AN EXAMINATION OF THE USE OF DIGITAL COMMUNICATIONS IN SAFETY-RELATED APPLICATIONS

Dr Andrew J White, BSc, PhD, CEng, MIEE
HM Specialist Inspector of Health and Safety (Electrical and Control Systems), Health and Safety Executive Bootle, UK; andrew.white@hse.gsi.gov.uk, 0151 951 3923

Digital communications are rapidly being incorporated into many types of control and safety systems, allowing information to be exchanged between items of equipment and with the outside world. Digital communications facilitate reduced wiring, increased flexibility of equipment positioning and distributed processing, deliver more information about processes, and aid the remote diagnosis of equipment faults. Whilst these attributes can have a positive effect on safety, there are other effects of using digital communications that can seriously, silently, and suddenly reduce safety margins. This paper explores the challenges of designing adequate safety-related systems that utilise digital communications, of assessing the attributes of such systems, and of maintaining their integrity. Topics discussed include ease of connectivity, the erosion of isolation between control and safety systems, managing safety-related systems, and the use of proven-in-use components. The paper refers to the requirements described in the functional safety standards IEC61508 and IEC61511, and suggests some key principles for the application digital communications to safety-related systems.

INTRODUCTION

This paper examines the use of digital communications in safety-related applications.

In the process industry it is common to refer to a safety-related system as a safety instrumented system, and elsewhere the term safety-critical system is used. This paper will used the term 'safety-related' to encompass all these.

Digital communications are being increasingly used in safety-related systems to transmit both safety-related and non safety-related information. Many component suppliers have incorporated digital communication capabilities in their products, with some products having multiple communication ports and the ability to communicate using a wide variety of protocols.

There are many communication devices in common use, some of which are general purpose and others that are designed for specific applications such as high-speed control or the transfer of data to and from remote input and output systems. Many communication systems have been developed to meet the needs of Information Technology (IT) systems, and are being applied to engineering systems due to their ease of use, low cost, and ability to connect a wide variety of devices.

This paper does not attempt to compare the capabilities of individual communication devices or their suitability for specific applications (the British Standards Institution publication 'Selecting the best fieldbus for your application' [1] may help with this), but it does aim to describe the challenges that digital communications can present when they are used in safety-related systems.

The paper describes the use of communication systems that are specifically designed for safety-related applications, and those that are not. It covers the practical difficulties that may be faced in the design of such systems, their use, and their maintenance. The paper describes the potential for digital communications to compromise safety functions and how the security of safety-related systems can be affected by the connectivity offered by communication systems, particularly where there is the risk of malicious attack.

BACKGROUND

The Health and Safety at Work etc. Act 1974 [2] places specific legal duties on those who control work activities. Section 2 of the Act states "It shall be the duty of every employer to ensure, so far as is reasonably practicable, the health, safety and welfare of all his employees . . ." including ". . . the provision and maintenance of plant and systems of work that are, so far as is reasonably practicable, safe and without risks to health". Section 3 of the same Act places a duty on an employer (including the self employed) to "ensure, so far as is reasonably practicable, that persons not in his employment who may be affected thereby are not thereby exposed to risks to their health and safety".

Section 6 of the same Act places a duty on designers, manufacturers, installers, importers or suppliers of articles for use at work to ensure that they are safe and without risk to health. Those involved in the design or integration of a safety-related system should take reasonably practicable measures to ensure the safety-related system does not expose people to undue risks to their health and safety.

The Consumer Protection Act 1987 [3] gives people injured by defective products the right to sue the producer of the product for damages without having to prove that the manufacturer was negligent. A defective product is one where the safety of the product is not such as persons generally are entitled to expect.

Employers and designers need to be sure they possess the competencies necessary to design a safety-related system that meets all legal requirements.

A safety-related system performs one or more safety functions which can each be implemented using an electrical, electronic or mechanical system, other means, or by a mixture of these. Failure of a safety function significantly increases the likelihood of injury or harm to the health of people. Where it carries safety-related information, a digital communication system may contribute to the operation of at least one safety function.

The action of all safety functions together should meet legal requirements, and individual safety functions need to be dependable enough to do the job required of them.

Simple non-programmable safety-related systems generally have well defined failure modes. The presence of a failure is often relatively easy to recognise by plant operators and to diagnose and repair by a maintainer without specialist knowledge or special tools. Programmable safety-related systems can, in certain circumstances, replace non-programmable ones and can provide additional features that may be beneficial. Programmable safety-related systems with communications can provide even greater functionality, reduce wiring, and simplify maintenance, although a great deal of attention needs to be paid to their design and use so that these benefits are realised and not outweighed by the difficulties of employing a complex system in safety-related applications.

SAFETY-RELATED SYSTEM DESIGN – GENERAL

The IEC 61508 [4] family of functional safety standards provides general guidance on how to design a safety-related system in any industry. BS IEC 61511 [5] describes how the principles of IEC 61508 can be applied to process plant, and BS EN 62061 [6] does the same for the machinery sector. These standards describe a safety lifecycle that, if adhered to, can provide confidence that all elements of the design process are adequately covered and that the safety-related system will work as the designer intended when it is called upon to do so. BS EN 50159 (IEC 62280) [7] provides general guidance on the use of communications in railway systems. The principles described in BS EN 50159 have been applied to a number of safety-related digital communication systems.

It is better to eliminate the hazards associated with a process if possible. This is often cost-effective, reduces design work, saves time, and leads to a safer system.

Functional safety standards require that each safety function be allocated a Safety Integrity Level (SIL) according to the level of risk reduction that has been attributed to it during the safety requirements allocation stage of the safety lifecycle. There are a number of ways to select a suitable SIL for a safety function including numerical calculation and qualitative estimation. A list of techniques is presented in IEC 61508.

The failure modes of all parts of a safety-related system should be assessed for their impact on the level of risk control provided by the system. Communication systems require correct interaction between many items of hardware and software (which themselves may fail to work correctly).

It should be noted that any safety-related system should be tested rigorously before it enters service. Theoretical performance alone does not guarantee the suitability of a safety-related system for its purpose.

Continuous and demand mode operation

A safety-related system can be designed to operate in either continuous or demand mode. Continuous mode safety functions are always acting to maintain an acceptable level of risk during normal operation. A dangerous failure of a continuous mode safety function will always result in a hazard or a demand on another safety function, and so it is common for additional layers of protection to be used. Demand mode safety functions take action only when a demand is placed upon them, with the aim of transferring the process to a specified safe state. They should have no effect upon the process until a demand is made.

Process safety time

To be effective, demand mode safety-related systems must be able to take action quickly enough to prevent intolerable risks to individuals or society. The process safety time is the time period between a failure occurring in the process or the basic process control system (with the potential to give rise to a hazardous event) and the occurrence of the hazardous event if the safety function is not performed [8]. The process safety time includes sensing delays, sensor data transmission time, decision-making, actuator command transmission time, actuator operation time, and delays in the effect of the actuator on the process. It is very important to note that delays in the transmission of safety-related information, such as those caused by communication errors and recovery, may significantly increase the response time of a safety function. The safety function must be able to act in significantly less than the process safety time even under fault conditions, or another safety measure should be available to take on this role.

Safety actions

Most demand mode safety actions involve the removal of power from a field device such as a valve or contactor, so allowing a simple mechanism such as a spring to open or close the valve, or open the contactor. Some safety-related systems may command outputs to a particular state to perform a safety action, and these functions should have a suitable safety integrity.

Many safety-related systems that implement demand mode safety functions utilise a relatively simple hardware mechanism called a watchdog to trigger a safety action in the event of a detected failure. It is essential that the watchdog and failure detection mechanism have a suitable integrity for the safety function being implemented by the device. However, some devices that incorporate digital communications allow the watchdog to be configured so that it is suitable for the application. This configuration may include the time before a safety action is taken, what event should lead to the triggering of the safety action, how many communication failures are allowed before a safety action is taken, or disabling of the safety action altogether. Clearly there is the potential for incorrect configuration to lead to a failure of the device to trigger a safety action within the time allocated for it. Section 9 on configuration describes some of the challenges posed by the correct configuration of safety actions.

Proof testing

It is essential that devices performing a safety action are proof tested regularly enough that there is little chance they will fail when a demand is placed on them. This may be difficult to accomplish as a complete test may trigger a safety action that is inconvenient or which increases the risks posed by a process. Thus proof testing may not be able to fully test the safety system, and other measures may be required to reduce the risk that untested devices will cause a failure of the safety function.

Communication devices may be continuously proof tested by the sending of test messages, but this may not exercise all parts of the communication network, or the devices connected to the network, and some faults may not be identified or correctly reported.

Sometimes communication systems are designed with a redundant or backup path that should be tested to ensure that it could transmit the quantity and type of data that may be required in the event of a failure of the primary system. It is also essential that the secondary communication path is triggered to act when it is required, and that data is not lost, replicated, or buffered during the switch from one path to another.

Diagnostic coverage

If faults that could affect the integrity of the safety-related system cannot be identified by the failure detection mechanism, there is the potential for a dangerous failure to go unrecognised, and for the safety-related system to fail to respond to a subsequent demand. IEC 61508 part 2, annex C described a process by which diagnostic coverage can be identified. However, users of communication systems may not know what can go wrong with the communication system, and how it can be detected, and so may be unable to identify the diagnostic coverage. Manufacturers of communication systems that are suitable for use in safety-related systems should be able to provide information on how the communication system must be used, in order to achieve the desired safety integrity.

It is often possible for a high integrity communication system to be used in a way that contravenes the manufacturers instructions, and significantly reduces the integrity of the safety-related system using it.

Even if faults that have occurred in a safety-related system have been detected, these may not be reported correctly or in a timely manner due to communication problems. This could lead to the safety-related system failing to act, or for process operators to be unable to identify why a plant shutdown has occurred. Such events can erode confidence in the safety system and could lead to the system being disabled or permanently overridden.

COMPLEXITY

Whilst digital communication systems can confer many benefits on an electronic control system, such as reduced wiring complexity, more flexibility in where items of equipment can be placed, the ability to have subsystems with their own autonomous control, and provide more information about the status of the process and the devices controlling and monitoring that process, they also result in increased complexity.

This increase in complexity results from the amount of hardware and software required by a communications system to operate, and the potential for there to be complex interaction between items of equipment. This complexity is of a different nature to the complexity of traditional wiring systems, and different skills are required to design, operate, maintain and modify such systems. Organisations without knowledge of the challenges posed by communication systems either have to rely on technical partners, or develop the knowledge in-house. The former can result in an organisation being locked-in to a particular equipment supplier, whilst the latter can take time to achieve and effort to maintain.

HARDWARE

The complexity of communication device hardware varies widely. Most communication devices incorporate a microprocessor, although some use field programmable gate arrays (fpga's) or proprietary integrated circuits. Some devices feature multiple processors and highly complex hardware.

All devices that incorporate communications capabilities have a means of converting between the transmitting media and internal voltage levels. There may be additional circuits that allow the device to store up communication messages, store configurations, detect power supply fluctuations, and detect failures.

Some devices could disrupt the communication of other devices on the network if there is an internal fault and may require hardware to isolate them from the network when this event is detected. This mechanism should have an integrity that meets the requirements of the most demanding safety function that is using the communication network.

The hardware will have limitations in how it can be used. This will include temperature specification, power supply voltages, media voltage range, etc, but may also contain limitations on such things as operating speeds when used in particular modes.

It is unlikely that the complex hardware in a communications device is completely free of design faults that could lead to a failure or unexpected behaviour. However, devices that have been designed and tested to a suitable design standard (such as IEC 61508 for functional safety) are more likely to have had potential failure modes identified, analysed, and eliminated, or measures put in place to reduce their impact.

SOFTWARE

Many communication devices contain significant amounts of software. Some of this is required to implement the communications protocol, whilst other software may provide other application functions such as sensing, calibration, display of information, etc.

It is very unlikely that the software within a communication device is completely free of design faults. A significant period of service in one communication environment will not necessarily mean that all parts of the software have been exercised. Changes in the configuration of the device, the type of information carried and loading variations could reveal design flaws that were previously hidden.

It should be particularly noted that the nature of the communication environment is likely to change significantly if devices are exposed to public communication networks. Communication devices that have worked reliably for years in a private network may very well fail when they are exposed to a greater diversity of communication traffic and specific challenges that exist on a public network (see section 7 on security).

As with hardware, software that has been designed in accordance with a suitable design standard is likely to contain fewer design flaws that could result in a failure of the communication device. Standards such as IEC 61508 (part 3) describe what programming techniques and measures are recommended for different SIL targets, including the use of limited variability programming languages, use of recovery blocks, and defensive programming techniques.

Application software should be able to identify communication faults and failures, and take appropriate action, or trigger a safety action, all with the necessary integrity.

COMPLEX INTERACTION BETWEEN DEVICES THAT USE COMMUNICATION SYSTEMS

Because devices that are connected by a communication system are able to exchange significant amounts of information, there is the potential for faults or failures in one system to propagate to another through data transfer. This could lead to a failure of the safety-related systems immediately, or months or years after the data transfer occurred.

Also where two or more items of plant are operating in a coordinated way, there is the potential for the communication to get out of step. This may be exacerbated where process operators are performing direct control and are expecting a timely response. A communication failure, even if it is only temporary, may result in a change in response time that may prevent the operator achieving the required degree of control.

Particular problems may occur where a safety-related system has been subjected to an upset such as a power supply failure, emergency stop, or an abnormal start-up or shutdown. Such events have the potential to cause large numbers of messages to be transmitted onto the communication network. Some of these messages may be repeating, or may be stimulated by the response of other parts of the system. It could be some significant time before the safety-related system reaches a stable state again and is ready to perform as designed.

Complex interaction between different parts of a safety-related system may make it difficult for process operators to understand what is happening when a process deviation occurs. Potentially large amount of information can be presented to the process operator, but this may make no sense if the operator does not have a clear mental model of how the system is supposed to work and how changes to one part of a system can affect others.

SAFETY-RELATED AND NON SAFETY-RELATED SYSTEM SEPARATION

Some parts of a system (e.g. communications) may be regarded as non safety-related if failure (or malfunction) of those parts has no impact on the likelihood of a safety function acting correctly. However, in reality, a wide range of external events can have an impact on a safety-related system. For example, the modification of a control system may result in a dramatic increase in the demand rate on the safety-related system, or type of demand for which the safety-related is not designed.

Where there is poor separation between safety-related parts and non safety-related parts of a system, then every part of that system should be treated as safety-related. Any modification without suitable safety analyses during installation, commissioning and maintenance could result in incorrect operation of safety functions. Separation may be physical where different components are used for safety-related and non safety-related functions, or functional where some of the same components are used for both, but where there are additional measures to protect the safety-related path from the effects of common cause failures.

Strongly separating the safety-related and non safety-related parts of a system allows greater effort to be focused on the safety-related parts and also means that there is less likelihood that a single event will cause a failure of both the control and safety systems.

In process industries it is common for a safety-related system to be one of a number of different layers of protection, each of which limits the potential for, and impact of, a hazardous event. The strengthening or weakening of a layer of protection adjacent to the safety-related system layer requires the capability of the safety-related system to be re-assessed for its suitability.

Where safety-related and non safety-related systems share resources such as sensors, power supplies, data storage devices, cabling, etc, care needs to be taken to ensure that a single failure does not affect both the non safety-related and safety-related systems. Similarly, if redundant safety-related systems are used, it is important that common cause failures do not affect both primary and secondary systems at the same time.

CREEPING CONNECTIVITY

A significant problem with safety-related communication systems is ‘creeping connectivity’. This is where safety-related information is mixed with non safety-related information by a communication network connection, or within a device that is connected to both the safety-related and non safety-related communication networks, generally with the best of intentions such as providing additional process information to operators or managers. This step may be well planned, and the potential for interaction between the safety-related system and non safety-related system may be recognised and minimised. However, it is common for business information networks to be regularly modified to meet the needs of the business. Because these networks are not safety-related, little or no effort may be employed to identify if the changes are likely to lead to a safety implication. Thus it may be that the business information network may be connected to a wide range of plant, may be used to carry a wide variety of data, and may even be connected to external networks. This provides the potential for external events to interfere with the safety-related information.

An organisation may expend significant resources on maintaining good internal control of its communication systems, but may have difficulty controlling the activities of individual employees or contractors, even if strict working practices are agreed. This may result in inadvertent connection to the internet with little or no firewall protection.

The dangers of creeping connectivity is one of the reasons why it is better for safety systems to remain isolated from business information networks.

USING ADDITIONAL INFORMATION IN A SAFETY-RELATED SYSTEM

A feature of using communication systems is that additional devices can usually be quickly and easily added to provide more information about the process. Sometimes a large amount of additional information can become available just by making a single communication link.

However, it is good practice for safety-related systems to be as simple as possible and it is important that the information being used by a safety function has adequate safety integrity. Use of information provided by a low integrity sensor, or derived from a number of low integrity sensors, for a safety function requires that the sensor(s) should be regarded as safety-related. Information from a sensor that has a quoted safety integrity, but with is passed through a low integrity communication device may not have adequate integrity for a safety function. This places all sorts of demands on the way that the sensor(s) and communication devices should be specified, commissioned, operated, maintained and modified, and also how higher integrity information should be transmitted from place to place.

It may be identified that there are specific safety benefits to be gained from providing additional information about the process. If these benefits outweigh the potential disadvantages of the additional connectivity on safety grounds, then a strong argument may be supportable. However, it is important to remember that the safety-related system must be protected from challenges to its correct operation throughout the life of the system. When more devices are used to perform safety functions, it will become more difficult to ensure that inadvertent or unauthorised changes are not made to these.

When changes are made to safety-related devices, the potential effect on the safety-related system should be identified by someone competent to do so. Similarly where there is a failure of a safety-related component or equipment, significantly more work may be required to ensure a replacement is suitable, is correctly configured and is adequately tested.

Having a larger number of safety-related devices connected to the safety system can have a significant impact on the amount of work required to operate, maintain, and perform future modifications on the system. This longer-term impact should be considered before connecting too many additional devices to the safety-related system.

REMOTE DIAGNOSIS OF EQUIPMENT

It is common for modern equipment to have the facility to be remotely interrogated so that failures or process problems can be identified. The manufacturer of the equipment may stipulate this to ensure service and efficiency contractual agreements are met, and this generally requires the equipment to be connected to a public or business network.

Unless there is strong protection within the equipment, there is the potential for remote interrogation to result in inadvertent, unauthorised or malicious changes to be made. Changes in state within the equipment could have an impact on equipment safety, and even the action of interrogating equipment may be disruptive enough to prevent normal operation.

For these reasons it is better for equipment only to be connected to the remote interrogation network when this is essential, and for the functionality of equipment to be fully tested after the interrogation has been completed.

It is important that software being used to perform an interrogation has adequate integrity and that measures are put into place to prevent inadvertent changes to the equipment being interrogated.

COMPETENCE

Safety-related systems that use communications should be designed to an appropriate standard such as IEC 61508 by personnel competent to do so. The additional challenges presented by the use of communications should be considered such that design, commissioning, operational and maintenance staff are able to attain the competence necessary to operate the safety-related system.

Where communication systems are used to transmit safety-related information, these should be maintained by staff who are competent to do such work, and who can identify potential threats to the correct operation of the system. Staff should be competent to perform any proof testing, according to the schedule produced by the system designer, and to identify actions that need to be taken on indication of a failure or problem.

SECURITY

Where access can be gained to communication networks, and the devices connected to them, damage may be caused inadvertently, or deliberately. Thus security is a significant issue where parts of a safety-related system are connected by a communication system.

Traditional methods of security such as preventing unauthorised staff from physically accessing safety-related plant may not be effective, as damage could be caused by simply connecting into a safety-related communication system, and the safety-related system may be spread over a larger physical area.

Where the extent of a communication system is regularly changed, and different devices are connected to it, there is the potential for new challenges to arise that only become evident when there is a problem. For this reason communication systems carrying safety-related information should, if possible, remain fixed in size, and with the same devices connected to them.

Safety-related devices connected to public networks may not be able to withstand the types of challenges present on these. The changing nature of electronic attack means that even if a device had adequate defences against electronic attack when it was installed, the challenge may have changed to the extent that the operation of the device may be disrupted. It should be noted that the nature of electronic attack can change very quickly, and that access to vulnerable networks has to be continuously managed. This may require a significant amount of resource.

FIREWALLS

Firewalls may be used in an effort to protect the devices on a communication network from disruption caused by external events, usually where there is a connection to a public network such as the internet.

Unfortunately a firewall has two requirements that potentially conflict with each other; that of speeding the transmission of legitimate communications traffic between the private and public network, and that of accurately identifying and preventing communications that may not be authorised, or which could result in damage to the private network, or devices connected to it.

The fast-changing nature of the threat posed by traffic on a public network means that firewalls have to be regularly monitored for potential breaches and upgraded to ensure protection is adequate. Some challenges may only become evident after they have caused damage to a system.

NISCC (National Infrastructure Security Co-ordination Centre) has produced a report that describes the challenges posed by communication systems on the control of industrial plant [9].

COMMON CAUSE FAILURES

Where many devices of a similar or identical design are used to perform control and safety functions there is the potential for a single failure to affect more than one device, possibly leading to the loss of control and safety systems at the same time. Identical communication devices used for control and safety systems could be vulnerable to particular communication messages, network loadings, or other factor that could cause simultaneous failures.

Where a communication network is extensive, a loss of power to a part of that network may result in a deluge of messages that may prevent communication to safety devices, or prevent process operators from viewing the plant status.

CONFIGURATION

Many devices that can connect to a communications system are highly configurable. This configurability gives the system designer flexibility to use the device in a number of different ways. Some devices may allow selection of measurement units, alarm settings, action on fault, scaling values, offset values, to name but a few. Where such a device is used to perform a safety function, an incorrect configuration could lead to a failure of the safety-related system to detect a demand, or for the wrong action to be taken, either by the device itself, or by a process operator who is using feedback from the device to decide which action to take. Even if more than one identical device is used for redundancy, the use of a faulty common configuration for several devices could lead to a failure to take the desired safety action.

It will generally be necessary for additional actions to be performed to confirm that the configuration is correct, such as reading back the programmed values, and checking the device is behaving as anticipated.

Where the configuration of a device is critical to the correct functioning of a safety device, measures need to be put in place to prevent inadvertent or unauthorised changes to the configuration. According to the extent of the communication network attached to the device, configuration may be done remotely significant distances from the physical location of the device. As a consequence traditional physical access controls such as locked cabinets and authorised areas may not be effective. Equally, as a wide variety of programming devices including desktop computers, laptop computers and even handheld computers may be used to make the configuration changes, there should be defences of sufficient integrity to prevent unwanted re-configuration. Methods such as the use of passwords, physical keys, or software keys need to have a suitable level of integrity that matches that required by the safety function using the configurable device.

AUTOMATED CONFIGURATION SYSTEMS

Some integrated systems provide the facility for the control system to identify when configuration changes have been made, possibly as a result of inadvertent re-configuration or as a result of a component being replaced due to a failure, and for the correct configuration to be downloaded. Where an automated system such as this is used, care must be taken to ensure that the automated configuration system has a suitable integrity, and that legitimate configuration changes are not overwritten by the automated system. Equally, data stored in a device that may be necessary for correct start-up may have been lost, necessitating a commissioning procedure to be undertaken.

PROOF TESTING, MAINTENANCE AND REPLACEMENT

Many field devices require regular proof testing and maintenance, particularly in safety-related applications. This may require the device to be placed in a test mode, or for the configuration to be temporarily changed. Sometimes this is done locally, and sometimes by remote means. It is important that the personnel who perform the tests and/or maintenance understands the potentially complex tests that may need to be carried out, and that they are able to restore the device to its normal settings. Mistakes can be made where a device has a number of very similar operational modes, or where an old device is replaced by a new one and the configuration has to be copied across by hand, or via an electronic system.

Sometimes devices need replacing. However, the new device may have different software or hardware which means that it will perform differently from the original. The previous configuration may not work in the same way, and simple tests may not reveal the differences. For this reason it is important for the version numbers of devices to be recorded.

Difficulties can arise where there have been changes to specifications and the old devices are no longer available, necessitating a change to a different version or device. In such cases the replacement needs to be treated as a modification, and a safety impact analysis carried out.

Where manufacturers of general purpose communication devices are not used to the demands of safety-related systems they may make changes to a device that could impair its performance, with no external indications being available to the user. The substitution of one component for another may have a significant impact on an attribute of the device that is being used for safety.

MODIFICATION

The configuration of many communication devices can be changed very easily once any security system that prevents it has been overcome. There may be a strong temptation in certain circumstances for process operators to 'tweak' the settings of a sensor to reduce operational problems. Where there are complex interactions between devices this may result in unintended effects that may include unexpected tripping or even failure of a safety function.

Modification to devices that are performing safety functions should only be carried out after an analysis of the impact of the change by personnel with the skills to make such an analysis.

POWER LOSS

Problems can arise where there is a failure of power, and where there is information about the state of a process stored in devices connected to a communication network. This could result in the transmission of old data, or data that is not synchronised with other data. As a consequence the safety-related system may fail to take safety action when required to do so, or to take action when there is no need. Process operators may rapidly lose confidence in a safety-related system that causes unwarranted shut downs and may ignore information or prevent the system from taking safety actions.

DISPLAY AND USE OF DATA

It is common for data received from digital communication systems to be used by process operators to make decisions regarding the best action to take in the event of a failure that could lead to a hazardous event. Such systems allow the flexibility of human intervention to provide a potentially better response than that which may be achieved by automated systems.

However, the information that is presented to operators may arrive from many different locations, via a variety of different communication paths, and pass through a number of different devices before it is displayed. There is the risk that some data may take considerably longer to make the journey from the measurement point than other data, particularly under fault conditions where communication networks may be more heavily loaded than usual. Where plant conditions are changing rapidly (again, likely under fault conditions), the operator may be presented with information that is out of step in time, and which may appear contradictory.

An extreme case of information delay is where there is a complete loss of communication with some parts of the plant that may lead to some data values 'freezing' whilst others continue to reflect measurements made on the plant. Under these conditions it is unlikely a plant operator will be able to identify the cause of problems and may have to take drastic action that may increase the risks posed by the plant to other personnel, and which could have significant financial consequences.

A significant challenge with safety-related systems that incorporate communications is the amount of information that can be used. There is the temptation to provide plant operators with a lot of information simply because it is available. Providing plant operators with too much information may cause confusion, result in poor decision making, and prevent timely decisions from being made. Thus a significant amount of effort needs to be put in only presenting the information needed at the time it is needed.

PROVEN-IN-USE AND PRIOR USE

Where there is a need to incorporate a safety-related device in a design, that device should ideally have been designed according to a suitable functional safety standard (such are IEC 61508). It may be the case that no such device is available. It is also possible that devices with no strong design documentation may have been successfully used for many years.

IEC 61508 (part 2, 7.4.7.5) describes a process whereby the use of a safety-related device may be justified on the basis of it having been 'Proven-in-use'. IEC 61511 uses the term 'Prior use' to mean a similar thing.

A justification to use a component in a safety-related system because it is Proven-in-use, or has Prior use can be very difficult to make as the device should have (IEC 61508) ". . . clearly restricted functionality and . . . adequate documentary evidence which is based on the previous use of a specific configuration . . . (during which time all failures have been formally recorded . . .)". The standard goes on to say that the documentary evidence shall be sufficient to demonstrate that the likelihood of a failure is low enough and that the previous conditions of use are close enough to the intended use that the required safety integrity can be achieved.

IEC 61511 (11.5.3.2) requires the evidence to include:

- consideration of the manufacturer's quality, management and configuration management systems;
- adequate identification and specification of the components or subsystems;
- demonstration of the performance of the components or subsystems in similar operating profiles and physical environments;
- the volume of the operating experience.

IEC61511 goes on to say that users may maintain lists of suitable components. However, the list should be regularly updated, and action taken in the event of a discovery that a particular component has a previously unsuspected failure mode, or other aspect that could compromise its performance.

The wide variety of communication network configurations, type of communication traffic, and variation of that traffic may make it very difficult to show that a particular communication device is indeed 'proven in use' for a specific application. To do this the manufacturer of the device has to know in what conditions each one is operating, and exactly what failures have occurred and for what reason. Many communication network devices are cheap enough to end users to simply throw failed devices away, rather than return them to the manufacturer for analysis and repair. Also, failures that can be fixed by a simple reset, may not even be correctly recorded. These factors may lead to significant under reporting of failures.

Where the device is intended for use in higher safety integrity applications it may be difficult to achieve enough running time in an appropriate environment for a proven in use argument to be legitimately made.

DATA TRANSMISSION MEDIA

There are a number of media types that are be used to transmit data. These include ordinary wires or special cabling, optical fibres, and via a radio signal (often termed wireless).

Radio communications are potentially susceptible to interference from plant equipment, particular meteorological conditions, and outside influences such as mobile telephones and other radio transmitters. Radio communications used for information technology networks may be insecure, allowing unauthorised access to safety-related communication systems.

For these reasons, it is not recommended that radio transmissions be used for safety-related applications, unless the potential for failed communications is worth the safety benefit to be gained, and communication cannot be achieved any other way.

SUMMARY

The ease of use of digital communication systems has increased dramatically over the last few years, with 'plug and play' functionality allowing users to configure a wide variety of communication architectures quickly and easily. It is this ease of use that can lead users to believe that the system is more capable than it actually is, and that a working system will continue to work. Prices continue to fall and capability continues to rise, making the use of digital communication systems a very attractive option when considering the design of new plant or upgrades to older plant.

Users of digital communications in safety-related systems should be aware of the capabilities of these, and take action to ensure that risks are properly controlled. Challenges caused by the intermixing of safety and non safety-related systems, external influences, long term maintenance and operational requirements can all have significant effects on the level of safety achieved, and the resources that have to be deployed to achieve this level of safety.

Suggested key principles for the application of digital communication systems

1. No single failure of a communication system should lead to a hazardous event. This may be the complete failure of a communication channel, or intermittent loss or corruption of information being carried by a communication channel.

2. Keep communication systems as simple as possible, in line with the aim of reducing the risk, so far as is reasonably practicable.
3. Only make connections between safety-related and non safety-related communication systems where detailed analysis shows there to be an overall reduction in risk, and where it can be shown that the risks can be maintained at an adequately low level despite the long term challenges posed by maintenance, modification, security, and human interaction.
4. Information used for safety-related purposes should have sufficient integrity for all safety functions that use that information.
5. Be aware that proven in use arguments for the adequacy of a communication system to carry safety-related information may not be appropriate where the intended usage differs in any way from a successful previous use. This may include the transmission of different data types, message lengths, communication channel loadings, environmental conditions, etc. Proven in use arguments need to be supported by sufficient data on usage conditions and failure rates.
6. Assess the potential for common cause failures to affect more than one part of the communication system at once, and take measures to ensure that the safety-related system is protected from this.
7. Where a safety function makes use of a communication system, there should be diagnostic capability of sufficient integrity to identify failures that may lead to a failure of that safety function to operate when it is needed.
8. Diagnostic checks of the correct functioning of any safety function that incorporates a communication system should be regular and thorough enough to ensure that risks are maintained at an adequately low level. Remote diagnosis should only be used when it is absolutely required.
9. Communication system devices should only be replaced by ones with identical software and hardware version numbers unless analysis has shown that a component of another type or version has the same capability as the component being replaced. Be aware that performance upgrades could reveal previously unanticipated failure modes that could adversely affect the ability of a safety-related system to respond to a demand.
10. Changes to communication system components that contribute to safety should only be performed after an adequately thorough safety impact analysis that considers configuration settings, start-up, operational, fault, shutdown, testing, maintenance, and other appropriate modes.
11. Care should be taken to ensure that the communication system devices are adequately protected against the effects of inadvertent or malicious re-programming, including the effects of power outages and loss of memory.

REFERENCES

1. Selecting the best fieldbus for your application – A guide to the evaluation of industrial fieldbus protocols. British Standards Institution, BIP 0014:2004, ISBN 0 580 44515 1.
2. Health and Safety at Work etc. Act 1974, sections 2, 3 and 6.

3. Consumer Protection Act 1987, http://www.opsi.gov.uk/si/si1987/Uksi_ 19871680_en_1.htm.
4. BS EN (IEC) 61508:2002, Functional safety of electrical/electronic/programmable electronic safety-related systems, parts 1 to 7.
5. BS IEC 61511:2003, Functional safety. Safety instrumented systems for the process industry sector, parts 1 to 3.
6. BS EN 62061:2005, Safety of machinery. Functional safety of safety-related electrical, electronic and programmable electronic control systems.
7. BS EN50159:2001, Railway applications – Communication, signalling and processing systems, part 1 Safety related communication in open transmission systems and part 2 Safety related communication in closed transmission systems.
8. BS IEC 61511-2:2003, 11.9.2, Definition of process safety time.
9. Good practice guide on firewall deployment for SCADA and process control networks: 2005, National Infrastructure Security Co-ordination Centre, http://www.niscc.gov.uk/niscc/docs/re-20050223-00157.pdf.