional standby generator should be installed. Either generator is capable of supplying all the power that is required. The overall mean time between failures of similar generators at other installations is about 2 years. The mean time between demands on the system is about one year.

The review group has decided that the multiple failure (loss of standby power when it is required) should happen no more often than once every 10000 years. How often should the generators be tested?

There are two standby devices, so the formula that we need for a risk target is

$$T_{ff} = M_{dev}\left[\frac{(n+1)M_{dem}}{M_{mf}}\right]^{1/n}$$

Substituting the values from the example:

$$T_{ff} = 2 \times \left[\frac{3 \times 1}{10000}\right]^{1/2} = 0.35 \text{ years}$$

The generators should be tested about every two weeks.

11.9 Key Points and Review

A parallel redundant protective system is one that is made of multiple protective devices, any one of which can prevent the multiple failure.

Using multiple protective devices can substantially increase the availability achieved for a given failure-finding task interval, compared with the availability of a single protective device.

Substituting a parallel protective system for a single device may make it possible to achieve tolerable levels of risk while ensuring that the required failure-finding intervals are feasible.

Failure-finding intervals based on availability, risk or cost are typically longer for a parallel redundant system than for an individual device.

Although very high availability figures may be theoretically attainable, in practice designers must be very careful to assess the effects of common cause failures. Real world availability levels may be substantially below calculated figures for many reasons, including common cause failures and because the protective system may be left in a disabled state after carrying out failure-finding tasks.

The equipment design and failure-finding task description must be carefully chosen to ensure that all devices can be tested and that they are all operational after the failure-finding task is completed.

The disadvantage of adding parallel protective devices is that the number of spurious operations increases as the number of devices increases. Therefore operational availability levels may drop if too many devices are employed.

12 Imperfect Testing

12.1 Introduction

Testing a protective system almost always disturbs it in some way. It may be necessary to disable the device, or part of it, in order to carry out the test. The valve leading to a low oil pressure switch may be closed so that a test can be carried out without closing down its associated lubrication system. A fire alarm's monitoring system is disconnected during a scheduled test so that the fire services are not called. However well conceived the task descriptions may be, and however well trained the maintenance personnel, there is a finite probability that the protective system will be left disabled or compromised immediately after the task has been carried out.

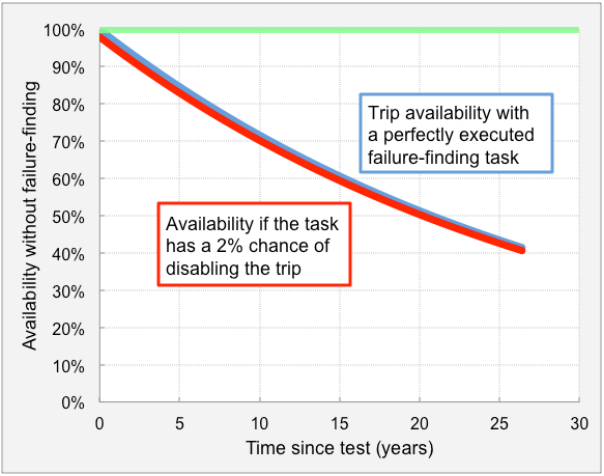
In other cases, the test stresses the protective device so that it may fail immediately after the test. Although the result of the test could suggest that the device is working, it would fail to operate if it were required to prevent a multiple failure. Any mechanical parts can wear during failure-finding; corrosion and erosion can result from contact with product or other fluids; switch contacts are stressed by each test carried out, making them more likely to fail each time they are tested.

A key point to remember is that, no matter how often it is tested, the availability of a protective device can never be higher than it is immediately after replacement or testing.

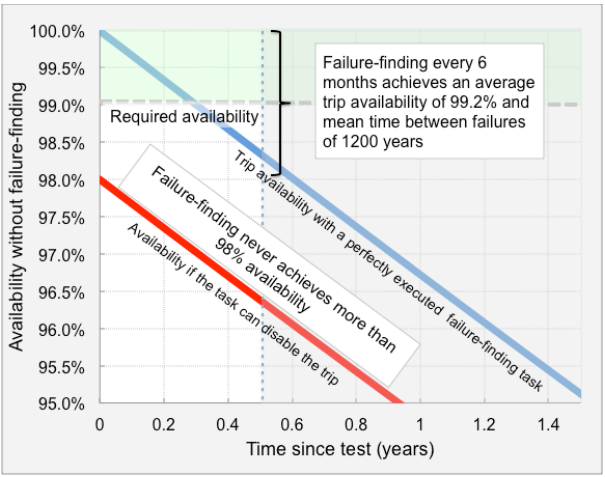
> A low pressure protective system is designed to trip a polymer production process if the system pressure falls below a set limit. Some pressure excursions happen during normal operation, so a time delay only shuts down the system if the pressure is too low for more than 5 seconds.
>
> The review group estimates that the mean time between demands on the shutdown system is about once every 10 years, and the protective system's MTBF is estimated to be about 30 years, although experience is limited. The cost of a multiple failure can vary significantly, with estimates between $10,000 and over $1m in the worst case. After a lot of discussion, the group agrees a risk-based mean time between multiple failures of 1000 years, giving a failure-finding interval of 6 months.

The most effective way to test the shutdown system is to isolate it from the process controller, then lower the system pressure to a point just below the set trip level for a few seconds and check that a shutdown signal is generated. Because shutting down production is expensive, this is normally carried out with the system online, with a technician monitoring the system pressure manually and able to trip production from a local switch if necessary.

The review group realised that there was a risk that the shutdown system could be left isolated when the test was complete. They took some time to discuss the chance of this happening, reviewing all the evidence available for similar systems, and concluded that the chance of the system being left non-operational was at least 2%.



With a tolerated mean time between multiple failures of 1000 years and demand rate of 1 in 10 years, the trip system needs to have an average availability of at least 99%, or 1% unavailability. Unfortunately there is a 2% chance that the device is disabled right after the test. Whatever the chosen task interval, failure-finding cannot achieve the availability that is needed.



12.2 A Warning

Imperfect testing is subject that has to take into account a complex interplay between task descriptions, equipment design and human error. It is impossible to give firm, general guidance on how to set up failure-finding where the unreliability of testing is a important consideration. The most that a book can do is to point out the areas of concern that you should consider.

If you are in any doubt at all, find an expert who can help to analyse your specific installation and validate your decisions.

12.3 Testing Disables the Protective System

Sometimes it is necessary to disable part or all of the protective system in order to carry out the test. This may prevent the system under test from shutting down production or taking some other undesirable action, but it introduces two problems.

- It does not test the whole of the system because part of it has been deliberately disabled
- The engineer carrying out the test might forget to re-enable part or all of the protective system, leaving it in a failed state after the task and exposing the organisation to the risk of a multiple failure

The examples below illustrate what could happen in some specific cases.

Test	Effect of leaving the system disabled
Disabling the water deluge during a fire system test	The fire alarm might activate, but the water deluge that should help to extinguish the fire would not be available, leading to additional damage and loss
Isolating relief valves to test them without pressurising the main line	If the isolations are not removed after the test, the system could leak or explode if its pressure reaches abnormal levels
Gasoline task high level switch output disabled for testing	If the switch is still disabled after the test, the tank could overfill and fail to shut down the feed pump. In the worst case this could lead to an explosion as it did in the Buncefield incident described in chapter 1.

12.4 Testing Stresses the Protective System

Failure-finding tasks are generally far more frequent than real demands on the protective system. Most of the stress or wear on parts of the protective system occurs through testing. Failure-finding itself can therefore become a source of both hidden and evident failure modes.

Failure modes that result from stress during a failure-finding task should be analysed in the same way as any others, taking into account the consequences (hidden or evident) and identifying appropriate monitoring,

Some examples of task-induced failure include

- Wear of moving parts in standby pumps and emergency generators
- Fatigue caused by pressure or temperature cycling during a test
- Corrosion or erosion of valves and pipework due to exposure to product or other fluids such as water, steam or air
- Damage to high voltage contacts caused by breaking or making a circuit under load
- Failure of electronics caused by current transients during testing

If the testing process itself can cause hidden failures, so that the device appears to work when it is tested but it is left in a failed state soon after the test, then the first action should be to review the testing procedure and the protective system design to try to eliminate the issues.

Only use the techniques described in this chapter for estimating modified failure-finding intervals if:

- Eliminating failure modes caused by testing is not possible
- The failure modes that could disable the device **occur completely at random**, with no relationship between the device's age and the probability of failure
- It is possible to estimate the worst case risk that the device will be disabled after testing

12.5 Failure-Finding Intervals with Imperfect Testing

Single device

The chance that the protective device is left inactive after the test, for example because its output is inhibited, is p. If $p = 0$, the device is never left in a failed state by the test; if $p = 1$, then it never works after testing.

With the usual assumptions:

- The protective device fails at random
- Demands on the protective device occurs at random
- The chance of the protective device being left disabled after the test is not related to its state before the test, its age, or to any other event
- The failure-finding interval is less than about 5% of each individual protective device's mean time between failures
- The mean time between demands on the protective device is much greater than the failure-finding interval

then the failure-finding interval required to achieve an average availability A is

$$T_{ff} = 2M_{dev}(1 - p - A)$$

For a risk-based calculation with average mean time between failures M_{mf}, the failure-finding interval is

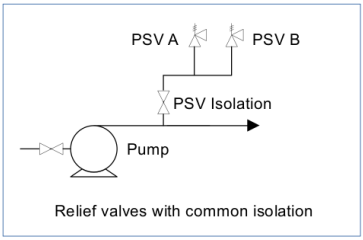
$$T_{ff} = \frac{2M_{dev}}{(1-p)}\left(\frac{M_{dem}}{M_{mf}} - p\right)$$

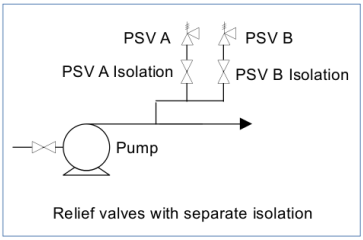
In either case the failure-finding interval will be negative if the chance of the protective device working after the failure-finding task is less than the required device availability. A negative task interval tells you that failure-finding is not effective and some other action must be taken, for example:

- Change the task so that the chance of the device being disabled by the task is reduced
- Redesign the protective device or replace it with a different type to make it more likely to work after a scheduled test
- Redesign the system of isolations to make mistakes less likely
- For risk-based calculations, significantly reduce the demand rate on the protective device

Parallel and voting systems

The failure-finding calculations for parallel and voting systems differ depending on whether all of the individual devices would be disabled after the task, or only one device is likely to be left disabled. In the example below, two parallel pressure relief valves protect a pump's output line. In the left hand diagram they are configured so that the pair is isolated for testing. If the test engineer leaves the isolation valve closed, neither PSV would be able relieve excess pressure in the line.





Each relief valve in the right hand example has its own isolation valve, so if the engineer forgets to open one of the valves, he or she might remember to open the second.

It is possible to write down formulae for failure-finding in both cases, and also to model more complex systems. They are not included in this book for two reasons.

- It is not possible to develop a general expression for the failure-finding task interval. The equations have to be solved numerically to determine the task interval.
- More importantly, the calculation depends strongly on *correlations*. Even in the simple example of two independently isolated relief valves above (the right hand diagram), it would probably be wrong to assume that an engineer will forget to open one of the valves at random. Our instinct says that if one valve is left closed, there is a significant risk that the other valve will be closed as well. The chance of this happening depends on details of the equipment, the task, and even the engineer.

12.6 Systems where Failure-Finding is Impractical

Failure-finding may be impractical for a number of reasons.

- Failure-finding would destroy or disable the device (for example, a car airbag or a pressure relief bursting disc)
- It is impossible to access the protective device
- The failure-finding task would be too hazardous or too expensive to carry out
- Demands on the protective device are so frequent that it is impractical to test the device at an interval that is less than the mean time between demands on it

Trip abuse

In almost all cases, very high demand rates are a sign of *trip abuse*, where operation of a protective device that was intended to be used only in an emergency becomes part of normal working. The right course of action is to reduce the demand rate so that failure-finding becomes feasible.

This could involve:

- Making the protected system more reliable by appropriate maintenance or improved design
- Changing the operating procedures to reduce or eliminate demands on the protective system
- Introducing an additional level of protection, such as an alarm that alerts operators before the trip is reached

If failure-finding is impossible

If failure finding cannot be carried out and the demand rate cannot be reduced far enough, we have to be certain that the risk of multiple failures is acceptable without testing.

To do this we calculate the rate of multiple failures without testing and compare it with the tolerable level of multiple failures. The mean time between failures with no failure-finding task, assuming that the protective system is repaired or replaced if a multiple failure occurs, is

$$M_{mf} = M_{dev} + M_{dem}$$

The derivation of this deceptively simple formula is given in the mathematical annex.

12.7 Testing Causes the Multiple Failure

Test procedures usually simulate demands on protective system: introducing smoke into a detector rather than actually starting a fire in an office, or interrupting the connection to a level switch to simulate an empty oil tank instead of draining the tank. Sometimes it is not possible to simulate fault conditions in a way that actually tests the protective system, and the only option is to generate a real demand and observe exactly what happens.

Now we have a problem. If the multiple failure effects are severe, not testing the protective system is unthinkable; on the other hand, the test itself is now inherently dangerous. The protective system may not operate at all, or it may operate incorrectly. Other, possibly unrelated, systems may fail during the test and lead to catastrophic consequences.

You can see just how complex real-demand testing can be by looking at this turbine over-speed test experience in a Cincinnati nuclear plant (Ornstein, 1995).

> The Salem Unit 2 plant was carrying out a steam turbine over-speed test.
>
> During tests an operator held down a lever to isolate the mechanical overspeed system from the auto-stop oil (AST) system, which would normally have tripped the reactor on high turbine speed. With the mechanical trip disabled, the turbine should still have been protected by a separate system designed to shut down the turbine through three redundant solenoid valves (SOVs) if it reached 103% of normal speed.
>
> Just after testing the mechanical over-speed system, with the equipment still in testing configuration, there was a brief dip in AST pressure that lasted for about 1.5 seconds. The pressure drop was interpreted as a turbine trip signal and initiated an automatic reactor scram. In turn, the reactor protection system signalled the solenoid valve ET-20 (one of the three ultimate turbine trip SOVs, still operational during testing) to trip the turbine. In addition, one of the AST system's pressure switches had been incorrectly set, so it failed to detect the initial turbine trip; if it had worked, it would probably have reduced the governor demand to zero and avoided what followed.
>
> When the brief oil pressure dip cleared, the hydraulic system repressurised and the steam valves began to open again. Unfortunately the reactor trip had started a 30-second timer; when it expired, circuit breakers opened to isolate the generator from the grid. The turbine was now unloaded with its steam valves close to the fully open position, and it began to over-speed.
>
> At 103% of normal speed, the ultimate protective system signalled two solenoid valves to close the governor valve. Both SOVs failed to respond. Meanwhile, for reasons that are not clear, the operator continued to hold the test lever, isolating the (functional) mechanical protective system from the turbine.

> When the turbine reached 2900 rpm, 160% of its design speed, blades broke off and penetrated the 1¼ inch thick steel casing, making holes more than two feet across. Some parts landed 100 yards away. A large section of the 1¼ inch casing landed on a truck 40 yards from the turbine. 100 condenser tubes were cut. Meanwhile, high vibration had caused the generator's hydrogen gas cooling system seals to leak, causing an explosion and starting a fire. The fire spread to the seal oil system and fire-fighters took several hours to extinguish it completely.
>
> The plant was shut down for six months for repairs at a cost of between $100m and $600m. It is interesting to note that the manufacturer had originally estimated the rate of turbine missile ejection events at 10^{-7} to 10^{-6} per turbine-year.

The incident inquiry cited the following primary causes.

- Lack of understanding of the sensitivity of hydraulic oil to contaminants
- Failure to understand that solenoid-operated valves have a limited design life
- Did not appreciate the need for individualised testing for redundant components
- Failure to provide backup when defeating protective equipment during tests
- Failure to provide operators with clear instructions on what to do if a test anomaly occurred
- Failure to consider human factors in a stressful test environment

The overall conclusions were that the incident resulted from inadequate turbine control and protective equipment maintenance and poor periodic testing of turbine control and protective equipment.

12.8 Human Issues

Human issues such as experience, attention to detail, the time of day when the task is carried out, and even mood, influence the risk of making mistakes while carrying out a scheduled failure-finding task. Some of these are discussed at greater length in Chapter 7.

Determining the exact level of risk involved is almost never possible. Often the best way to proceed is by estimating the worst case risk, using the available information from a number of sources.

- Any experience of incorrect maintenance completion on the protective system
- "Near miss" reports when engineers or supervisors have found disabled protective systems on operational equipment
- If there are several similar protective systems that are accessible, check whether any of them are currently disabled

- Experience on similar equipment either within the same plant or elsewhere
- Comments from test engineers on the current or proposed failure-finding task

12.9 Key Points and Review

A protective system that has been tested successfully may be left in a failed state due to mistakes made by the tester, stress on the system caused by testing, and other factors.

Even a low level of risk that the device will be left disabled could have an impact on failure-finding intervals.

It is difficult to estimate the chance of disabling protective systems. Use worst case risk estimates if necessary.

If failure-finding is impractical because of a high demand rate, action should be taken to reduce the demand rate

In rare cases it is impossible to carry out failure-finding because the demand rate on the protective system is too high. The default risk of a multiple failure can be calculated and compared with the organisation's maximum tolerable risk.

The impact of imperfect testing depends on several factors including human error rates. Consult an expert who can help with the analysis and auditing of failure-finding tasks.

13 Practical Analysis Guidance

13.1 Introduction

This chapter brings earlier threads together and provides a set of notes to guide you through analysis of a hidden function using RCM.

Is condition monitoring, overhaul or scheduled discard applicable?

Don't jump immediately to failure-finding for a hidden failure. It is better, if possible, to deal with the failure through condition monitoring, scheduled restoration or scheduled discard because these prevent the failure rather than allowing the device to fail. Only start to ask the failure-finding question if the group has already answered *No* to the first three questions.

Is it possible to check whether the device has failed?

Ask this question before starting the failure-finding calculation. If the answer is *No*, the answer to the failure-finding question is *No* and you do not need to carry out the calculation at all.

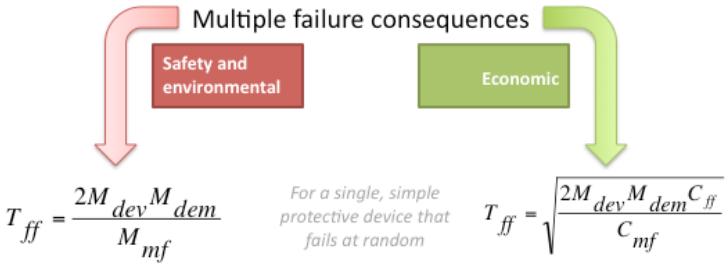
If the answer is *Yes*, take a deep breath.

Risk or Economic?

Describe the multiple failure to the group using the names of the protected system and protective device.

"The multiple failure is that the lubrication system low-low level trip has failed and the oil level drops. The pump does not shut down and it could be damaged so badly that it has to be replaced."

Ask whether there are any safety or environmental consequences associated with the multiple failure. Select the formula appropriate to the system that you want to analyse.



Write down the terms

Write down all the terms in the formula you have chosen. You don't have to write these on a flipchart or board yet, but you need to know exactly what information you need so that you don't miss essential data or waste the group's time by trying to find information that isn't needed at all.

Simple Risk Formula	Simple Economic Formula
M_{dev}	M_{dev}
M_{dem}	M_{dem}
M_{mf}	C_{ff}
	C_{mf}
T_{ff}	T_{ff}

Translate the terms

Write the terms at the left of a flipchart page or board.

First describe each element's role in failure-finding ("This is the mean time between failures of the protective device"). Then translate it into the terms of the system you are analysing and write the description next to the mathematical symbol as shown below.

Simple Risk Formula

Term	What it means	Value
M_{dev}	Mean time between failures of the relief valve (how often, on average, it jams shut)	
M_{dem}	How often on average we call on the relief valve because the boiler goes overpressure	
M_{mf}	How often we are willing to accept that boiler blows up because the relief valve is jammed closed	
T_{ff}	How often we will test the relief valve	

Simple Economic Formula

Term	What it means	Value
M_{dev}	Mean time between failures of the motor overload trip (how often, on average, it would fail to a state in which it cannot trip the motor)	
M_{dem}	How often on average we call on the motor overload because the motor is stalled	
C_{ff}	How much it costs to check that the overload is operational every time we carry out the test	
C_{mf}	How much it would cost if the multiple failure occurred; i.e. that the motor is stalled, the overload does not trip it and the motor burns out	
T_{ff}	How often we will test the motor overload trip	

Get the Values

Now you know what the terms in the equations refer to, you need to fill in the values. This is probably the most difficult part of the whole process. Focus here on getting the information and, more importantly, recognising what you do not know. The steps shown here are for guidance only, and you should consult your own organisation's safety analysis procedures where appropriate.

M_{dev}

You are trying to find out the failure rate of the protective device (alarm, trip, overload, relief valve) if it were left by itself without maintenance.

You can start by asking if the device is checked at the moment, and if they have ever found it failed.

1 "Do you check this alarm/trip/relief valve at the moment?"

2 "If you do, have you ever found it failed when you did a check?"

3 "How many times have you found it failed? Over what period?"

Then calculate M_{dev}

$$M_{dev} = \frac{Period}{Number\ of\ failures}$$

If this does not work, ask:

1 "Are there any other alarms/trips/relief valves like this one on site?"
2 "If there are, have you ever found any in a failed state?"
3 "If so, how many times over what period?"

Then calculate M_{dev}

$$M_{dev} = \frac{Period \times Number\ of\ devices}{Number\ of\ failures}$$

If that doesn't produce any information you can use, ask

"Is there anywhere (manufacturer, generic data, elsewhere...) where we could get this information?"

Always treat generic data obtained from books or industry-wide surveys with caution. Mean time between failure data usually depend on the operating context of the protective device and you must ensure that the data are appropriate for your own operating context. Look out for any factors that might decrease the reliability of the protective system that you are analysing compared with other applications.

Example factors that could influence device reliability data

Environment	Reliability may be significantly influenced by vibration, erosion, temperature, corrosion, product state and so on
Duty	Is the device used in a clean area, or is it exposed to dirty gases, liquids or powdered solids?
Testing regime	Is the testing regime a significant source of stress to the device? Could the initial commissioning and testing regime influence reliability?
Construction (materials, quality, initial testing)	Do the materials used to make the device under analysis differ from those that contributed to the generic data?
Reporting bias	Are the generic data taken from devices used in a similar application or industry? Is it possible that the generic data includes only devices that are significantly more or less reliable than typical items?

M_{dem}

You are trying to find out how often the protective device has to operate for real (not on test).

You can start by asking if the protective device has ever been needed.

1 *"Have you ever activated this alarm/trip/relief valve?"*

2 *"How many times have you done it? Over what period?"*

Then calculate M_{dem}

$$M_{dem} = \frac{Period}{Number\ of\ activations}$$

If this does not work, ask:

1 *"Have you ever had any near misses which might have needed the alarm/trip/relief valve?"*

2 *"If there have been, what is the chance of a near miss turning into an incident (i.e. the multiple failure)?"*

Then calculate M_{dem}

$$M_{dem} = \frac{Period}{Number\ of\ near\ misses\ \times\ Chance\ of\ an\ incident}$$

If this does not work, ask:

1 "Are there any other alarms/trips/relief valves on systems like this one on site?"

2 "If there are, have you ever activated those?"

3 "If so, how many times over what period?"

Then calculate M_{dem}

$$M_{dem} = \frac{Period\ \times\ Number\ of\ systems}{Number\ of\ activations}$$

If that doesn't produce any information you can use, ask

4 "Is there anywhere (other users of similar systems...) where we could get this information?"

You cannot obtain "generic" demand data except for incidents that originate outside the organisation or where the reasons for the demand are not dependent on operating context (e.g. lightning strikes, loss of main utility power supply or various types of human error).

M_{mf}

There are no ways to estimate this figure.

Ask

"How often would the organisation be willing to have the boiler go overpressure with the relief valve jammed closed so that the boiler explodes?"

1 Ensure that you consult senior management if the multiple failure has safety or environmental consequences

2 Remember that there may be many hidden failure modes on site which ultimately have safety or environmental consequences. If there are 100 failure modes like this one on site and M_{mf} for each is 10000 years, the mean time between multiple failures for the whole site is 100 years, which may well be unacceptable.

Since you cannot carry out a risk-based calculation without this figure, you might expect that a section called "Practical Analysis Guidance" would give you some figures. So what stops us from simply giving you a few handy guidelines in this book?

1 The lawyers. Anyone who gives you a statement that "one multiple failure in so many years is acceptable" assumes some of the legal responsibility for the failure-finding policy that you choose. That doesn't matter too much, unless something goes wrong; in that case, you could (probably correctly) point out that someone else bears part of the responsibility for the maintenance policy.

2 We don't know the operating context of your equipment. It may be operating in an isolated environment where the multiple failure could hurt no one. It may be that the multiple failure could hurt or kill hundreds of people, some of them innocent bystanders who are not employed by the organisation.

3 The tolerance of risk in your organisation. Some processes are inherently more dangerous than others, although that does not necessarily mean that the more dangerous processes should be shut down. Consider a potentially fatal incident aboard a spacecraft, a fighter aircraft and a civilian aircraft, and it is immediately apparent that different standards apply. Although the operators of all three craft would like to reduce the risk to zero, it is recognised that some operating environments carry more risk than others.

The following figures are intended to give a feeling for the levels of risk that are typically encountered by an individual, and show how the operating context can change an organisation's view of what is acceptable. All figures are rough estimates, and of course depend on an individual's behaviour and lifestyle.

Description	Individual's approximate risk of death per year
Lightning strike	1 in 20 million (typical)
Industrial accident (typical light industry)	1 in 20000—1 in 100000
Domestic accident	1 in 10000
Road accident (car)	1 in 17000 (UK); 1 in 6500 (USA)

In general, the rate of fatality to individuals in light industry and office work is lower than that due to domestic accidents. Without careful thought, it may be tempting to conclude that any maintenance procedure which reduces the risk of a multiple failure to less than this figure might be acceptable. However, in reality the decision is more complex. Consider the following risks.

Description	Individual's risk of death
Offshore oil worker, typical, while on platform/rig	1 in 1000 to 1 in 10000 per year
Space exploration, NASA space shuttle	1 in 100 (approximate) per mission

Why are these risks considered acceptable, when they are so much higher than other risks to which the individuals are routinely exposed during the rest of their lives? It is accepted by the participants that the risks are normal *within the context of these activities*. The risks of space exploration are accepted by those who participate. Considering work on offshore installations, the situation is more complex. Although it may be possible to reduce the level of risk to offshore workers while they working on a platform, these risks would then be far less than those associated with risks such as those from helicopter transport between the shore base and the platform. Rather than expending additional resources on reducing risks *at work*, money could be more effectively spent on reducing the risks associated with *travelling to work*.

If no one is willing or able to set a figure, the most important and obvious rule is to *err on the side of safety*. The following approach may be useful, but it should be used with caution to give you an idea of whether failure-finding is going to be appropriate, and to indicate whether your current testing interval is potentially dangerous. Be aware that most individuals are not used to dealing with levels of risk in their jobs or personal lives, and our assessments as human beings can be wildly inaccurate when the probability of an event occurring is low.

Use this approach as a final resort, and only ever to check whether redesign needs to be considered.

1 Find M_{dev} and M_{dem} as usual

2 Assume a practical failure-finding interval

3 *Calculate* M_{mf}. For a single, simple protective device, this is

4 $M_{mf} = \frac{2\,M_{dev}M_{dem}}{T_{ff}}$

5 Ask

"Is this mean time between multiple failures tolerable?"

If it is very short and the failure-finding interval cannot be substantially reduced, redesign may be appropriate; if the failure mode has safety or environmental consequences, you should ensure that action is taken urgently.

Plug the Numbers into the Formula

Complete the table with the figures, plug all the figures into the formula and calculate the failure-finding interval.

Remember to express all the times in the same units (usually years).

Term	What it means	Value
M_{dev}	Mean time between failures of the relief valve (how often, on average, it jams shut)	70 Years
M_{dem}	How often on average we call on the relief valve because the boiler goes overpressure	100 Years
M_{mf}	How often we are willing to tolerate that boiler blows up because the relief valve is jammed closed	100000 Years
T_{ff}	How often we will have to test the relief valve	0.14 Years (6 weeks)

Term	What it means	Value
M_{tive}	Mean time between failures of the motor overload trip (how often, on average, it would fail to a state in which it cannot trip the motor)	100 Years
M_{ted}	How often on average we call on the motor overload because the motor is stalled	25 Years
C_{ff}	How much it costs to check that the overload is operational every time we carry out the test	$20
C_{mf}	How much it would cost if the multiple failure occurred; i.e. that the motor is stalled, the overload does not trip it and the motor burns out	$3500
FFI	How often we will test the motor overload trip	5 Years

Checks and Balances

It isn't all over just yet. Most of the formulae in this book are only valid for a certain range of values. If you're outside their range of validity, your failure-finding values may be incorrect. Worse, failure-finding might not be the right option at all for this failure mode.

Formula Validity

Work out the average unavailability

$$\overline{U} = \frac{T_{ff}}{2\,M_{dev}}$$

Now check if U is greater than about 0.05 (5%). If it is greater than 0.05, the formula is outside its range of validity.

Task Feasibility

Check the availability figure that you calculated above.

$$\overline{U} = \frac{T_{ff}}{2\,M_{dev}}$$

If it is very low — exactly how low depends on the task you have in mind — you must seriously question whether the task can be done well enough to guarantee that level of unavailability.

Example

A pressure switch is used to shut down a process if the pressure in a reaction vessel rises above 10 000 kPa. If the pressure switch failed to operate and the vessel pressure rose violently, the vessel could explode and cause a reportable environmental incident. Although no-one is usually present in the area, it is possible that maintenance or operations personnel could be injured or even killed in the incident.

After some discussion, the review group decides to calculate a failure-finding interval for the switch based on the following data:

Term	What it means	Value
M_{dev}	Mean time between failures of the pressure switch (how often, on average, it fails closed)	250 Years
M_{dem}	How often on average we call on the pressure switch because the reactor vessel goes overpressure	10 Years
M_{mf}	How often we are willing to accept that reaction vessel blows up because the pressure switch has failed	50000 Years
T_{ff}	How often we will have to test the pressure switch to achieve (based on risk formula)	5 Weeks

The task the group has in mind is to isolate the pressure switch at a local block valve and attach a small pump to pump up the isolated leg to the trip pressure.

The facilitator calculates the unavailability $U = T_{ff}/(2\ M_{dev}) = 0.0002$ (0.02%).

The group thinks that the chance of leaving the pressure switch disabled after the test is not less than 0.02% (one in 5000) and might be as high as one in 1000, so that task as envisaged is not feasible.

If you find that the required availability is very high, some of the options available to you are:

- Look for some other, more fool proof way to test the device. This might involve testing a whole instrumentation loop rather than isolating sections of it to carry out individual tests.
- Make the test instructions more detailed and incorporate double-checking at the end of the procedure ("...After completing the test, the supervisor is to check independently that the block valve to the pressure switch has been opened. He/she must sign the job sheet to confirm that the check has been made.").
- Conclude that failure-finding is not a viable failure management strategy and look at alternative maintenance or design strategies.

$M_{dem} >> T_{ff}$

Ensure M_{dem} is much bigger than T_{ff}. If it is less than the failure-finding interval or about the same value, failure-finding is likely to be ineffective since the device is being operated just about as often as it is being tested. T_{ff}/M_{dem} should be around 0.25 or lower.

Example

> An overhoist switch protects a crane from being raised too far and possibly dropping its load. The operator currently hits the limit about once per day. Therefore failure-finding at an interval greater than one day is unlikely to have much effect on the availability of the device.

This condition usually indicates **alarm** or **trip abuse**: the system is poorly controlled and the alarms and trips are being used as control systems, not as emergency systems.

No one ever said that failure-finding would be easy, but by knowing the basics and applying them consistently, you can not only reduce your stress level but also keep the motivation and attention of your review group.

13.2 Key Points and Review

Follow a structured approach to failure-finding calculations.

Consider alternatives to failure-finding which could prevent failure of the protective system altogether.

Ensure that can state clearly the protective device, the demand, and the multiple failure for each failure mode that you analyse.

Choose the equations to be used based on risk or economic requirements. Write down each term of the equation together with a description of how it relates to the system that you are analysing.

Be systematic when sourcing data, and ensure that the information that you obtain applies to the operating context of the equipment.

Take extreme care when establishing tolerable multiple failure rates. Remember that your organisation may be responsible for hundreds or thousands of failure modes that have serious consequences. Consult at whatever level is necessary within your organisation to obtain robust figures. Consider all applicable statutory requirements and external regulatory bodies.

Ensure that you convert all data to the same units (years, months, hours etc.) before applying the formulae.

Check the failure-finding interval to ensure that the formulae are valid and that failure-finding is applicable to the failure mode that you have analysed.

A Mathematical Annex

A.1 Notation

The following symbols are used throughout the text.

A	Availability
u	Unavailability
$R(t)$	Survival function
$F(t)$	Probability that the system has failed at time t
T	Failure-finding interval
λ	Failure rate of an individual protective device in such a way that it does not provide the required protection
μ	Demand rate on the protective system
ρ	Failure rate of an individual protective device in such a way that it sends a spurious trip signal
L	Is the rate of multiple failures
r	Is the spurious trip rate of the entire protective system
n	The number of parallel independent protective devices making up a protective system
k	In a k of n voting system, the number of protective devices which must vote in order for the protective system to operate
C_{ff}	The cost of a failure-finding task every time it is carried out
C_{mf}	The cost of a multiple failure every time it occurs
α	Probability that a single protective device is capable of operating immediately after a test

A.2 Approximations

Unless otherwise stated, all results in this document are only valid if all the following approximations apply.

- $T \ll \frac{1}{\lambda}$ (in practice, $\lambda T \leq 0.5$)
- $T \ll \frac{1}{\mu}$
- Both the demand on the protective system, and the failures of the protective system itself, occur at random

A.3 Linearity of the Survival Curve

The probability that a single protective device is in a functional state at time t is given by the survival curve

$$R(t) = e^{-\lambda t}$$

This can be expanded as

$$R(t) = 1 - \lambda t + \frac{(\lambda t)^2}{2!} - \frac{(\lambda t)^3}{3!} + \ldots$$

For small values of λt, the survival curve is approximately linear, since terms in $(\lambda t)^2$ and higher can be ignored.

$$R(t) \approx 1 - \lambda t$$

A.4 Availability

The following formulae for the availability of protective systems are fundamental to the development of the more complex failure-finding equations considered later in the text. Those below deal purely with the availability of a protective system, without considering the rate of demands on the system or the ultimate consequences if the protective system fails when it is needed.

Single Device

If a device fails at a random rate λ, then provided that we are certain that the device is functional at time $t = 0$, the probability that it will operate at time $t > 0$ is given by the survival curve $R(t)$:

$$R(t) = e^{-\lambda t}$$

The instantaneous unavailability of the protective device is

$$u(t) = 1 - e^{-\lambda t}$$

Thus if the device is restored to working condition at regular intervals T, the average unavailability of the device over that interval is

$$\bar{u}(t) = \frac{\int_0^T u(t)dt}{T}$$

Under the approximations listed in section 3, the average unavailability of the protective device is

$$\bar{u}(t) = \left(1 + \frac{1}{\lambda}e^{-\lambda T} - 1 - \frac{1}{\lambda}\right)/T \approx \frac{\lambda T}{2}$$

If the target availability A of a protective system is known, then the required failure-finding interval is

$$T = \frac{2(1-A)}{\lambda}$$

This formula is based on an *average* availability figure: the chance that the device is in a failed state at the end of the failure-finding interval *T* is obviously higher than at the start. Therefore the *instantaneous risk* of a multiple failure at the end of the period is higher than at the start of the period. Under the approximations given in section 3, the probability that the device is in a working state drops linearly from 100% over the failure-finding interval. For systems employing more than one protective device in parallel, the climb in unavailability is steeper, climbing as a higher power of the time elapsed since the last failure-finding task was carried out.

The spurious trip rate of the protective system is independent of the testing interval:

$$r = \rho$$

Parallel Protective Devices

These systems consist of several identical parallel protective devices, any of which alone can provide full protection when a demand is placed on the system. In this section it is assumed that all devices are tested and if necessary repaired when the failure-finding task is carried out.

The instantaneous probability that the device is disabled (unavailable) at a time *t* after the last test is

$$u(t) = \left(1 - e^{-\lambda t}\right)^n$$

where *n* is the number of parallel protective devices employed. Under the approximations stated in section 3, the average availability over the failure-finding interval *T* is

$$\bar{u}(T) = \frac{(\lambda T)^n}{n+1}$$

As in the section above, this represents the average availability over time. The *instantaneous* availability of the protective system is higher than the average availability at the start of the period, but lower at the end. The rise in unavailability is nonlinear: quadratic, cubic and so on depending on the number of parallel devices. If the failure-finding interval is lengthened, the unavailability (and hence the potential multiple failure rate) increases as the *n*th power of the testing interval.

Given a target availability *A*, and a system consisting of *n* parallel devices tested at the same time, the required testing interval is

$$T = \frac{\sqrt[n]{(n+1)(1-A)}}{\lambda}$$

There is some advantage in staggering the tests of individual devices if it is technically feasible to do so. For example, instead of checking both switches in a system of two parallel level switches every year, one could check one switch at the first visit and the second six months later. For a given overall test interval the average system availability is increased; equivalently, the overall test interval is longer for a given target availability.

The spurious trip rate of the protective system is independent of the testing interval:

$$r = n\rho$$

Heterogeneous Parallel Systems

If the parallel system consists of several different types of protective device with individual failure rates λ_1, λ_2,... then the system unavailability with testing interval T is

$$\bar{u}(T) = \frac{\lambda_1\lambda_2\lambda_3\ldots\lambda_n T^n}{n+1}$$

and the testing interval corresponding to a target average availability A is

$$T = \sqrt[n]{\frac{(n+1)(1-A)}{\lambda_1\lambda_2\lambda_3\ldots\lambda_n}}$$

The spurious trip rate of the protective system is independent of the testing interval:

$$r = \rho_1 + \rho_2 + \ldots + \rho_n$$

Voting Systems

Voting systems also consist of n parallel protective devices, but at least k of them must "vote" to trip before the protective system as a whole is activated. These are known as *k-of-n voting systems*. Although the reliability of the protective system is lower than a corresponding 1-of-n system, the advantage of the configuration is that the number of spurious operations of the protective system is generally much lower than that of the corresponding 1-of-n system.

A k-of-n voting system is unable to provide protection if k-1 or fewer of its individual constituent devices are in a working state (or equivalently that n-k+1 are in a failed state). The probability that i such devices are in a failed state is

$$\frac{n!}{(n-i)!\,i!}\left(1-e^{-\lambda t}\right)^i\left(e^{-\lambda t}\right)^{(n-i)}$$

So the probability that the overall system is in a failed state is given by

$$F(t) = \sum_{i=n-k+1}^{n} \frac{n!}{(n-i)!\,i!}\left(1-e^{-\lambda t}\right)^{i}\left(e^{-\lambda t}\right)^{(n-i)}$$

This formula assumes that the spurious trip rate is small ($\rho T << 1$). Under the approximations described in section 3, this can generally be approximated by

$$F(t) = \frac{n!}{(n-k+1)!\,(k-1)!}\left(1-e^{-\lambda t}\right)^{n-k+1}$$

If the failure-finding task tests each constituent device, and each device is tested at the same time, the average system availability is

$$\bar{u}(t) = \frac{n!}{(n-k+2)(n-k+1)!\,(k-1)!}(\lambda t)^{n-k+1}$$

The failure-finding interval required to achieve a protective system availability A is therefore

$$T = \frac{1}{\lambda}\left[\frac{(1-A)(n-k+2)(n-k+1)!\,(k-1)!}{n!}\right]^{1/(n-k+1)}$$

The rate of spurious trips depends on the configuration of the protective device. There are two principal designs: in the first, an alarm is annunciated when *any* of the individual devices detects a fault condition; in the second, there is no warning of a fault until sufficient devices vote to send a trip signal.

In the first case, there is only a rate of spurious *alarms* $n\lambda$. There are no spurious trips provided that spuriously failed devices are diagnosed and repaired sufficiently quickly.

In the second case, anything up to k-1 devices may be in a spuriously failed state without any indication. In this case the rate of spurious trips depends on the failure-finding interval T. The rate of spurious trips (under the usual approximations concerning the failure-finding interval and failure rates) is

$$\frac{n!}{(n-k)!\,k!}\rho^{k}T^{k-1}$$

A.5 Multiple Failure Rates and Risk-Based Calculations

In general the average multiple failure rate of any configuration is given by the demand rate on the protective device times its average unavailability:

$$L = u\mu$$

The following formulae are derived from those in the previous section, re-arranged so that a failure-finding interval can be calculated from a target multiple failure rate.

Single Device

The average multiple failure rate *L* is given by

$$L = \frac{\mu\lambda T}{2}$$

So the failure-finding interval *T* for a given target multiple failure rate *L* is

$$T = \frac{2L}{\mu\lambda}$$

Parallel Protective Devices

The average multiple failure rate *L* is given by

$$L = \mu\frac{(\lambda T)^n}{n+1}$$

So the failure-finding interval *T* for a given target multiple failure rate *L* is

$$T = \frac{1}{\lambda} \times \sqrt[n]{\frac{(n+1)L}{\mu}}$$

Heterogeneous Parallel Systems

The average multiple failure rate *L* is given by

$$L = \mu\frac{\lambda_1\lambda_2\lambda_3\ldots\lambda_n T^n}{n+1}$$

The failure-finding interval *T* for a given target multiple failure rate *L* is

$$T = \sqrt[n]{\frac{(n+1)L}{\mu\,\lambda_1\lambda_2\lambda_3\ldots\lambda_n}}$$

Voting Systems

The average multiple failure rate *L* is given by

$$L = \mu\frac{n!}{(n-k+2)(n-k+1)!\,(k-1)!}(\lambda T)^{n-k+1}$$

The failure-finding interval *T* for a given target multiple failure rate *L* is

$$T = \frac{1}{\lambda} \times \sqrt[n-k+1]{\frac{L(n-k+2)(n-k+1)!\,(k-1)!}{\mu\,n!}}$$

A.6 Economically Optimised Failure-Finding Intervals

If a multiple failure has only economic consequences, it is possible to choose a failure-finding interval which represents a balance between testing too infrequently, where the *risked expenditure* on multiple failures is increased, and testing too frequently, where the risk of a multiple failure is reduced, but the *actual expenditure* on testing is increased.

Given a rate of multiple failure $L(T)$ and a multiple failure cost C_{mf}, the risked expenditure on multiple failures is

$$C_{mf}\, L(T)$$

The rate of expenditure on failure-finding is

$$\frac{C_{ff}}{T}$$

and the total cost is

$$C = C_{mf}\, L(T) \; + \; \frac{C_{ff}}{T}$$

The optimum interval is determined by finding the interval T which minimises the total expenditure.

Single Device

The economic optimum failure-finding interval for a single protective device is

$$T = \sqrt{\frac{2\, C_{ff}}{\lambda\mu\, C_{mf}}}$$

Parallel Protective Devices

The optimum failure-finding interval for parallel protective devices is

$$T = \sqrt[n+1]{\frac{(n+1)\, C_{ff}}{n\, \lambda^n\, \mu\, C_{mf}}}$$

Heterogeneous Parallel Systems

The optimum failure-finding interval for parallel protective devices is

$$T = \sqrt[n+1]{\frac{(n+1)\, C_{ff}}{n\, \lambda_1\lambda_2\lambda_3 \ldots \lambda_n\, \mu\, C_{mf}}}$$

Voting Systems

The optimum failure-finding interval for voting systems is

$$T = \sqrt[n-k+2]{\frac{(n-k+2)(n-k)!\,(k-1)!\;C_{ff}}{n!\;\lambda^{n-k+1}\,\mu\,C_{mf}}}$$

A.7 Maximum Allowed Unavailability

The above formulae are based on an average unavailability of the protective system. Immediately after a test the system is almost 100% available; however, toward the end of the failure-finding interval the availability of the system is less than the average availability, so the instantaneous risk of a multiple failure is higher than calculated in section 5. In some circumstances it may be preferable to base calculations on the minimum allowed unavailability of a protective system, or equivalently on the maximum allowed instantaneous risk of multiple failure.

Unless otherwise stated, all formulae in this section are exact provided that the mean time between demands on the system is much longer than the failure-finding interval.

Single Device

The failure-finding interval based on a maximum unavailability u_{max} or minimum availability A_{min} is

$$T = -\frac{1}{\lambda}\ln(1-u_{max}) = -\frac{1}{\lambda}\ln(1-A_{min})$$

Parallel Protective Devices

The failure-finding interval based on a maximum unavailability u_{max} is

$$T = -\frac{1}{\lambda}\ln(1-u_{max}^{1/n})$$

Heterogeneous Protective Devices

The failure-finding interval based on a maximum unavailability u_{max} under the usual approximation

$$T \ll \frac{1}{\lambda}$$

is

$$T = \frac{u_{max}}{\lambda_1\lambda_2\lambda_3\ldots\lambda_n}$$

Voting Systems

The failure-finding interval based on a maximum unavailability u_{max} is

$$T = -\frac{1}{\lambda}\ln\left(1 - \sqrt[n-k+1]{\frac{(n-k+1)!\,(k-1)!}{n!}u_{max}}\right)$$

A.8 Maximum Allowed Multiple Failure Rate

The formulae below calculate failure-finding intervals based on a maximum allowed multiple failure rate L_{max}.

Unless otherwise stated, all formulae in this section are exact provided that the mean time between demands on the system is much longer than the failure-finding interval.

Single Device

The failure-finding interval based on a maximum multiple failure rate L_{max} is

$$T = -\frac{1}{\lambda}\ln\left(1 - \frac{L_{max}}{\mu}\right)$$

Parallel Protective Devices

The failure-finding interval based on a maximum multiple failure rate L_{max} is

$$T = -\frac{1}{\lambda}\ln\left(1 - \sqrt[n]{\frac{L_{max}}{\mu}}\right)$$

Heterogeneous Protective Devices

The failure-finding interval based on a maximum multiple failure rate L_{max} under the usual approximation

$$T \ll \frac{1}{\lambda}$$

is

$$T = \frac{L_{max}}{\mu\,\lambda_1\lambda_2\lambda_3\ldots\lambda_n}$$

Voting Systems

The failure-finding interval based on a maximum multiple failure rate L_{max} is

$$T = -\frac{1}{\lambda} ln\left(1 - \sqrt[n-k+1]{\frac{(n-k+1)!\,(k-1)!}{n!}} \frac{L_{max}}{\mu}\right)$$

A.9 Multi-Level Protective Systems

Multi-level systems are common in applications which incorporate an *alarm* level and a *trip* level. This section examines the failure-finding options for such systems. Although calculations can be carried out for any combination of the configurations described in the preceding sections, it is assumed for simplicity that each level consists of a single protective device. The analysis in this section also makes the following assumptions:

- Failure of each level is independent (there are no common cause failures which would affect both systems simultaneously)
- If the lower level (alarm) system operates correctly, there is no demand on the higher level system
- If the higher level system operates correctly, there is no multiple failure

The second assumption should be considered carefully since, in some protective systems, the initiating incident may develop so quickly that the ultimate protective system may be activated even if the first level responds correctly. For example, a liquid storage vessel's alarm system may operate correctly, but if a sudden surge of liquid arrives the operators may have insufficient time to take preventive action before the high level shutdown system is activated.

The following notation is used below:

T_1 Failure-finding interval for the alarm (lower level) system

T_2 Failure-finding interval for the trip (higher level) system

λ_1 Failure rate of the alarm system

λ_2 Failure rate of the trip system

μ Demand rate on the lower level system

L_1 Multiple failure rate of the alarm system (demand rate on the trip system)

L Overall multiple failure rate of the system

C_{ff1} The cost of a failure-finding task on the alarm system every time it is carried out

C_{ff2} The cost of a failure-finding task on the trip system every time it is carried out

C_{mf1} The cost of a multiple failure of the alarm system (if the trip system operates)

C_{mf2} The cost of a multiple failure of the alarm and trip systems

Risk-based: Alarm and Trip Checked Together

In this case the alarm and trip circuits are checked at the same time and at a common failure-finding interval T. This is often the case when checking vessel level detectors because each device can be checked simply by filling or emptying the vessel.

If the devices are checked at the same interval T and the checks on alarm and trip loops are carried out at the same time, the average unavailability of the combined alarm and trip system is

$$\bar{u} = \frac{\lambda_1 \lambda_2 T^2}{3}$$

so the failure-finding interval needed to establish an overall multiple failure rate L is

$$\bar{T} = \sqrt{\frac{3L}{\mu\, \lambda_1 \lambda_2}}$$

Economic and Risk Based

This situation is typical in many production environments. Failure of the alarm level results in an economic loss C_{mf1} if the trip operates and stops the process. If the trip does not operate then the business is exposed to safety or environmental consequences: the maximum acceptable rate of the ultimate multiple failure is L.

Typically the alarm system is less costly to test than the trip level and often the reliability of the trip system is higher than that of the alarm level. It is not surprising that the optimum failure-finding intervals for the two levels are usually different.

Assuming that both levels of protection can be tested without significantly increasing the risk that either level is disabled following the test, the maximum rate of multiple failures can be achieved in many different ways. The alarm system could be checked very often, resulting in few demands on the second protective system and hence a low rate of expenditure on production downtime. Although production downtime costs and testing costs for the second level system would be low, the cost of failure-finding the alarm system would be high. An alternative strategy would be to check the alarm less often. This reduces failure-finding costs on that system but increases expenditure on lost production through activation of the trip level and results in higher failure-finding costs at the trip level to achieve the same ultimate multiple failure rate.

The following calculations are based on an approximation. The assumption is made that failure-finding tasks at the two levels are carried out independently and that the tasks are not related (for example, that the trip tests are not carried out at the same time as alternate alarm tests). In reality this is not likely to be the case: however, the greatest deviation is not significant in comparison with errors in the reliability and demand data used in this type of calculation.

The failure rate at level 1 is

$$L_1 = \frac{\mu\lambda_1 T_1}{2}$$

The ultimate failure rate at level 2 is

$$L = \frac{\mu\lambda_1 T_1}{2} . \frac{\lambda_2 T_2}{2}$$

The rate of expenditure, taking into account risked costs and the cost of testing at alarm and trip levels is

$$\frac{C_{ff1}}{T_1} + \frac{C_{ff2}}{T_2} + \frac{\mu\lambda_1 T_1 C_{mf1}}{2}$$

The alarm failure-finding interval which minimises this total cost is

$$T_1 = \sqrt{\frac{4LC_{ff1}}{\mu\,\lambda_1(C_{ff2}\lambda_2 + 2LC_{mf1})}}$$

The trip failure-finding interval which achieves the target multiple failure rate *L* is

$$\frac{C_{ff1}}{T_1} + \frac{C_{ff2}}{T_2} + \frac{C_{mf1}\,\mu\,\lambda_1 T_1}{2} + \frac{C_{mf2}\,\mu\,\lambda_1\lambda_2 T_1 T_2}{4}$$

Two-Level Economic System

This configuration is similar to that described above except that the consequences of failure of the alarm (first level) and trip (second level) are both economic. Typically the consequences of alarm failure are relatively small, but the consequences of failure of the trip level are severe.

The total cost including failure-finding tasks and multiple failures at both levels is

$$\frac{C_{ff1}}{T_1} + \frac{C_{ff2}}{T_2} + \frac{C_{mf1}\,\mu\,\lambda_1 T_1}{2} + \frac{C_{mf2}\,\mu\,\lambda_1\lambda_2 T_1 T_2}{4}$$

A.10 Protective Devices Disabled after the Test

In order to carry out a failure-finding task, it is common for the protective system to be disturbed in some way. Individual sensors or control units may be deliberately disabled in order to prevent a shutdown caused by the scheduled test. If there is a risk that the protective system will remain disabled immediately after the test, or if the test might stress the protective system in such a way that it is rendered non-functional immediately after the test, the unavailability introduced by the test must be taken into account when calculating the failure-finding interval.

Single Device

A device fails at a random rate λ, but we can no longer be certain that it is operational at time $t = 0$. The probability that the device operates immediately after the test is α; the probability that it will operate at a later time $t > 0$ is given by the survival curve $R(t)$:

$$R(t) = \alpha e^{-\lambda t}$$

The instantaneous unavailability of the protective device is

$$u(t) = 1 - \alpha e^{-\lambda t}$$

Thus if the device is restored to working condition at regular intervals T, the average unavailability of the device over that interval is

$$\bar{u}(t) = \frac{\int_0^T u(t)dt}{T}$$

Under the approximations listed in section 3, the average unavailability of the protective device is

$$\bar{u}(t) = \frac{1}{T}\left(T + \frac{\alpha}{\lambda}e^{-\lambda t} - \frac{\alpha}{\lambda}\right) \approx (1 - \alpha) + \frac{\alpha\lambda T}{2}$$

If the target availability A of a protective system is known, then the required failure-finding interval is

$$T = \frac{2(\alpha - A)}{\lambda\alpha}$$

Notice that in this case the failure finding interval becomes negative when $\alpha < A$, so that a failure-finding task cannot be selected. This is reasonable because the required availability is greater than the availability of the device immediately after the scheduled failure-finding task has been carried out.

A.11 Multiple Failures without Failure-Finding

It is sometimes impractical to carry out failure-finding, possibly for one of these reasons.

- It is impossible to access the protective device
- The failure-finding process itself would be too hazardous
- Failure-finding would destroy the device (for example, a car airbag or a pressure relief bursting disc)
- Demands on the protective device occur very frequently and it is impossible or impractical to carry out failure-finding at an interval less than the mean time between demands

If failure finding cannot be carried out, it is essential to ensure that the default risk of multiple failures is acceptable. To do this we calculate the rate of multiple failures without testing and compare it with the tolerable level of multiple failures.

Assumptions

This section makes the usual assumptions of random failure of the protective device and protected system. However, it makes no assumptions about the relative magnitude of the mean time between failures of the two systems. The analysis also does not assume that the survival curve of either the protective device or the protected system remains in the region where a linear approximation is valid.

The mathematical treatment below is applicable to a single protective device and a single protected system. It can be expanded to deal with more complex systems.

Model

The model used in this section is slightly more complex than that used for simple failure finding interval calculations.

The system under analysis is assumed to consist of a single protective device and a single protected system. Both systems fail at random: failure of the protected system is evident, while failure of the protective device is hidden. A multiple failure occurs if the protected system fails while the protective device is in a failed state. Initially and after a multiple failure, both the protective device and the protected system are fully operational. If the protected system fails while the protective device is operational, it is assumed that the protected system is repaired to "as new" condition without delay.

Because a multiple failure can only occur if the protective device has already failed, the system must be in one of three states.

S_1 The protective device and the protected system are operational

S_2 The protective device is in a failed state, but the protected system is still operational

S_3 The protective device and protective system have both failed (a multiple failure has occurred)

In the model below, S_1,S_2 and S_3 are used to denote the probability that the system is in the state with that name.

The transition rates between these states are defined as follows.

λ_{12} The transition rate from state S_1 to state S_2. This is the failure rate of the protective device

λ_{23} The transition rate from state S_2 to state S_3. This is the failure rate of the protected system

This notation is used to derive the overall mean time between multiple failures. Once the result is obtained, it will be re-expressed in the usual terms.

Initial Conditions

Initially the protective device and protected system are both operational, so the system has 100% chance of being in state S_1.

$$S_1 = 1$$

$$S_2 = 0$$

$$S_3 = 0$$

Transition Equations

The rates of change of occupation of the states are given by the following equations. Since there is no automatic repair of the protective device unless a multiple failure occurs, there are no transitions from state S_2 to state S_1 or from state S_3 to state S_2.

$$\frac{dS_1}{dt} = -\lambda_{12}S_1$$

$$\frac{dS_2}{dt} = +\lambda_{12}S_1 - \lambda_{23}S_2$$

$$\frac{dS_3}{dt} = +\lambda_{23}S_2$$

The first equation is easily solved to yield the exponential survival curve.

$$S_1 = e^{-\lambda_{12}t}$$

Substituting into the second equation for S_1:

$$\frac{dS_2}{dt} = +\lambda_{12}S_1 - \lambda_{23}S_2 = \lambda_{12}e^{-\lambda_{12}t} - \lambda_{23}S_2$$

Rearranging:

$$\frac{dS_2}{dt} - \lambda_{12}e^{-\lambda_{12}t} + \lambda_{23}S_2 = 0$$

This differential equation can be solved by multiplying through by the integrating factor

$$e^{\lambda_{23}t}$$

to give

$$e^{\lambda_{23}t}\frac{dS_2}{dt} - \lambda_{12}e^{-\lambda_{12}t}e^{\lambda_{23}t} + \lambda_{23}e^{\lambda_{23}t}S_2 = 0$$

which can be rearranged as the differential of a product

$$\frac{d}{dt}\left(e^{\lambda_{23}t}S_2\right) = \lambda_{12}e^{\lambda_{23}t-\lambda_{12}t}$$

Integrating:

$$e^{\lambda_{23}t}S_2 = \frac{\lambda_{12}}{\lambda_{23}-\lambda_{12}}e^{(\lambda_{23}-\lambda_{12})t} + c$$

Finally, rearranging again,

$$S_2 = \frac{\lambda_{12}}{\lambda_{23}-\lambda_{12}}e^{-\lambda_{12}t} + ce^{-\lambda_{23}t}$$

Under the initial conditions, $S_2 = 0$ when $t = 0$, giving

$$c = -\frac{\lambda_{12}}{\lambda_{23}-\lambda_{12}}$$

So we have

$$S_2 = \frac{\lambda_{12}}{\lambda_{23}-\lambda_{12}}e^{-\lambda_{12}t} - \frac{\lambda_{12}}{\lambda_{23}-\lambda_{12}}e^{-\lambda_{23}t}$$

Since S_1, S_2 and S_3 represent probabilities, we know that

$$S_1 + S_2 + S_3 = 1$$

State S_3 represents the multiple failure; substituting in the above equation and rearranging, we have

$$S_3(t) = 1 - e^{-\lambda_{12}t} - \frac{\lambda_{12}}{\lambda_{23} - \lambda_{12}}\left(e^{-\lambda_{12}t} - e^{-\lambda_{23}t}\right)$$

This is the time dependent probability of a multiple failure. We now need to use this calculate the mean time between multiple failures.

The failure density curve for multiple failures is given by

$$\frac{dS_3}{dt} = \lambda_{12}e^{-\lambda_{12}t}\left(1 + \frac{\lambda_{12}}{\lambda_{23} - \lambda_{12}}\right) - \frac{\lambda_{12}}{\lambda_{23} - \lambda_{12}}\lambda_{23}e^{-\lambda_{23}t}$$

To calculate the mean achieved life or mean time between multiple failures, this expression is multiplied by time and integrated from zero to infinity.

$$M_{mf} = \int_0^\infty \lambda_{12}te^{-\lambda_{12}t}\left(1 + \frac{\lambda_{12}}{\lambda_{23} - \lambda_{12}}\right)dt - \int_0^\infty \frac{\lambda_{12}}{\lambda_{23} - \lambda_{12}}\lambda_{23}te^{-\lambda_{23}t}dt$$

$$M_{mf} = \frac{1}{\lambda_{12}}\left(1 + \frac{\lambda_{12}}{\lambda_{23} - \lambda_{12}}\right) - \frac{1}{\lambda_{23}^2}\frac{\lambda_{12}\lambda_{23}}{\lambda_{23} - \lambda_{12}}$$

$$M_{mf} = \frac{\lambda_{23}^2 - \lambda_{12}^2}{\lambda_{12}\lambda_{23}(\lambda_{23} - \lambda_{12})} = \frac{\lambda_{23} + \lambda_{12}}{\lambda_{12}\lambda_{23}}$$

Translating this into the notation used in the remainder of this chapter, this becomes

$$M_{mf} = \frac{\mu + \lambda}{\mu\lambda}$$

or, in terms of the mean time between failures of the protective device, M_{dev}, and the mean time between failures of the protective system, M_{dem}

$$M_{mf} = M_{dev} + M_{dem}$$

B Equation Summary and Reference

B.1 Assumptions

The following assumptions have been made in deriving the formulae in this section unless otherwise noted.

$T_{ff} << M_{dev}$	The failure-finding interval T_{ff} is very much less than the mean time between failures of the protective device M_{dev} (typically $T_{ff} < 0.05\ M_{dev}$)
$T_{ff} << M_{dem}$	The failure-finding interval T_{ff} is very much less than the mean time between failures of the protected system M_{dem} (typically $T_{ff} < 0.1\ M_{dem}$)
$A > 0.95$	The required average availability of a single protective device, or if there are several in a parallel or voting configuration, each protective device, is greater than 95%
Random failure	Failures of both the protective device and protected system occur at random, with no relationship between time and the probability of failure, and no correlation between any failure and a subsequent failure
Common cause and common mode failures	Where more than one device is deployed in a protective system, there are no common cause or common mode failures that could affect more than one simultaneously

B.2 Definitions

Term	Meaning
M_{dev}	Mean time between failures of a single protective device
M_{dem}	How often on average we call on the protective system to operate
M_{mf}	The minimum tolerable mean time between multiple failures
A	Minimum average availability of the protective system.
C_{ff}	Cost of carrying out one failure-finding task
C_{mf}	Cost of the multiple failure if it occurs
n	The number of parallel protective devices. n = 1 for simple systems.
k	In a voting system, the number of protective devices that must "vote" in order to initiate protective action. Typical k and n values are "2 of 3" and "3 of 5".
p	Probability that the protective device is operational immediately after a check. (1-p) is the probability that a single protective device is disabled by the check.
T_{ff}	The interval between scheduled tests of the protective system

B.3 Availability, One Device

$$T_{ff} = 2M_{dev}(1 - A)$$

Item	Term	Meaning
Target	A	Minimum average availability of the protective system.
Data	M_{dev}	Mean time between failures of a single protective device
Output	T_{ff}	How often we will have to test the protective system

B.4 Availability, Parallel Devices

$$T_{ff} = M_{dev}[(n+1)(1-A)]^{1/n}$$

Item	Term	Meaning
Target	*A*	Minimum average availability of the protective system.
Data	M_{dev}	Mean time between failures of a single protective device
	n	Number of parallel protective devices
Output	T_{ff}	How often we will have to test the protective system

B.5 Availability, Voting System

$$T_{ff} = M_{dev}\left[\frac{(1-A)(n-k+2)(n-k+1)!\,(k-1)!}{n!}\right]^{1/(n-k+1)}$$

Item	Term	Meaning
Target	A	Minimum average availability of the protective system.
Data	M_{dev}	Mean time between failures of a single protective device
	n	Number of parallel protective devices
	k	Number of protective devices which must vote in order to initiate protective action
Output	T_{ff}	How often we will have to test the protective system

B.6 Risk-based, One Device

$$T_{ff} = \frac{2M_{dev}M_{dem}}{M_{mf}}$$

Item	Term	Meaning
Target	M_{mf}	The minimum tolerable mean time between multiple failures
Data	M_{dev}	Mean time between failures of a single protective device
	M_{dem}	How often on average we call on the protective system to operate
Output	T_{ff}	How often we will have to test the protective system

B.7 Risk-based, Parallel Devices

$$T_{ff} = M_{dev}\left[\frac{(n+1)M_{dem}}{M_{mf}}\right]^{1/n}$$

Item	Term	Meaning
Target	M_{dmf}	The minimum tolerable mean time between multiple failures
Data	M_{dev}	Mean time between failures of a single protective device
	M_{dem}	How often on average we call on the protective system to operate
	n	Number of parallel protective devices
Output	T_{ff}	How often we will have to test the protective system

B.8 Risk-based, Voting System

$$T_{ff} = M_{dev}\left[\frac{M_{dem}(n-k+2)(n-k+1)!\,(k-1)!}{n!\,M_{mf}}\right]^{1/(n-k+1)}$$

Item	Term	Meaning
Target	M_{dmf}	The minimum tolerable mean time between multiple failures
Data	M_{dev}	Mean time between failures of a single protective device
	M_{dem}	How often on average we call on the protective system to operate
	n	Number of parallel protective devices
	k	Number of protective devices which must vote in order to initiate protective action
Output	T_{ff}	How often we will have to test the protective system

B.9 Economic, One Device

$$T_{ff} = \sqrt{\frac{2M_{dev}M_{dem}C_{ff}}{C_{mf}}}$$

Item	Term	Meaning
Target	M_{dmf}	The minimum tolerable mean time between multiple failures
Data	M_{dev}	Mean time between failures of a single protective device
	M_{dem}	How often on average we call on the protective system to operate
	C_{ff}	Cost of carrying out one failure-finding task
	C_{mf}	Cost of the multiple failure
	n	Number of parallel protective devices
Output	T_{ff}	How often we will have to test the protective system

B.10 Economic, Parallel Devices

$$T_{ff} = \left[\frac{(n+1)M_{dem}C_{ff}M_{dev}^{n}}{n\,C_{mf}}\right]^{1/(n+1)}$$

Item	Term	Meaning
Target	M_{dmf}	The minimum tolerable mean time between multiple failures
Data	M_{dev}	Mean time between failures of a single protective device
	M_{dem}	How often on average we call on the protective system to operate
	C_{ff}	Cost of carrying out one failure-finding task
	C_{mf}	Cost of the multiple failure
	n	Number of parallel protective devices
Output	T_{ff}	How often we will have to test the protective system

B.11 Economic, Voting System

$$T_{ff} = \left[\frac{(n-k+2)(n-k)!\,(k-1)!\,M_{dem}C_{ff}M_{dev}^{n-k+1}}{n!\;C_{mf}}\right]^{1/(n-k+2)}$$

Item	Term	Meaning
Target	M_{dmf}	The minimum tolerable mean time between multiple failures
Data	M_{dev}	Mean time between failures of a single protective device
	M_{dem}	How often on average we call on the protective system to operate
	C_{ff}	Cost of carrying out one failure-finding task
	C_{mf}	Cost of the multiple failure
	n	Number of parallel protective devices
	k	Number of protective devices which must vote in order to initiate protective action
Output	T_{ff}	How often we will have to test the protective system

B.12 Availability, One Device, Test Disables the Device

$$T_{ff} = 2M_{dev}(1-p-A)$$

Item	Term	Meaning
Target	A	Minimum required average availability
Data	M_{dev}	Mean time between failures of a single protective device
	p	The probability that the failure-finding task disables the protective device.
Output	T_{ff}	How often we will have to test the protective system

B.13 Risk, One Device, Test Disables the Device

$$T_{ff} = \frac{2M_{dev}}{(1-p)}\left(\frac{M_{dem}}{M_{mf}} - p\right)$$

Item	Term	Meaning
Target	M_{dmf}	The minimum tolerable mean time between multiple failures
Data	M_{dev}	Mean time between failures of a single protective device
	M_{dem}	How often on average we call on the protective system to operate
	p	The probability that the failure-finding task disables the protective device.
Output	T_{ff}	How often we will have to test the protective system

C References

BOEMOE (2011) *FUSCG/BOEM Marine Board of Investigation into the marine casualty, explosion, fire, pollution and sinking of mobile offshore drilling unit Deepwater Horizon*. Bureau of Safety and Environmental Enforcement. Available at: www.bsee.gov/newsroom/library/archive/deepwater-horizon-reading-room

Cantu *et al*. (2003) *Brain injury-related fatalities in American football, 1945-1999.* Neurosurgery 52:846-853, 2003.

Chase, Nancy L, Sui, Xuemei, Blair, Steven Noel (2008) *Swimming and All-Cause Mortality Risk Compared With Running, Walking, and Sedentary Habits in Men,*. International Journal of Aquatic Research and Education Vol 2 No. 3

Davidson TM, Laliotis AT (1996) *Snowboarding injuries-A four-year study with comparison with alpine ski injuries*. West J Med 1996; 164:231-237

Fortun, Kim (2001) *Advocacy after Bhopal*. University of Chicago Press. ISBN 0-226-25720-7

Gouttebarge V, Ooms W, Tummers T, Inklaar H (2014), *Mortality in international professional football (soccer): a descriptive study*. World Football Union unpublished report, 2014.

Harvard Medical School (2010) *Marriage and men's health*. Harvard Health Publishing, Cambridge, MA, USA. Available at: https://www.health.harvard.edu/mens-health/marriage-and-mens-health

Hobbs, A (2008) *An Overview of Human Factors in Maintenance*, Australian Transport Safety Bureau. Report AR-2008-055

IARC (2012) *A Review of Human Carcinogens, Part E*. IARC Monographs on the evaluation of carcinogenic risks to humans, volume 100. International Agency for Research on Cancer. Lyon.

Judiciary of England and Wales (2010) *Case of Regina v Total (UK) Limited, Hertfordshire Oil Storage Limited, Motherwell Control Systems (2003) Limited, TAV Engineering Limited and British Pipeline Agency Ltd*. Judiciary of England and Wales.

Lichtenstein Saray, Slovic Paul, Fischhoff Baruch, Layman Mark, Combs Barbara (1978) *Judged Frequency of Lethal Events*, Journal of Experimental Psychology, vol 4 no 6, p551-578

Moubray, John (1997). *Reliability-centered Maintenance*. Industrial Press. New York, NY. 1997

NJ FACE (2009). *A 60-year-old Hispanic maintenance worker killed when a pressure vessel exploded.* New Jersey Fatality Assessment and Control Evaluation Project FACE 08-NJ-003

Nowlan, FS and Heap, HF (1978) *Reliability-centered Maintenance*. Dolby Access Press. Los Altos, CA, USA.

Ornstein H (1995). *Operating Experience Feedback Report--Turbine-Generator Overspeed Detection Systems*, Safety Programs Division, Office for Analysis and Evaluation of Operational Data, US Nuclear Regulatory Commission, Washington DC, April 1995

Perrow, Charles (1984). *Normal Accidents*. New York: Basic Books. ISBN 0-465-05142-1

Ranter, Harro (2017) *ASN data show 2017 was safest year in aviation history*. Aviation Safety Network. Flight Safety Foundation. 30 December 2017. Available at: news.aviation-safety.net/2017/12/30/preliminary-asn-data-show-2017-safest-year-aviation-history.

Regina v Total (UK) Limited and others (2010): Regina v Total (UK) Limited, Hertfordshire Oil Storage Limited, Motherwell Control Systems (2003) Limited, TAV Engineering Limited and British Pipeline Agency Ltd, Mr Justice Calvert Smith (Judiciary of England and Wales)

Sakata *et al.* (2012). R Sakata, P McGale, E J Grant, K Ozasa, R Peto, S C Darby *Impact of smoking on mortality and life expectancy in Japanese smokers: a prospective cohort study*, British Medical Journal 2012;345:e7093

Smith, DJ (2017) *Reliability, Maintainability and Risk*. Butterworth-Heinemann. ISBN 978-0081020104

Svenson, Ola (1981). "Are We All Less Risky and More Skillful Than Our Fellow Drivers?" (PDF). *Acta Psychologica*. **47** (2): 143–148. doi:10.1016/0001-6918(81)90005-6.

Taylor AJ, McGwin G, Valent F, et al (2002) *Fatal occupational electrocutions in the United States*. Injury Prevention 2002;8:306-312.

Turk et al. (2008) Br. J. Sports Med. 2008;42;604-608. *Natural and traumatic sports-related fatalities: a 10-year retrospective study*

UKHSE (1999) *Reducing error and influencing behaviour*. The Stationery Office. Norwich, UK. ISBN 978 0 7176 2452 2. Available at: www.hse.gov.uk/pUbns/priced/hsg48.pdf

UKHSE (2010) *Manufacturer find over former soldier's death*. UK Health and Safety Executive (North West) release HSE/NW/22BCR

UKHSE (Undated) *HSE's Land Use Planning Methodology* (no date). Available at http://www.hse.gov.uk/landuseplanning/methodology.pdf

UNDP and UNICEF (2002). *The Human Consequences of the Chernobyl Nuclear Accident: A Strategy for Recovery*. Retrieved September 2011 from chernobyl.undp.org/english/docs/strategy_for_recovery.pdf

US Bureau of Labor Statistics (2012) *National Census of Fatal Occupational Injuries in 2011*. US Bureau of Labor Statistics report USDL-12-1888. Available at: www.bls.gov/news.release/archives/cfoi_09202012.pdf

US Fire Administration (1987) *College dormitory fire*. US Fire Administration Technical Report Series USFA-TR-006. Available at: https://www.usfa.fema.gov/downloads/pdf/publications/tr-006.pdf

US NRC (2009) *Backgrounder on the Three Mile Island Accident*. Available at: www.nrc.gov/reading-rm/doc-collections/fact-sheets/3mile-isle.html

USA Today (2019) *After eating raw rodent's kidney for good health, couple die of bubonic plague*, USA Today, 8 May 2019

USCG/BOEM (2010). Marine Board of Investigation into the marine casualty, explosion, fire, pollution and sinking of mobile offshore drilling unit Deepwater Horizon, with loss of life in the Gulf of Mexico 21-22 April 2010. Retrieved September 2011 from www.uscg.mil/hq/cg5/cg545/dw/exhib/7-22-10.pdf

Viscusi W. Kip, Magat WA and Huber J (1987). The RAND Journal of Economics, Vol. 18, No. 4 (Winter, 1987), pp. 465-479

WHO (2018) *Global status report on road safety 2018*. World Health Organisation. Available at: www.who.int/publications/i/item/9789241565684

D Biography

Since 1991, the author has helped to implement Reliability-centred Maintenance methods in the defence, chemical, oil and gas, food, manufacturing and other sectors, in locations including the UK, USA, Canada, Europe and the Gulf states.

He was CEO of Information Science Consultants Ltd (ISC), a UK-based organisation which specialises the management of risk and the application of reliability-centred methods. ISC was acquired in 2007 by IFS Defence Ltd and BAe Systems.

Mark holds an MA and PhD from the University of Cambridge and is a Fellow of the Royal Statistical Society, London.