SCC.444 Information System Risk Management

An analysis of the HRO approach to risk management from an information security perspective

Ben Goldsworthy, 32098584
b.goldsworthy@lancaster.ac.uk

Abstract

In this essay, the relavance of the theory of High-Reliability Organisations (HRO) to the provision and risk assessment of information security services is assessed. The two main variants of HRO theory are discussed, and an HRO's potential approach to information security is detailed. This is followed by an analysis of the flaws in, and proposed alternatives to, HRO theory. Finally, the pros and cons of each of these competing theories in their application to information security are considered. Ultimately, a synthesis of them all is proposed as the most useful, as each brings something useful to the table, and the urgency of further research is reinforced.

I. Introduction

'High-Reliability Organisations' (HROs) are a subset of safety-conscious organisations, comprising of any organisation which, due to the nature of its work, is in the position that any failure in risk management may pose a threat to the safety of a large number of stakeholders. A commonly-cited example of an HRO is that of a nuclear power plant, in which operation is nearly fault-free despite the incredible preponderance of risk that they are daily exposed to; recently, however, medical instutions have begun adopting elements of the HRO methodology. In this essay, the elements of the HRO approach to risk management within information security will be detailed and critically assessed. First, the HRO theory will be detailed. Secondly, a picture will be sketched of what an HRO's approach to information security may look like. Thirdly, the limitations of the HRO theory are discussed and a number of alternatives proposed. Fourthly, these alternatives shall be applied to the information security perspective and their relative strengths and weaknesses pointed out.

II. What is an HRO?

The concept of the 'High-Reliability Organisation' was first developed by Roberts and Rousseau (1989) in response to a perceived lack of coverage in the literature of organisation research for this specific subset of organisations. Roberts and Rousseau considered HROs to be 'a subset of high-risk organisations designed and managed to avoid [catastrophic] incidents'. A crucial characteristic of the HRO was identified as being their approach to learning. 'Most organizations engage in trial and error and other experimental forms of learning every day', write Roberts and Rousseau. 'In high-reliability organizations, the cost of this kind of learning can far exceed the value of the lessons learned', and 'such organizations can destroy themselves and perhaps a larger public, entirely wiping out evidence that might be used for learning'. This seems obvious—one would be hard-pressed to find a proponent for a trial-and-error approach to nuclear safety.

Roberts and Rousseau go on to specify eight primary characteristics of an HRO:

1) hypercomplexity;
2) tight coupling and reciprocal interdependence across many units and levels;
3) extreme hierarchical differtiation;

4) large numbers of decision makers in complex communication networks;
5) degree of accountability that does not exist in most organizations;
6) high frequency of immediate feedback about decisions;
7) compressed time factors; and
8) more than one critical outcome that must happen simultenously.

Roberts (1990) further developed the theory of this new class of organisation, citing P. Shrivastava (1986) in noting that 'the number of organisation capable of killing large numbers of people is growing', and that the set of HROs are those that have nonetheless 'enjoyed a record of high safety over long periods of time'. Finally, she posed the question by which one may identify an HRO: 'how many times could this organisation have failed resulting in catastrophic consequences that it did not?' If the answer to this quesiton is 'on the order of tens of thousands of times the organisation is "high reliability"'.
This view of the HRO has been challenged since its proposal. Rochlin (1993) proposed a change of approach from that of identifying HROs by the reliability they *achieve*, towards that of identifying them by the reliability they *seek*. This potentially serves to better identify the full range of HROs currently operating. The logical result of the ideal HRO is that they never experience a lapse of reliability. Couple with the fact that HROs tend to be overrepresented within relatively new fields (e.g. nuclear power), there is the risk that an entire industry of HROs maybe so judicious in their continuing reliability that they, collectively, have yet to experience any such catastrophe. In this situation, the risks of the field may not been known and the HRO's reliability within it overlooked. Would we consider nuclear power so risky had we not had our Chernobyls, our Three Mile Islands, our Fukushimas?
There have been further recent developments in the field of HRO study. Weick, Sutcliffe, and Obstfeld (2008) proposed the concept of 'collective mindfulness'—that is, the cultural development within an organisation of constant, collective awareness at all levels. This was further developed by Weick and Sutcliffe (2015), who clarified the theory as 'mindful organizing'. They identified five key principles of collective mindfulness:

1) preoccupation with failure;
2) reluctance to simplify;
3) sensitivity to operations;
4) commitment to resilience; and
5) deference to expertise.

The contemporary definition of an HRO may be found in the words of the US Nuclear Regulatory Commission (2010, p. 2), which state that the term covers any organisation 'that operates and manages processes with the potential to adversely affect human life or the environment'. This definition may strike the reader as similar to that of 'Critical National Infrastructure' (CNI), defined by the Center for the Protection of National Infrastructure (2017) as 'those facilities, systems, sites, information, people, networks and processes, necessary for a country to function and upon which daily life depends[, including] some functions, sites and organisations which are not critical to the maintenance of essential services, but which need protection due to the potential danger to the public'. In practice, the two definitions have broad overlaps—the canonical example usually given of both is that of the nuclear power plant—but CNI tends to refer to the equipment itself, whilst HRO refers to the organisation running it and its management processes. In this sense, Heysham nuclear power station is an example of CNI, but EDF Energy is (hopefully) an HRO.
The US Nuclear Regulatory Commission (2010, p. 7) incorporates elements of Weick and Sutcliffe's theory of 'mindful organizing' by declaring that HROs possess an 'HRO safety culture', comprising of 'professional leadership attitudes [...] that manage potentially hazardous activities to maintain risk to people and the environment as low as reasonably achievable, thereby assuring stakeholder trust.' They argue that it is essential that this culture be leader-led as, otherwise, 'production practices will overcome those aimed towards prevention' (ibid., p. 16). Quoting Peter F. Drucker, they point out that 'management is "doing things right"; leadership is "doing the right things"'. In this case, the potential price of doing the wrong things well is so high that providing the correct leadership is of the utmost

importance.

III. What might an HRO's approach to information security look like?

There appears to be next to no research on the HRO theory as applied to information security. Instead, it is left to us to apply Weick and Sutcliffe's five principles ourselves. In this section, we shall do exactly that.

A. Preoccupation with failure

Jacobson (2015) writes the 'HROs do not ignore any failure, no matter how small, because any deviation from the expected result can snowball into tragedy.' Lowers & Associates (2017) state that 'HROs do NOT [...] assume that if a control in place succeeds in containing a failure, everything is right. They look deeper into an incident to find underlying causes.' This can be applied with little effort to the information security field. Akamai (2018) revealed recently that some 48% of login attempts are malicious. Much of this malicious activity is relatively benign—automated web crawlers trying `password:password` against any login forms they find—but within that statistic is, presumably, human actors. To use the current terminology, some of these login attempts are likely to be the behaviour of advanced persistent threats (APTs), potentially during the active reconnaisance phase of the intrusion kill chain (Hutchins, Cloppert, and Amin 2011). An IT HRO would have to have in place some sort of technique for investigating all such intrusion attempts in order to determine whether they are the mere prelude to a larger threat, or an unavoidable consequence of operating on the Internet.

B. Reluctance to simplify

'High Reliability Organizations are complex by definition and they accept and embrace that complexity.' (Jacobson 2015) This suggests that an HRO would approach information security not in isolation, but as a tightly integrated part of numerous other systems. For example, IBM (2014) claimed that as many as 95% of security incidents involve human error. This demonstrates that the issue of information security is multidiscipinary—in this case alone, it touches on elements of employee education, human resources and the need for layered security. In an HRO, we would expect an institutional appreciation of this notion.

C. Sensitivity to operations

The crucial aspect of this principle is that, regardless of how well-intentioned or well-designed it may be in *theory*, any organisational behaviour or proces is useless if not also accompanied by a continous observation of the *actual* effects. The insights gleaned by this two-way process can then help to further develop the original policy in order to make it even more efficacious. A good example of this would be a company that implemented rigid password creation requirements—'must be x number of characters long', 'must contain upper- and lower-case characters', and so on—but then failed to look around the office and check that employees weren't sticking their passwords on Post-It notes by their computers in order to remember them.

D. Commitment to resiliency

Weick and Sutcliffe state that 'the signature of an HRO is not that it is error free, but that errors don't disable it.' The concepts of layered security—wherein a 'series of different defenses should each be used to cover the gaps in the others' protective capabilities'—and defence-in-depth—wherein 'technological components [...] are regarded as stumbling blocks that hinder the progress of a threat [...] until either it ceases to threaten or some additional resources [...] can be brought to bear'—receive much attention within information security literature (Perrin 2008). That resiliency is a fundamental

concern of information security is further evidenced by the prevelance of redundancy, and thus resiliency, in the design of the very protocols that underpin all such technology, particularly networking protocols. Therefore, we might expect this principle to manifest in an organisation by means of layered security, including fine-grained access control and multi-factor authentication, coupled with multiple firewalled networks, airgapping where appropriate, and so on. That any one of these elements of security can be breached is not an issue; the intention is that all of them being breached, simulteneously, is extremely unlikely.

E. Deference to expertise

The final defining principle of the HRO is that the opinion of the subject matter expert is afforded the respect that their expertise warrants, as opposed to their authority and seniority within the orginisation. From the information security perspective, this would mean an organisation in which the lowly IT support worker's concerns about the CEO's password security, for example, were acted on promplty and effectively. That these concerns did not come from the CTO would be immaterial.

IV. What are the limitations of the HRO approach?

As with all theories, the concept of the HRO is not without its detractors. A principal criticism if that of mistakingly lauding reliability as the primary desirable goal for an organisation. As Hopkins (2007) writes, 'reliability, especially reliability of supply, is not always equivalent to safety. Indeed the two may pull in opposite directions.' The example that Hopkins provides is that of the power plant in a situation in which 'safety may [...] depend on shutting off supply.'

There is also the epistemological question of how one can identify what one does not know to look for. This falls within the Rumsfeldian territory of 'unknown knowns' and 'unknown unknowns' (Rumsfeld 2002). Rashid et al. (2016) have analysed the case of the former—'knowledge that is "unknown" to the requirements engineer, and so does not make its way into requirements, but is "known" in that it exists in known security breaches'—but the latter is a more intractable problem.

As mentioned previously, Roberts and Rousseau's original definition of an HRO relies upon successful identification of all industries of potential catastrophic risk. Would we possess, as a species, so full an appreciation of the risk posed by the pandemic outbreak of an incurable disease had we not experience of the Black Death and the Spanish Flu? Would we posesses so awesome a fear of nuclear weaponry, to the degree that its continued unuse is allegedly held in check only by the threat that to use it would invite it to be used on you in return, had we not pictures and testimony from the blasted ruins of Hiroshima and Nagasaki? Would we posess such a wariness around nuclear power had we not the Chernobyl sarcophagus to visit? If not, then we must consider what the unknown unknowns may turn out to be, before the existential threat that they pose in the shadows has a chance to materialise. This must, clearly, be a prerequisite to determining how reliable the organisations responsible for its provision have historically been.

Weick and Sutcliffe's journey-focused, rather than destination-focused, approach to defining the HRO as an ideal to which any organisation may aspire is perhaps more useful for our purposes. However, the democratising effect of this redefinition serves to devalue, somewhat, the supposed uniqueness of the HRO. Whereas the original HRO theorists saw fit to declare that industries such as railroad and petroleum were precluded from harbouring HROs for various reasons (Hopkins 2007, p. 6), Weick and Sutcliffe allows the definition to be applied to everyone and, in doing so, make it somewhat less meaningful. As Marais, Dulac, Leveson, et al. (2004) puts it, 'it is difficult to think of any low reliability organisations'.

Ultimately, one of the basic tenets of capitalism is that any low reliability organisation (unless otherwise supported, for example via a monopoly) will be rejected by the market in favour of a high reliability one. With this in mind, Weick and Sutcliffe's step-by-step checklist serves more of a function as some sort of 'how-to' guide for aspiring managers and business owners to improve their organisation's reliability,

whereas Roberts and Rousseau's exclusive club is more accurately thought of as a inaccurately-named identification not of 'high reliability organisations', but of 'unexpectedly reliable organisations within a specific set of industries'. The former is easy to apply to the information security approach of an organisation, but is too soft to be able to draw any particularly interesting conclusions from. The latter, however, does not lend itself to extensibility into fields unpredicted by its authors (such as many within modern information security).

Any theory must have an oppositional yang to its yin, and that of the HRO is no different. HRO theory takes the optimistic approach that 'that extremely safe operations are possible, even with extremely hazardous technologies, if appropriate organization design and management techniques are followed'. On the flipside, the Normal Accidents Theory (NAT) prposed by Perrow (2011) 'presents a much more pessimistic prediction: serious accidents with complex high technology systems are inevitable.' (Sagan 1995) In this theory, nothing that an organisation does can halt a serious accident once the system in which the organisation operates reaches a critical threshold of complexity—it can only hope to delay it.

NAT can be applied productively to some elements of information security. The recently-discovered Spectre and Meltdown vulnerabilities, which fundamentally affect an entire decade's worth of CPUs, could be argued as a result of the field of chip manufacture reaching this threshold of complexity (for more from the current author on these vulnerabilities, see Goldsworthy (2018)). That it has been thusfar catastrophic not in the 'screaming and melting flesh' sense, but rather the 'very expensive class-action lawsuit and the potential death of Moore's Law' sense, might suggest how a catastrophic IT accident may look. Indeed, 'whilst NAT has previously focused on the consequences of physical accidents', Nunan and Di Domenico (2017) propose 'a new form of system accident that we label *data accidents*.' Some, such as S. Shrivastava, Sonpar, and Pazzaglia (2009), claim that 'the two theories [HRO and NAT] appear to diverge because they look at the accident phenomenon at different points of time', and that they are, in fact, complementary rather than competetive in nature. When taking this approach, the issue becomes that both theories are unfalsafiable (Rosa 2005). S. Shrivastava, Sonpar, and Pazzaglia summarises thus:

> If a tightly coupled complex system were to succeed in avoiding an accident, NAT proponents would attribute the safe outcome to the system in question being not complicated enough. Similarly, in the event of an accident in a highly reliable organization, HRT proponents would argue that the accident occurred because the organization had ceased being reliable in that it had not followed recommended processes.

Finally, there are those who propose a third way. For example, Leveson et al. (2009) 'believe that the debate between NAT and HRO can become a more productive three-way conversation by including a systems approach to safety emerging from engineering disciplines.' They propose a systems theoretical approach in which safety is viewed not as a component of building a system, but as an emergent property of one properly built. In addition to this, control rather than component failure is claimed to be the cause of most major complex system accidents.

V. Which approach best fits, from an information security perpective?

Though the initial subject of this analysis, the HRO approach is substanitally flawed. It suggests—depending on which particular flavour is being considered presently—either that an organisation, operating within a specific subset of industries, aim to be reliable above all else, or that it try to adhere as much as possible to an ideal set of organisational techniques. The former case does not say much about whether the organisation is actually secure; an attacker can have compromised a company's network years prior and gone unnoticed in the time since, exfiltrating sensitive information. In this case, the reliability of the company is unlikely to be effected, but one would be unlikely to call their situation ideal. When the CIA (confidentiality, integrity, availability) triad is applied, it becomes clear that the reliability of an organisation is only one way in which it can be impacted by a cyber incident. The latter strain of HRO theory, as previously mentioned, is almost too generalised to be of any use in

the information security, or any other, perspective. However, as demonstrated earlier in the analysis, it *can* be applied to get a broad sense of what good reliability practice within information security may look like.

Meanwhile, the NAT approach states that accidents *will* happen at a certain level of complexity, and that we must accept that. Pessimism certainly has its place, but when considering the stakes with which we are dealing here, capitulation to their inevitibility seems a near-suicidal response. The main claim also does not appear to be borne out in reality—Leveson et al. (2009) point out that 'there has never been an accidental detonation of a nuclear weapon in the 60+ years of their existence.' The takeaway from NAT that can, however, be applied in information security is that no one element of security should be considered inviolable. Layered security and defence-in-depth must be considered in order to ensure that any incident can be minimised and stopped.

Ultimately, the systems-driven approach appears to be the most relevant here. Safety cannot be an intention, but a consequence. Risk assessment is an imperfect process, and so one should not rely on its proposals as an end themselves, but as a means to that end. Only once all components of the system are as secured as they can be—the passwords complex, the authentication multi-factor, the firewalls established—can safety be expected to arise.

However, there is a major obstacle in the application of any of the previous theories to an information security perspective. A catastrophic accident at a nuclear power plant will be down to a chance occurance of certain conditions, and the primary actor in causing the accident will be the laws of physics. In an information security perspective, a breach is far more likely to be down to a human-led attack. In the former case, the conditions conducive to the accident may be almost achieved, but never fully, and so the accident will fail to happen many times. In the latter, stopping the attacker from using one approach will likely only drive them, if they are sufficiently motivated, to find another attack. Where there's a will, there's usually a way, and there's no will quite like that of an someone with an axe to grind. As the IRA said to Ms Thatcher after the Brighton hotel bombing failed to kill her, 'remember we only have to be lucky once—you will have to be lucky always.'

VI. Conclusion

The theory of the HRO has some application within information security, as we have seen. However, the theory is not without its factionalism, nor its rivals. These rivals, such as NAT and systems theory, are also capable of productive application to the understanding of organisational information security and risk management. Ultimately, however, the most accurate theory likely lies in some synthesis of some or all of the disparate theories. In other words, yes, accidents may be inevitible, but some organisation structures are more resilient to them than others. That these systems are more or less resilient to these accidents may be as much an accidental side-effect of their design as it is an intentional result. In any case, in the light of emerging technologies such as artificial intelligence which some, such as Prof. Stephen Hawking and Elon Musk, claim may pose an previously unconsidered existential threat to humanity, it is clear that there has never been a more pressing time for further research into this field.

REFERENCES

Akamai (2018). *State of the Internet / Security: Q4 2017 Report*. State of the Internet. URL: https://content.akamai.com/us-en-PG10413-q4-17-soti-security-report.html.

Center for the Protection of National Infrastructure (2017). *Critical National Infrastructure*. URL: https://www.cpni.gov.uk/critical-national-infrastructure-0.

Goldsworthy, Ben (2018). 'Meltdown & Spectre'.

Hopkins, Andrew (2007). 'The problem of defining high reliability organisations'. In: *National Research Center for Occupational Safety and Health Regulation. January*.

Hutchins, Eric M, Michael J Cloppert, and Rohan M Amin (2011). 'Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains'. In: *Leading Issues in Information Warfare & Security Research* 1.1, p. 80.

IBM (2014). *2014 Cyber Security Intelligence Index*. Cyber Security Intelligence Index. URL: https://www.ibm.com/developerworks/library/se-cyberindex2014/index.html.

Jacobson, Greg (2015). *5 Principles of a High Reliability Organization (HRO)*. URL: https://blog.kainexus.com/improvement-disciplines/hro/5-principles.

Leveson, Nancy et al. (2009). 'Moving beyond normal accidents and high reliability organizations: a systems approach to safety in complex systems'. In: *Organization studies* 30.2-3, pp. 227–249.

Lowers & Associates (2017). *5 Principles of High Reliability Organizations*. URL: http://blog.lowersrisk.com/5-principles-hros/.

Marais, Karen, Nicolas Dulac, Nancy Leveson, et al. (2004). 'Beyond normal accidents and high reliability organizations: The need for an alternative approach to safety in complex systems'. In: *Engineering Systems Division Symposium*. MIT Cambridge, MA, pp. 1–16.

Nunan, Daniel and Marialaura Di Domenico (2017). 'Big data: a normal accident waiting to happen?' In: *Journal of Business Ethics* 145.3, pp. 481–491.

Perrin, Chad (2008). *Understanding layered security and defense in depth*. URL: https://www.techrepublic.com/blog/it-security/understanding-layered-security-and-defense-in-depth/.

Perrow, Charles (2011). *Normal accidents: Living with high risk technologies*. Princeton university press.

Rashid, Awais et al. (2016). 'Discovering "unknown known" security requirements'. In: *Proceedings of the 38th International Conference on Software Engineering*. ACM, pp. 866–876.

Roberts, Karlene H (1990). 'Some characteristics of one type of high reliability organization'. In: *Organization Science* 1.2, pp. 160–176.

Roberts, Karlene H and Denise M Rousseau (1989). 'Research in nearly failure-free, high-reliability organizations: having the bubble'. In: *IEEE Transactions on Engineering management* 36.2, pp. 132–139.

Rochlin, Gene I (1993). 'Defining high reliability organisations in practice: a taxonomic prologue'. In: *New challenges to understading organisations*. Ed. by Karlene H Roberts. New York: Macmillan, pp. 11–32.

Rosa, Eugene A (2005). 'Celebrating a Citation Classic—and More: Symposium on Charles Perrow's Normal Accidents'. In: *Organization & Environment* 18.2, pp. 229–234.

Rumsfeld, Donald H (2002). *DoD News Briefing - Secretary Rumsfeld and Gen. Myers*.

Sagan, Scott Douglas (1995). *The limits of safety: Organizations, accidents, and nuclear weapons*. Princeton University Press.

Shrivastava, Paul (1986). *Bhopal*. New York: Basic Books.

Shrivastava, Samir, Karan Sonpar, and Federica Pazzaglia (2009). 'Normal accident theory versus high reliability theory: a resolution and call for an open systems view of accidents'. In: *Human relations* 62.9, pp. 1357–1390.

US Nuclear Regulatory Commission (2010). *HRO Safety Culture Definition: An Integrated Approach*. URL: https://www.nrc.gov/about-nrc/regulatory/enforcement/hro-sc-collins.pdf.

Weick, Karl E and Kathleen M Sutcliffe (2015). *Managing the unexpected: Resilient performance in an age of uncertainty*. 3rd ed. John Wiley & Sons.

Weick, Karl E, Kathleen M Sutcliffe, and David Obstfeld (2008). ‘Organizing for high reliability: Processes of collective mindfulness’. In: *Crisis management* 3.1, pp. 31–66.