Inadequate Legal, Regulatory and Technical Guidance for the Forensic Analysis of Cyber-Attacks on Safety-Critical Software

Chris W. Johnson

School of Computing Science, University of Glasgow, Glasgow, Scotland, UK, G12 8RZ.
johnson@dcs.gla.ac.uk; http://www.dcs.gla.ac.uk/~johnsons

Abstract

National and international organisations including NIST and ENISA have published guidance that is intended to help organisations respond to, and recover from, cyber incidents. They provide detailed information about contingency planning, about the processes needed to gather and analyse evidence, about appropriate ways to disseminate the findings from forensic investigations. Legal frameworks, including the Federal Rules of Evidence, also help companies to identify ways of preserving a chain of evidence with the digital data gathered in the aftermath of a cyber-attack. It is essential that companies apply these guidelines to increase their resilience to future attacks. However, they provide the least support where they are needed the most. Existing guidelines focus on corporate office-based systems; they cannot be applied to support companies dealing with cyber-attacks on safety-critical infrastructures. This is an important omission. It is impossible to immediately disconnect infected systems where they provide life-critical functions. There are conflicts between the need, for instance, to preserve the evidence contained in volatile memory and the requirement to return safety-critical applications to a safe state before any forensic work can begin. The following pages identify the problems that arise when applying legal, regulatory and technical guidance to the cyber security of safety-critical applications. The closing sections focus on techniques that can be used to support the forensic analysis of cyber incidents and promote recovery from attacks without placing lives at risk[1].

Keywords: safety-critical systems; contingency planning; cyber-security; forensic analysis; incident response.

1 Introduction

A small but growing number of incidents have affected safety-critical applications, including Air Traffic Management infrastructures, Fire and Rescue dispatch systems, Maritime monitoring applications, power generation and distribution systems (Johnson, 2012a, US Department of Transport 2009). As far as we can tell, these events have not resulted in loss of life. We have been protected by a range of defences including human monitoring; firewalls; software and hardware diversity. However, there are no grounds for complacency (ENISA, 2011, NIST, 2012). The increasing sophistication of malware and our reliance on a small number of common software infrastructures is combining to undermine existing defences. The growing use of Commercial off the Shelf (COTS) software creates new vulnerabilities for safety critical systems (Johnson, 2012). Linux variants have replaced specialist operating systems in many application areas. Similarly, the internet protocol stack is common across many safety-related infrastructures. Satellite Based Augmentation Systems have been certified for use in safety related applications across Europe and North America (Johnson and Atencia-Yepez, 2010). For the first time, this enables the integration of enhanced GPS data into primary systems in air traffic management, maritime surveillance, fire and rescue, command and control etc. The widespread dissemination of technical information about these common infrastructure components increases concerns about the probability and consequences of cyber threats against safety-critical systems.

[1] This is a draft position paper; comments and criticisms that should be sent to Johnson@dcs.gla.ac.uk.

Several organisations, including the US National Institute of Standards and Technology (NIST) and the European Network and Information Security Agency (ENISA) have recently published guidance on leading techniques for the forensic analysis of and recovery from cyber-incidents (NIST, 2012, ENISA, 2011). This paper argues that we must extend existing guidance to provide support that is tailored to the particular needs of safety-critical infrastructures. In more conventional office-based applications it is, typically, possible to disconnect an infected system while a forensic analysis is conducted without corrupting available evidence. However, this is seldom possible in safety-related infrastructures. We cannot disconnect electricity consumers, including healthcare providers, water supply companies, and food distribution organisations, for the many days that might be required in order to complete detailed forensic analyses. In other cases, we will have to balance the risks associated with shutting down a safety-critical system using an infected primary application or risk using back-up systems that might also have been subjected to a cyber-attack. Such risk assessments reflect the complex interactions between safety and security. They are non-trivial because it is difficult to determine the probability of cross-infections between primary and secondary systems; it can also be difficult to quantify the additional risks created by using back-up systems that are typically less familiar and may offer reduced functionality to system operators.

A small number of guidelines already focus on the recovery from cyber-incidents in safety-critical applications. For instance, the US Nuclear Regulatory Commission (2010) has developed Cyber-Security architectures for Nuclear Facilities. Their guidance includes a section on incident response in the context of a safety-critical application. This makes it clear that the operators of nuclear facilities must provide for the timely detection and response to incidents; the mitigation of any future cyber-attacks; the correction of exploited vulnerabilities, and the restoration of systems, networks and equipment affected by an attack. Licensees must also develop a contingency plan that identifies roles and responsibilities during the recovery process. NRC guidance helps to mitigate the impact of cyber-attacks. However, it leaves many questions unanswered. For instance, what processes are needed to gather forensic evidence in the aftermath of a cyber-incident without undermining the safety of operators or the general public? What are the legal implications of cyber-attacks on safety critical infrastructures? Can networks and systems be regarded as a crime scene, following an incident involving national critical infrastructures? How can we facilitate information sharing about the causes, consequences and recovery actions from previous incidents in order to mitigate the risks of future attacks on safety-critical systems?

2. International Frameworks for the Recovery from, and Reporting of, Cyber-Incidents

This paper focuses on the tensions that exist between the need to maintain safety and at the same time coordinate the recovery from cyber incidents. The US Federal Information Security Management Act (2002) provides the wider context for these recovery actions. Compliance with the provisions of FISMA, involves a number of key activities that are summarised as follows:

1. Identify and categorize the information that is to be protected;
2. Select minimum controls to secure that information and document in a security policy;
3. Refine controls using a risk assessment procedure to assess the probability and consequence of loss;
4. Document the controls in a system security plan that implements the policy in item 2;
5. Implement security controls in appropriate information systems to mitigate risks;
6. Assess the effectiveness of the security controls once they have been implemented;
7. Determine any residual agency-level risk to the mission or business case;
8. Authorize the information system for operation and maintain controls;
9. Monitor and audit the security controls on a continuous basis.

FISMA requires that agencies have “procedures for detecting, reporting, and responding to security incidents”. These procedures gather evidence and document the causes of an attack or violation; they also help to exchange lessons learned in the aftermath of a incident. The Act provides a generic framework that does not directly

address the potential impact of a cyber-incident on public safety. The risk-based approach that is embedded within FISMA is entirely focused on the development of cost-effective strategies for cyber-security rather than the reduction of safety-related hazards to the users and operators of complex systems.

The US National Institute of Standards and Technology (NIST) coordinate the technical implementation of FISMA. Figure 1 provides an overview of the key processes in NIST's (2012) incident management strategy. The following pages focus on the challenges that arise when implementing the NIST lifecycle for safety-critical infrastructures. Companies and regulatory organisations must be provided with techniques that aid containment, eradication and recovery without posing increased risks to the users or operators of safety-related applications.

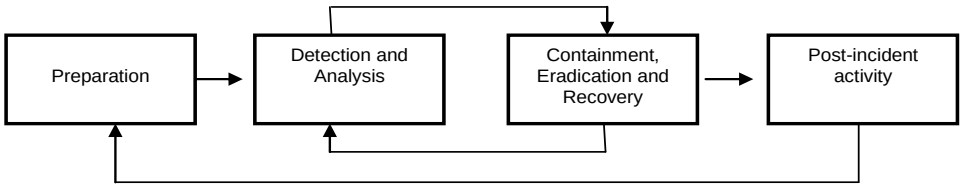

Figure 1: NIST Incident Response Lifecycle

In Europe, the legislative context is set by EU Directive 2009/140/EC. A key element within the directive has become known as 'Article 13a' on the security and integrity of public communication networks. Paragraphs 1 and 2 of Article 13a require service providers to ensure the security and integrity of their networks and to ensure continuity of service. Paragraph 3 requires that service providers report significant security breaches and losses of integrity to national regulatory agencies. They must then forward summaries to the European Network and Information Security Agency (ENISA). EU Directive 2009/140/EC focusses on the resilience of operators irrespective of whether their services are being used in safety-critical infrastructures or in mass market applications. However, the European Commission has recently proposed the extension of obligatory reporting requirements as part of the European Union's 2013 Cyber-Security Strategy[2]. This initiative increases the significance of our work; we must develop exiting guidance to support the recovery and reporting of cyber-incidents across transportation, healthcare, food and water supply, power generation and distribution etc.

3. Limitations of Cyber-Incident Guidelines in Safety-Critical Infrastructures

The NIST guidelines support many different application areas. In consequence, they are drafted at a relatively high-level of abstraction. There is often an implicit focus on more conventional information processing applications:

> "Incidents targeting IT systems typically impact the business functionality that those systems provide, resulting in some type of negative impact to the users of those systems. Incident handlers should consider how the incident will impact the existing functionality of the affected systems. Incident handlers should consider not only the current functional impact of the incident, but also the likely future functional impact of the incident if it is not immediately contained. Incidents may affect the confidentiality, integrity, and availability of the organization's information. For example, a malicious agent may exfiltrate sensitive information. Incident handlers should consider how this information exfiltration will impact the organization's overall mission. An incident that results in the exfiltration of sensitive information may also affect other organizations if any of the data pertained to a partner organization". (NIST, 2012)

[2] Article 34 of the strategy explores the extension of security breach notification provisions.

In safety-related systems, the business impact is often secondary to the consequences for what safety standards such as IEC 61508 refer to as 'equipment under control'. The exfiltration of data is, typically, less of a concern than the safety of application processes. For instance, altering the configuration of surveillance hardware can have a disastrous impact on air traffic management infrastructures at hub airports. Small changes in electronic healthcare records not only create administrative concerns but also have the potential to result in iatrogenic injuries. Not only do cyber-incidents create direct safety concerns, there are secondary effects. For example, by undermining existing safety arguments, malware invalidates the regulatory approval that is a prerequisite for the operation of most critical infrastructures. Cyber-attacks invalidate assumptions about processor, memory and network utilisation. It is hard to guarantee that a safety-critical process will always have the resources that it requires if a system has been compromised by code with unknown provenance. Existing NIST guidance provides general recommendations on the recovery process. It does not consider how organisations can ensure that their networks are sufficiently secure for regulators to permit return to service:

> "The size of the incident and the type of resources it affects will determine the amount of time and resources that must be spent on recovering from that incident. In some instances it is not possible to recover from an incident (e.g., if the confidentiality of sensitive information has been compromised) and it would not make sense to spend limited resources on an elongated incident handling cycle, unless that effort was directed at ensuring that a similar incident did not occur in the future. In other cases, an incident may require far more resources to handle than what an organization has available. Incident handlers should consider the effort necessary to actually recover from an incident and carefully weigh that against the value the recovery effort will create and any requirements related to incident handling". (NIST, 2012)

The recovery from a cyber-incident is further complicated by an immediate need to preserve the safety of end-users and operators. There are a host of detailed technical concerns that have yet to be addressed in any sustained manner. For instance, diverse, secondary systems can be used to recover in the immediate aftermath of an incident, especially if the secondary systems are immune from the source of an initial attack. It is for this reason that some safety-critical organisations have developed fall-back applications running under MS Windows™ when the primary infrastructures rely on Linux variants. However, this additional assurance must be balanced against the costs associated with two redundant infrastructures. The relatively low likelihood of cyber-incidents must be offset against an increased probability of routine design, implementation and maintenance flaws when development budgets support two different versions of the same application.

The European Network and Information Security Agency (2009) has provided similar guidance to that published by NIST. It is generic and can support a wide range of different industries but there are problems in adapting it to support the recovery from cyber-incidents that involve safety critical systems. Rather than reiterate the weaknesses that were identified in previous paragraphs, the remainder of this section identifies the benefits from developing specific guidance for transportation, healthcare, food and water supply, power generation and distribution etc. ENISA (2011) distinguishes between three different types of cyber-incident reporting systems, although most schemes cover two or more of these objectives:

- *Incident reporting systems that support emergency response.*
  These systems support real-time information sharing and coordination during emergency situations. There are obvious benefits if these systems can be extended to a wider class of safety-related applications. For instance, many different Air Traffic Management, National Emergency and Rescue coordination centres, maritime monitoring systems and some military applications all now operate very similar VOIP communications infrastructures. A breach on one end user of these technologies might have very similar effects on communications in very different safety-related applications. ENISA make it clear that the organizers of future eCommunications reporting schemes will have to cooperate with national crisis management centres and bring in representatives of other sectors to support crisis management. At present there is little or no coordination to ensure that information about an on-going

cyber-incident involving one safety-critical industry is communicated to the users of similar software infrastructures in another safety-related area;

- *Incident reporting systems that support incident prevention.*
  Rather than focusing on real time dissemination about on-going events, incident reporting systems can provide information on how to prevent the recurrence of earlier violations. Sector-wide information about previous threats guides security risk management. It is important not to underestimate the potential benefits of such incident reporting systems. There are elaborate schemes for the dissemination of safety information, including NASA's Aviation Safety Reporting System (ASRS) or the UK Confidential Human Factors Incident Reporting Programme (CHIRP). However, there are no similar systems for the exchange of information about cyber-security threats to safety-critical industries (Johnson, 2012a). In contrast, many companies remain confused about their reporting obligations. For instance, the UK operates a Mandatory Occurrence Reporting Scheme. This "is intended to record reportable occurrences which endangered or which, if not corrected, would have endangered an aircraft, its occupants or any other person". The UK regulator interprets this to include information about a broad range of cyber-threats. However, there are very few (no?) examples of security violations being reported through this scheme;

- *Incident reporting systems that support legal actions.*
  Many companies only invest in reporting systems to meet legal and regulatory obligations. These systems are often starved of investment and lack management support. Recommendations fail to prevent the recurrence of an adverse event. There is a danger that insufficient information will be obtained about the causes of an incident for other operators to benefit from participation in the scheme. Further problems complicate security incident reporting in safety-related applications where regulators often lack the audit mechanisms to verify that companies are meeting their legal obligations to report cyber-incidents.

The reminder of this paper focuses on the forensic analysis of cyber-incidents in safety-critical infrastructures. This is justified because incident reporting systems offer limited benefits if organisations cannot accurately identify the causes and consequences of security violations.

4. Triggering the Forensic Analysis of Cyber-incidents in Safety-Critical Industries

A number of mechanisms can trigger the forensic analysis of cyber-incidents involving safety-critical applications. NIST (2006) advocate the use of several different software systems to automatically detect the "precursors and indicators" to an incident. The use of several different services can increase situation awareness and provide warnings about multiple threats. However, the NIST advice to exploit diverse monitoring systems can undermine safety-critical software engineering. In order to obtain regulatory approval for the installation of software into safety-related systems, companies must demonstrate the reliability of their code within its intended context of use. Each additional software service incurs significant costs that far exceed those for most other application areas. Companies must still show that intrusion detection and prevention systems, antivirus software, and file integrity software do not contain routine bugs that might threaten safety. This is particularly important for defensive applications that use external servers to periodically update malware definitions etc. In such circumstances, safety engineers would continually be engaged in a test and re-test cycle to ensure that new versions of protection and detection systems could safely be integrated into critical operating environments.

A number of further problems complicate the automated detection of cyber-incidents in safety-critical applications. It is important to profile 'normal behaviour' so that deviations can be reported. A deep knowledge

of normal operation can be gained by reviewing logs and through the routine analysis of system behaviour. However, many safety-critical systems do not routinely have the level of monitoring implemented, for example by financial institutions. Networks that have experienced few operational problems will often not be analysed to any significant extent. There are numerous reasons for this. The most obvious is that safety-related engineering is guided by risk-based techniques – resources are focussed on those applications that are most likely to have a significant impact on safe and successful operation. Attention tends to focus on those areas that cause the greatest problems for operations rather than on areas that might be most vulnerable to cyber-attacks. Many companies also question the need to maintain logs which are very unlikely to be used given the relatively low reported frequency of cyber incidents (Johnson, 2012). There is also a justified concern that the introduction of additional audit mechanisms will increase complexity and might undermine the resilience of safety-critical systems.

Sub-contractors can also submit cyber-incident reports. An attack on their infrastructure can be propagated to the companies that employ them. Subcontractors play an increasingly important role across many safety-critical industries. Few service providers have the technical capacity to maintain the growing array of components required within SESAR or the European Train Control Systems. Unfortunately, there are many barriers to the vertical reporting of security violations up the supply chain. Contracting companies have significant concerns about the legal and commercial implications of admitting cyber incidents on their future business. In other areas, Cloud based architectures offer the benefits of virtualisation, especially where safety related processes rely on large amounts of less critical operational data. In such circumstances, it is hard for end users who experience the safety-related consequences of a security breach to trace the technical causes of particular violations. Both ENISA (2009a) and NIST (2011) provide valuable guidance on how to deal with these issues, neither considers the role of sub-contractors and Cloud service providers in gathering evidence about cyber incidents that involve safety-critical applications. This is an important omission; lives may depend on the timely provision of information about the scope and extent of any violation.

5. Immediate Actions and Incident Containment

Earlier sections argued that the difficulty of certification has limited the introduction of automated intrusion detection systems and of anti-viral products within many safety-critical environments. There is also a culture of coping with degraded modes of operation (Johnson, Kirwan and Licu, 2009). In consequence, engineering teams will try to use a range of ad hoc 'solutions' rather than diagnose security violations (US General Accounting Office, 1998). These problems are compounded by the barriers to cyber- incident reporting in safety-critical industries. Cultural, commercial, legal and regulatory problems make it difficult to exchange lessons about previous security violations. In consequence, engineers and managers face enormous uncertainty when cyber-incidents are detected in complex, safety critical systems:

- *Who to notify?*
  It is important to inform a range of internal stakeholders. Conventionally these include the chief information officer, the head of information security, local security officers, other incident response teams within the organization, sub-contractors and system owners. Additionally, there is a requirement to inform safety management and the head of operations. This is important because security engineers often lack the technical insights necessary to understand the impact of particular violations on operational practices; for example it takes years of experience to fully understand the ways in which air space are operated. Conversely, operational staff find it difficult to understand the implications of cyber-incidents when individuals do not have a detailed understanding of software engineering. It is also important to communicate with external agencies, including regulators and national security agencies. Similar communications issues arise. For example, it is impossible to provide a complete guarantee that systems are free from infection. This follows Dijkstra's maxim that testing can provide the presence of

bugs and not their absence. By analogy, anti-viral systems will prove the presence of malware but cannot establish their absence. Most regulatory agencies are struggling to retain key staff as the fiscal constraints of an economic downturn limit their ability to recruit well-trained engineers. In such circumstance, it is vital to discuss the regulatory response to a cyber-incident before any adverse event occurs. This helps to ensure that regulators have competent staff who are sufficiently well trained to provide suitable guidance to companies during the recovery process;

- *What systems are affected?*
  In order to contain an incident, it is important to determine the extent of a security violation. This is non-trivial even when infections seem to be isolated within corporate systems rather than primary control systems. Many safety-critical organisations have adopted architectures that separate office systems from those applications that interact with equipment under control. This provides the reassurance that safety-related systems are not connected to the public Internet. Unfortunately, such arguments ignore the ways in which existing malware has been designed to bridge the divide between corporate information systems and control applications. The isolation between these networks is also difficult to sustain when training, development and corporate networks all draw data from their internal 'secure' networks. This has led US Federal agencies to question whether it is possible to draw any clear dividing lines between public facing servers and internal control systems (US Department of Transport, 2009). In consequence, evidence of an attack may be found on corporate networks but we cannot exclude the possibility that the scope of an incident may not extend to operational applications.

- *How to contain an incident?*
  In order to contain a cyber-incident, it is important to identify common factors behind the symptoms of an adverse event. For instance, it may be possible to isolate an infection to components sharing a particular network, or operating system, or sub-contractor. However, there are obvious dangers. For example, if the forensic analysis focusses too narrowly on one sub-contractor then they may miss the cross-infections associated with other companies. Similarly, if logs and records are not available for critical systems then it will be very difficult to take appropriate action to diagnose and contain the scope of an outbreak. These actions may include isolating network components or limiting their interaction with other critical applications until fall-back systems are in place so that primary systems can be shut-down in safety.

- *Is it safe to maintain operations or to perform an emergency shut-down?*
  Once an incident has been detected, it is vital to determine whether or not it is safe to continue operation. Redundancy is often used to maintain safety during routine system failures. If a primary application fails then a secondary system can be brought on-line. However, the high costs of certification mean that these secondary systems are often "moth balled" versions of earlier control systems. Using the fall-back drastically reduces capacity. Alternatively, companies will re-use the same software in primary and secondary systems. This provides hardware redundancy but both primary and secondary systems will share the same vulnerabilities to cyber-attacks. Even if primary and secondary applications are written by different development teams using diverse techniques, they are very likely to use the same operating systems and network protocols. It is very difficult to be sure secondary systems are sufficiently resilient to maintain safety after a cyber-attack. In most cases, operations are restricted to emergency shut-down procedures which may, themselves be compromised by the presence of malware;

- *How long has the attack lasted?*
  It is important to determine how long a violation has continued. The sooner an attack is detected then the easier it is to contain. If a violation or intrusion has continued for some time, engineering teams will have to carefully reconstruct compromised systems from available archives. They then have to re-

integrate the cleaned version with the changes that have occurred to many other hardware and software components. The time and effort required to restore compromised systems creates knock-on concerns. Safety can be compromised when necessary updates to other applications are postponed while engineering teams focus their effort on the recovery from a cyber-attack;

- *How to cope with existing safety concerns?*
  Many contingency plans and exercises assume that cyber-incidents occur under ideal situations; all members of staff are available and every subsystem is working correctly. In reality, violations are detected at almost any time. Many of the problems that exacerbate recovery from cyber incidents stem from the other routine systems failures that characterise everyday operation in safety-critical industries. The additional problems created by cyber-incidents can stretch engineering resources in ways that have not previous been envisaged by companies or regulatory agencies. For example, during the routine installation of updates to a primary application, it is acceptable to use secondary and fall-back systems for a limited period of time. This involves an increased level of risk in safety-critical systems because engineers and operators are typically less familiar with fall-back infrastructures. If a cyber-incident is detected during the operation of a fall-back system then it is difficult to restoring the primary application until the causes of the security violation are identified. This increases exposure to a raised level of risk from the use of fall-back systems beyond the time limits normally deemed acceptable by companies and regulators.

This list provides a partial enumeration of the concerns that arise after a cyber-incident has been detected in safety-critical applications. NIST (2006) identify further concerns. For example, they urge companies to prepare for cyber events that include multiple forms of attack that are identified at the same time in different areas of an organisation:

> "…because of resource limitations, incidents should not be handled on a first-come, first-served basis. Instead, organizations should establish written guidelines that outline how quickly the team must respond to the incident and what actions should be performed, based on relevant factors such as the functional and information impact of the incident, and the likely recoverability from the incident" (NIST, 2006).

It is particularly important that the teams involved in incident recovery document and justify these decisions so that others can learn from their actions. These documents can also be shown to regulators and national security agencies in the aftermath of a cyber-incident. Unfortunately, there is little guidance on how to make such risk-based decisions especially when they can impact public safety. This again illustrates the need for more sustained work to prepare for the recovery from security violations in critical infrastructures.

6. The Cyber-Forensics of Safety-Critical Systems

Forensic analysis is a critical stage in the recovery from cyber-incidents in safety-critical systems. Without an understanding of the causes and consequences of previous violations, it is difficult to accurately assess the risks of future incidents (NIST, 2006). Forensic analysis involves the preservation and study of information associated with computational systems and networks. The aim is to identify what happened, who was involved and to make recommendations that avoid any recurrence of any adverse effects. Hence, forensic analysis builds on network monitoring and system debugging. At a more detailed level, forensic analysis involves the identification, retrieval, preservation, interpretation and presentation of evidence relating to the abuse of computer systems. Evidence includes files and logs derived from monitoring a compromised system. It can also include paper and digital documentation, describing the processes used to secure key assets.

There are two different types of forensic analysis. Investigations can focus on the computer systems that were used during the course of a criminal activity or on the computational infrastructures that were the target of a crime. These distinctions become blurred, for instance when insiders use corporate networks to launch their attacks. The following sections argue that forensic techniques cannot easily be applied to analyse the causes and consequences of cyber-attacks on safety-critical infrastructures. In particular, existing approaches focus on gathering evidence but do not consider the need to protect the lives of end users and system operators.

Bassett, Bass and O'Brien (2006) propose a number of generic processes that support the forensic analysis of computer systems:

1. Protect subject computer system from alteration, data corruption, virus infection, and physical damage;
2. Uncover all files: normal, hidden, deleted, encrypted, password-protected;
3. Recover as many of the deleted files as possible;
4. Reveal the contents of hidden and temporary files;
5. Access the protected and encrypted files, if legal;
6. Analyse all relevant data, including data located in unallocated file space and file slack;
7. Print out a listing of all relevant files, and provide an overall opinion on the system examination;
8. Provide expert testimony or consultation, if required.

Bassett et al focus on the forensic analysis of file systems. However, in safety-critical systems it is important to gather forensic evidence through network monitoring. This is essential when an attack might affect sensor data or actuator commands that have a profound impact on the application processes that are being controlled.

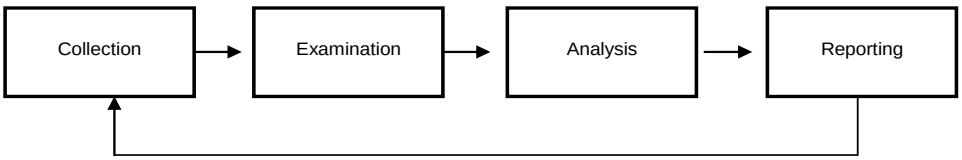

Figure 2: NIST Guidance on Forensic Processes

NIST (2006) have developed a more generic description of the key activities involved in the forensic analysis of cyber-attacks on computation infrastructures. These are illustrated in Figure 2 and can be summarised as follows:

- **Collecting** identifying, labelling, recording, and acquiring data from the possible sources of relevant data, while following procedures that preserve the integrity of the data. This stage can be compared to phases 1-2 of the model proposed by Bassett et al;

- **Examining** collected data using a combination of automated and manual methods, and assessing and extracting data that is of interest. This must preserve the integrity of the data. The second stage of the NIST model can be compared to phases 3 and 4 of the previous approach;

- **Analysing** the results of the examination, using legitimate methods to identify the causes and consequences of an attack. This stage is similar to phases 5 and 6 of the forensic model proposed by Bassett et al;

- **Reporting** the results of the analysis. Describing and justifying the methods used, explaining how tools and procedures were selected, determining what other actions need to be performed and providing recommendations for policies, procedures, tools, and other aspects of the forensic process. This can be compared to phases 7 and 8 of the previous approach.

Reith, Carr and Gunsch (2002) extend the NIST model. They stress the need to develop an approach strategy that minimises the impact on other members of the public and other users of computational systems during a forensic investigation. They also argue that investigators must consider how to return evidence to its proper owner. The Reith et al taxonomy is intended to cover the same general mix of 'conventional' office-based systems as the NIST guidance. However, these two additional phases are important in the context of safety-critical systems. For example, the 'approach phase' can ensure that the gathering of evidence has a minimal impact on other safety-related applications that remain unaffected by a cyber-incident. If the forensic analysis disrupts normal operation – for example in adjacent air traffic control sectors, then the response to an incident might be equally as dangerous as the attack itself. If the primary sector is closed during the forensic analysis then investigators must plan for and mitigate any knock-on increases in workload for the adjacent sectors. Similarly, the 'return of data' is important given that forensic evidence is likely to be commercially sensitive and to have an impact on future security. Inadvertent disclosure could do long term damage to the participation of organisations in subsequent investigations and undermine the future safety of application processes.

Previous sections argued that existing guidelines must be extended if they are to help safety-critical industries recover from cyber-attacks. For example, NIST and ENISA provide valuable recommendations that will speed the restoration of conventional 'office based' services. However, their frameworks tend to focus on the consequences for computational infrastructures and arguably do not consider the impact of an attack on the safety of complex application processes. The same observations apply to existing guidelines for the forensic analysis of cyber incidents. For example, NIST and ENISA both recommended intrusion detection through the introduction of network monitoring tools. However, great care must be taken that these systems do not undermine the real-time requirements of safety-critical applications. We must, therefore, extend existing forensic guidelines to ensure that stakeholders can gather necessary evidence without endangering safety. Equally, we must not inadvertently lose forensic evidence that might otherwise have been preserved without loss of safety - for instance by saving volatile memory prior to an emergency shut-down procedure.

The final element of the NIST (2006) model for forensic analysis; focuses on reporting. Safety-critical companies must carefully consider the forensic information that they release into the public domain, to their competitors and also to some government agencies. Both NIST and ENISA identify the dangers in sharing too much forensic information in the aftermath of a cyber incident, especially when reports disclose commercially sensitive information. Cyber incident reports can also disseminate details about vulnerabilities that persist in other companies. There is a natural concern to withhold any details that might motivate subsequent attacks. Although these concerns are the same for the victim of any cyber incident, they can have a more profound impact on safety-critical national infrastructures. The increasing integration of service providers both within and across national borders relies on mutual trust. For instance, energy transmission companies continually exchange data with their neighbours in response to numerous changes in the balance between supply and demand (Johnson, 2008). Admitting that networks and systems have suffered a cyber-attack creates a concern that industry partners will act to isolate the possible source of any contamination, creating significant knock-on effects across Europe and North America. Similarly, if members of the public learn that safety has been compromised by an attack then it may take months or years to restore their confidence in the underlying architectures. Hence the dissemination of forensic information takes on an added importance in the case of safety-critical systems. Existing guidance tends to focus more on the importance of technical communication about the causes and consequences of a cyber-incident rather than ensuring that the public and media get a proportionate view of the risks created by an attack.

7. Securing the Evidence of Cyber-Attacks on Safety-Critical Infrastructures

The providers of safety-critical services must consider the legal context of any forensic investigation. The US Department of Justice (2008) argue that compromised systems should be treated like any other crime scene. First responders should:

- "Follow departmental policy for securing crime scenes.
- Immediately secure all electronic devices, including personal or portable devices.
- Ensure that no unauthorized person has access to any electronic devices at the crime scene.
- Refuse offers of help or technical assistance from any unauthorized persons.
- Remove all persons from the crime scene or the immediate area from which evidence is to be collected.
- Ensure that the condition of any electronic device is not altered.
- STOP! Leave a computer or electronic device off if it is already turned off".

Some of these guidelines are applicable to safety-critical infrastructures. The US Department of Justice stress that the "first responders' primary consideration should be officer safety and the safety of everyone at the crime scene". There are obvious concerns about accepting help from unauthorised personnel. Others requirements are less easily applied to safety-related systems. For instance, it is hard to enforce a requirement to "remove all persons from the crime scene or the immediate area from which evidence is to be collected". In a crowded Air Traffic Control centre, the recovery from a cyber-incident requires cooperation between safety management as well as engineering and operational teams. It requires technical input from sub-contractors and external service providers. Similarly, cold stand-by applications often provide redundant support when primary control systems are compromised by adverse events. These techniques contradict the Department of Justice guidelines to "leave a computer or electronic device off if it is already turned off". Rather than revising the guidance, companies might change their policies and practices to support the forensic analysis of safety-critical applications. For instance, efforts could be made to partition cold-standby systems to ensure minimal interference between primary and backup systems during any handover. In either event, further work is required to ensure that the existing guidance provides pragmatic support for the recovery from cyber incidents in this class of applications.

It is difficult to apply US Department of Justice guidelines to safety-critical systems because they focus on the forensic analysis of systems involved in committing a crime rather than safeguarding evidence when infrastructures have been the target of a cyber-incident. This is illustrated by the priority actions to be taken by a first responder:

- "Look and listen for indications that the computer is powered on. Listen for the sound of fans running, drives spinning, or check to see if light emitting diodes (LEDs) are on.
- Check the display screen for signs that digital evidence is being destroyed. Words to look out for include delete, format, remove, copy, move, cut, or wipe.
- Look for indications that the computer is being accessed from a remote computer or device.
- Look for signs of active or on-going communications with other computers or users such as instant messaging windows or chat rooms.
- Take note of all cameras or Web cameras (Web cams) and determine if they are active".

(US Department of Justice, 2008)

Such guidance is useful in conventional investigations, especially where first responders have minimal experience in digital forensics or network management. However, the individuals and teams who respond to cyber incidents in safety critical systems, typically, possess a far more detailed understanding of the infrastructures that they operate. Rather than identifying which systems are live, the focus is likely to be more on determining whether it is possible to gather sufficient evidence to diagnose the scope of an attack without undermining the safety of operators or the general public. Not only must they gather evidence about the nature of a cyber incident, they

must also document their actions during the recovery process to reassure regulators that lives were not endangered during the response.

The forensic evidence that is gathered about the causes and consequences of a cyber-attack can inform future legal proceedings, regulatory actions and internal disciplinary procedures. Safety-critical organisations must, therefore, establish a chain of custody to avoid allegations that evidence has been mishandled or deliberately altered. This involves:

- logging every person who had physical custody of the evidence;
- documenting the actions that they performed and at what time;
- storing the evidence in a secure location when it is not being used;
- making a copy of the evidence and only analyse the duplicate copy;
- verifying the integrity of the original and the duplicate evidence.

Many safety-related organisations can re-use existing tools and techniques to preserve the chain of evidence recommended by the US Department of Justice (2008). Companies working in safety-critical industries already have contingency plans that can be used to preserve the evidence required by boards of enquiry in the aftermath of major accidents. Existing monitoring systems and replay facilities already support more routine engineering activities and staff training. These tools can be redeployed in the aftermath of a cyber incident. However, there are potential risks. Allocating simulation and replay facilities to forensic investigations can have knock-on effects when other staff cannot use them to implement routine updates and bug fixes. The resulting delays can undermine operational safety. There are other concerns. Simulation and replay facilities have themselves been a target for cyber-attacks. This compromises the ability of companies to conduct a forensic analysis of any subsequent attacks on primary systems and may also increase operational risk by delaying updates to safety-related control systems. There is also a concern that simulation and replay tools are more vulnerable than primary systems because they are often classified at the same level of security as corporate office systems and yet data is continually transferred between these tools and operational systems, frequently by USB sticks.

It is still rare for safety-related organisations to provide formal reports of cyber incidents to industry regulators. In many cases, organisations are so anxious to resume service provision that they immediately try to restore systems without preserving any evidence about the source of an attack (Johnson, 2012). The immediate destruction of evidence creates significant concerns for future security, especially when the initial detection heralds the first in a series of incidents. In such circumstances, the preservation of forensic information protects the data needed to establish the source of an infection. NIST (2006) summarise this in their guidance for the forensic analysis of cyber security incidents; “if it is unclear whether or not evidence needs to be preserved, by default it generally should be preserved… evidence that seems insignificant now may become more important in the future. For example, if an attacker is able to use knowledge gathered in one attack to perform a more severe attack later, evidence from the first attack may be key to explaining how the second attack was accomplished”. As a rule of thumb, they have advised that such data should be kept for up to three years before organisations are in a position to accurately determine whether or not it should be deleted.

Data retention creates considerable problems for forensic analysis in safety-critical industries. In more conventional office environments, hardware can be mothballed so that companies can access data that was collected several years before. However, most process industries integrate dozens of complex hardware and software systems that require careful integration and configuration. It is, therefore, not always possible to ensure that legacy data can always be interpreted by future systems. Typically, it is possible to replay or upload process data using conversion routines. However, the use of translation tools can undermine the evidential rules that guide forensic analysis. They alter the source data that was collected in the aftermath of a cyber-attack. Incident response teams must, therefore, retain access to a range of tools and resources that can be used to analyse copies of incident data in a way that does not undermine original evidence. This is not simply a matter of

mothballing obsolete equipment; it also involves careful archiving of network and application parameters, port configurations etc. At present, most safety-critical industries would find it hard to justify the investments required to support such forensic infrastructures.

8. Legal Perspectives on the Cyber-Forensics of Safety-Critical Infrastructures

NIST (2006) urge organisations to contact the legal agencies that must respond to cyber incidents; "one reason many security-related incidents do not result in convictions is that some organizations do not properly contact law enforcement… incident response team should become acquainted with its various law enforcement representatives before an incident occurs to discuss conditions under which incidents should be reported to them, how the reporting should be performed, what evidence should be collected, and how it should be collected". Very few companies contact law enforcement agencies before a cyber incident occurs (US Department of Justice, 2008). One reason for this is the widespread confusion over who to consult. In the United States, there are Federal investigatory agencies (e.g., the Federal Bureau of Investigation [FBI] and the U.S. Secret Service), district attorney offices as well as both state and local (e.g., county) law enforcement. Regulatory agencies introduce an additional layer of complexity in safety-related industries. In Europe, responsibility lies with national bodies such as the UK Civil Aviation Authority or the Office for Nuclear Regulation. Very few of these regulators provide any guidance about the legal interactions that must take place during the forensic analysis of a security incident. There is also a need to draft letters of agreement between regulatory and state security agencies to establish roles and responsibilities in the aftermath of a cyber-attack. These help to avoid "turf wars" when engineers are working to ensure the safety of compromised systems.

The legal and regulatory framework for the investigation of cyber incidents is likely to become more and more complex with the increasing cross-border integration of national critical infrastructures. The European Commission has encouraged the interoperability of national rail networks. They have promoted the development and integration of smart grids and the creation of the single European skies network. We have already glimpsed the confusion that this can create in the investigation of conventional system failures. The Viareggio accident occurred in Italy. It involved a train that was composed of wagons registered in Poland and in Germany, some of which were owned by an Austrian company under lease to a US corporation. The subsequent jurisdictional issues led to an investigation that has taken years not months to complete. The web of international interdependencies can be far more complex in some cyber incidents. Safety-related organizations in one country can draw operational data from servers located in a second state that are attacked by systems in a third nation that are remotely controlled by attackers in a fourth state (US Department of Justice, 2004).

External consultants can help companies recover from cyber-incidents. They provide a level of expertise in digital forensics that is missing from many safety-critical organisations. Their testimony can also be critical in subsequent litigation. Their evidence can convince regulators that organisations took appropriate measures in the aftermath of security violations. Companies often find it difficult to identify objective measures of expertise. In the United States, the Frye test defined admissible expert testimony in terms of techniques that are generally accepted as reliable in the relevant scientific community. Although there are some 'generally accepted' techniques in digital forensic, very few of them have been used in any sustained way within safety-critical applications. Significant questions remain about the use of these tools when lives may be at risk. However, the Daubert standard has gradually replaced the Frye test; introducing a more flexible concept of 'reliability'. The Judge must find it, more likely than not, that an expert's methods are reliable when applied to the facts in a case. These concepts have been embodied within the Federal Rules of Evidence:

> "A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if:

a) The expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;

b) The testimony is based on sufficient facts or data;

c) The testimony is the product of reliable principles and methods; and

d) The expert has reliably applied the principles and methods to the facts of the case".

(Federal Rules of Evidence, Rule 702)

If individuals and teams have not been trained to use reliable "principles and methods" for the forensic analysis of cyber-attacks then it is unlikely that their evidence will be admissible, even in cases that involve commercial or national security. If existing "principles and methods" cannot be applied then new techniques must be developed to support an effective response to future incidents involving safety-critical applications.

The Federal Rules of Evidence create a framework that enables courts to determine whether digital evidence is admissible. They can also be used to determine whether an original digital data source must be used or whether a copy is acceptable. Article IX deals with the authentication and identification of evidence. Article X deals with recordings, including logs derived from computing networks etc. Investigators must establish the reliability of the computer equipment used to create a recording. They must document the measures taken to insure the accuracy of the data when it was first entered. They must also document the precautions taken to prevent loss of digital evidence, this includes the processes and procedures used to archive and later retrieve system logs. Investigators must document the techniques that were used to verify the reliability of any computer programs used to process the data etc. Each of these requirements creates further burdens for safety-critical companies. Senior managers are often reluctant to allocate finite resources to develop procedures and train staff to protect evidence when their company may not be the target of a cyber-attack. It is tempting to assume that the 'clean up' process will be handled by a government funded CSIRT (Computer Security Incident Response Team). This reliance on CSIRTs is misplaced. They, typically, understand the legal and forensic challenges of incident recovery. However, they have little understanding of the regulatory or technical processes used to preserve the safety of complex industries.

In the UK, the Association of Chief Police Officers (2007, 2011) has published guidelines for the handling of evidence in forensic investigations. The aim is to support law enforcement officers in the aftermath of a cyber-incident. They have structured their guidance around four common principles:

1. No action taken by law enforcement agencies or their agents should change data held on a computer or storage media which may subsequently be relied upon in court;
2. In circumstances where a person finds it necessary to access original data held on a computer or on storage media, that person must be competent to do so and be able to give evidence explaining the relevance and the implications of their actions;
3. An audit trail or other record of all processes applied to computer-based electronic evidence should be created and preserved. An independent third party should be able to examine those processes and achieve the same result;
4. The person in charge of the investigation (the case officer) has overall responsibility for ensuring that the law and these principles are adhered to.

The ACPO guidance illustrates strong similarities between the approaches introduced across both sides of the Atlantic. For instance, Point 4 encapsulates assumptions that are also embedded within the guidance from the US Department of Justice (2008). The official in charge of an investigation must ensure that the investigation of a cyber-incident meets the principles that are intended to guide forensic analysis. This has profound implications

in safety-critical systems where, for instance, there can be conflicts between a desire to meet forensic principles and the need to maintain levels of safety and levels of service. In practice, it is likely that senior investigators will have to work closely with systems and safety engineers. We urgently need more guidance on the roles and responsibilities of these key staff in the aftermath of cyber incidents affecting national critical infrastructures. In related work, we have described the importance of conducting multi-party drills to facilitate these interactions (Johnson, 2012) – law enforcement officers do not always understand the complex interactions that might undermine public safety, conversely systems engineers typically have a limited understanding of the legal issues summarised in this paper.

9. Conclusions and Further Work

Safety-critical systems are vulnerable to a range of threats. The reliance on mass-market, Commercial off-the Shelf (COTS) operating systems and network protocols undermines previous assumptions about the probability and consequences of an attack. The lack of integration between safety-critical development practices and security engineering techniques creates further vulnerabilities. Safety standards encourage service providers to conduct extensive validation and verification tests before software updates can be integrated into operational systems. These safety requirements delay the introduction of new malware definitions. Even if we can find ways of updating cyber-security systems without undermining the safety of application processes, we still cannot guarantee that a process will be resilient to all potential attacks.

A number of organisations, including ENISA, NIST, the UK Association of Chief Police Officers and the US Department of Justice provide guidance for organisation to respond to, and recover from, cyber incidents. Their pioneering work increases the resilience of many different industries. However, the existing guidelines all focus on corporate office based systems. They do not consider the impact of cyber-attacks on complex, safety related applications. In consequence, first responders are urged to disconnect systems and thereby preserve forensic evidence even though this might have a profound impact on public safety. Similarly, there is no specific guidance on the complex risk-based decisions that must be made when managers choose between shutting down an infected system and starting a less familiar secondary application that might already have been compromised.

Much remains to be done. Future work intends to use the work of ENISA and NIST as the starting point for generic guidance for the forensic analysis of cyber-incidents across several different safety-critical industries. This future research poses a host of practical and theoretical challenges. In particular, there are considerable differences between avionics applications and healthcare systems, between nuclear generation facilities and process control systems. A key research question is, therefore, whether it is possible or useful to develop generic guidelines that might support forensic analysis across such a diverse range of industries. The use of COTS operating systems and network infrastructures across transportation, healthcare, power distribution creates a need for consistent approaches to cross-modal attacks. There are no established practices and procedures for identifying the impact that a single attack vector could have on the safety of many different industries

Acknowledgements

Thanks are due to Brad Glisson, George Grispos and Tim Storer for the informal discussions that led to this paper. All mistakes and omissions are entirely due to the author.

References

Association of Chief Police Officers (2007), Good Practice Guide for Computer-Based Electronic Evidence, 2007, ACPO. http://www.met.police.uk/pceu/documents/ACPOguidelinescomputerevidence.pdf

Association of Chief Police Officers (2011), Managers Guide: Good Practice and Advice Guide for Managers of e-Crime Investigation, 2011, ACPO. http://www.acpo.police.uk/documents/crime/2011/201103CRIECI14.pdf

R. Bassett, L. Bass and P. O'Brien (2006), Computer Forensics: An Essential Ingredient for Cyber Security, Journal of Information Science and Technology, 3(1) 2006.

European Network and Information Security Agency (ENISA) (2009), Good Practices on Reporting Security Incidents, Heraklion, Greece, December 2009. http://www.enisa.europa.eu/activities/Resilience-and-CIIP/Incidents-reporting/good-practice-guide-on-incident-reporting/good-practice-guide-on-incident-reporting-1

European Network and Information Security Agency (ENISA) (2009a), Cloud Computing: Benefits, risks and recommendations for information security, Heraklion, Greece, November 2009. http://www.enisa.europa.eu/activities/risk-management/files/deliverables/cloud-computing-risk-assessment

European Network and Information Security Agency (ENISA) (2011), Technical Guidelines on Reporting Incidents: Article 13a Implementation, Heraklion, Greece, December 2011. http://www.enisa.europa.eu/activities/Resilience-and-CIIP/Incidents-reporting/Technical%20Guidelines%20on%20Incident%20Reporting/incidents-reporting-to-enisa/technical-guideline-on-incident-reporting

C.W. Johnson (2008), Understanding Failures in International Infrastructures: A Comparison of Major Blackouts in North America and Europe. In R.J. Simmons, D.J. Mohan and M. Mullane (eds.), Proceedings of the 26th International Conference on Systems Safety, Vancouver, Canada 2008, International Systems Safety Society, Unionville, VA, USA, 2008.

C.W. Johnson (2012), Preparing for Cyber-Attacks on Air Traffic Management Infrastructures: Cyber-Safety Scenario Generation. In Proceedings of the 7th IET Conference on Systems Safety and Cyber-Security, Edinburgh, Scotland, 15-18 October 2012, IET, Savoy Place, London, 2012.

C.W. Johnson (2012a), CyberSafety: On the Interactions Between CyberSecurity and the Software Engineering of Safety-Critical Systems. In C. Dale and T. Anderson (eds.), Achieving System Safety, Springer Verlag, 85-96, London, UK, ISBN 978-1-4471-2493-1, 2012.

C.W. Johnson and A. Atencia Yepez (2010), Safety Cases for Global Navigation Satellite Systems' Safety of Life (SoL) Applications. In Proceedings of the Fourth International Association for the Advancement of Space Safety, Huntsville Alabama, NASA/ESA, Available from ESA Communications, ESTEC, Noordwijk, The Netherlands, ISBN 978-92-9221-244-5, SP-680, 2010.

C.W. Johnson, B. Kirwan and T. Licu (2009), The Interaction Between Safety Culture and Degraded Modes: A Survey of National Infrastructures for Air Traffic Management , Risk Management, (11)3:241-284, 2009.

Reith, M., C. Carr, and G. Gunsch (2002), An examination of digital forensic models. International Journal of Digital Evidence, 2002. 1(3): p. 1-12.

U.S. Department of Justice (2004), Forensic Examination of Digital Evidence: A Guide for Law Enforcement, 2004.

U.S. Department of Justice (2008), Office of Justice Programs, Electronic Crime Scene Investigation: A Guide for First Responders, Second Edition, Washington DC, 2008. http://www.nij.gov/publications/ecrime-guide-219941/

U.S. Department of Transport (2009), Report on Review of Web Applications Security and Intrusion Detection in Air Traffic Control Systems, FAA Report Number FI-2009-049, Washington DC, USA, May 2009.

U.S. General Accounting Office (1998), Air Traffic Control: Weak Computer Security Practices Jeopardize Flight Safety (Letter Report, 05/18/98, GAO/AIMD-98-155), 1998.

U.S. National Institute of Standards and Technology (NIST) (2006), Guide to Integrating Forensic Techniques into Incident Response, Special Publication 800-86, Gaithersburg, Maryland, 2006.
http://csrc.nist.gov/publications/nistpubs/800-86/SP800-86.pdf

U.S. National Institute of Standards and Technology (NIST) (2011), Guidelines on Security and Privacy in Public Cloud Computing, NIST Special Publication 800-144, Gaithersburg, Maryland, December 2011.
http://csrc.nist.gov/publications/nistpubs/800-144/SP800-144.pdf

U.S. National Institute of Standards and Technology (NIST) (2012), Computer Security Incident Handling Guide (Draft), Special Publication 800-61 Revision 2 (Draft), Gaithersburg, Maryland, 2012.
http://csrc.nist.gov/publications/drafts/800-61-rev2/draft-sp800-61rev2.pdf

U.S. Nuclear Regulatory Commission (2010), Cyber Security Programs for Nuclear Facilities, Office Of Nuclear Regulatory Research, Regulatory Guide 5.71, January, 2010.
http://pbadupws.nrc.gov/docs/ML0903/ML090340159.pdf