Flexibility vs. Structure: How to Manage Reliably Continuously Emerging Threats in Malware Protection

Antti Salovaara
Aalto University School of Business
antti.salovaara@aalto.fi

Kalle Lyytinen
Case Western Reserve University
kjl13@case.edu

Esko Penttinen
Aalto University School of Business
esko.penttinen@aalto.fi

Abstract

High Reliability Organizations (HROs) operate in risky and safety-critical environments where failure avoidance overrides cost efficiency and other traditional performance measures. Research on military, air traffic control, and similar domains has identified five key HRO characteristics: preoccupation with failure, reluctance to simplify interpretations, sensitivity to operations, commitment to resilience, and underspecification of structures. There are fewer studies on digital technologies' role in HRO operations. We address this gap with a case study in a leading malware (e.g., anti-virus) protection firm, which must establish high reliability in its digital operations. While the daily influx of millions of samples and the continuous mutation of malware attacks requires large-scale automation in malware protection, it also calls for continuous fine-tuning and re-engineering through human intervention. We examine the constant balancing of automated and human effort driven by the preoccupation with potential hidden vulnerabilities. This provides a starting point for conceptualizing "digital HROs" as a new research domain for organizational research.

1. Introduction

High reliability organizations (HROs) must operate in nearly error-free manner. They are necessary and common in hazardous business domains such as air traffic control systems or aircraft carriers. A common feature among HROs is that these organizations remain invisible to the public until they face a failure [20]. When a major failure takes place, the consequences are fatal and reach a high visibility. Studying HROs is important, because they provide a window on a distinctive set of processes that foster organizational effectiveness under trying conditions [29].

Most HRO studies to date have been conducted in command-and-control organizations such as aircraft carriers [22], air traffic control [18] or in time-critical infrastructure components such as power plants (e.g., [4,14]). These organizations can be considered extreme cases of HROs, given that their operations are life-critical: failures lead to deaths and natural disasters. One element in the study of HROs has been the effect of digital technologies to HRO based operations [5,13]. Most of these studies have analyzed the impact of IT based operations for organizational reliability in construction [13] and military [9]. We know of no study that has looked at HROs in domains where all operations are fully digital such as in Internet business or software security.

In this paper we seek to address this gap and examine the extent to which HRO-related properties are manifest in organizations whose risks are not in their physical operations, but buried in digital material and related operations. Digital material is highly abstract, flexible and demands attention to detail. At the same time it often comes in huge volumes. Nonetheless, digitized processes are critical to many organizations and failures to guarantee 'right' operations in digital material may halt organizations. Because of the intangible and networked nature, automaticity and high volume of digital operations, dependency on third-party infrastructures, hidden nature of software failures, and other factors, management of digital risks may be different from risks that have been observed in "traditional" HROs. Examples of situations where failures in digital operations have brought entire organizations to their knees include suspected computer virus infestations in South Korean banks and broadcast companies in 2013 (www.bbc.co.uk/news/world-asia-21855051) and the UK coast guard in 2004 (news.bbc.co.uk/1/hi/technology/3682803.stm).

The perceived lack of understanding of how high-reliability organizations operate in the digital domain motivated us to investigate reliability of operations in malware protection business. This business covers anti-virus software development and related services. We conducted a case study on how a malware protection company balances between automation of its operations and manual interventions to cope with the vast influx of anti-virus data samples received from its clients' computers. This is further challenged by a continuously mutating nature of the viruses and other malware. Reacting to these threats demands high reliability and gives us a motivation to define a "digital HRO" and identify its key operational characteristics.

2. High reliability organizations and collective mindfulness

Understanding organizational sources of critical accidents and disasters has intrigued researchers since 1980s. Two streams of research have sought to conceptualize and shed light on the organizational aspects of such accidents: High Reliability Organization (HRO) and Normal Accident Theory (NAT). NAT argues that regardless of the effectiveness of management and operations, accidents in systems characterized by tight coupling and interactive complexity will be inevitable ("normal"), because they cannot be foreseen or prevented [16]. Tight coupling reduces the ability to recover from small failures before they escalate. In contrast, loose coupling allows more efficient recovery ([16], page 160). The literature on HROs, on the other hand, takes a less deterministic stance and argues that high-risk organizations can function safely despite the ever present potential of hazards within complex systems [23].

Collective mindfulness has been defined as a distinguishing characteristic of such HROs. It has been traditionally used to characterize operations in nuclear plants, air traffic control, hostage negotiations, or organizations dealing with any high-risk tasks [2,16,19,20,26,27,28,30]. Despite dealing with high interactive complexity and tight coupling, HROs can consistently display nearly failure-free operations. The success of HROs is largely attributed to the five (cognitive) processes, which jointly "create a rich awareness of discriminatory detail and facilitate the discovery and correction of errors capable of escalation into catastrophe" [29,30,31]. They are:

1) *Preoccupation with Failure*. HROs display a chronic concern about failures, or potential surprises, and interpret close calls as cautionary lessons and opportunities for learning [29]. In the case of malware operations such preoccupation is inherent in the task.

2) *Reluctance to Simplify Interpretations*. HROs maintain divergent points of view and are prone to healthy skepticism, which minimizes their "blind spots" and enables them to attend to small anomalies and early warnings before they escalate. In the case of malware operations small differences may account for significant effects.

3) *Underspecification of Structures*. HROs form "garbage cans" [6] in which problems fluidly migrate to experts capable of solving them. The malware operations need to address issues of scope and scale yet be flexible in addressing new challenges.

4) *Commitment to Resilience*. HRO members cope with "surprises in the moment" by swiftly generating a variety of new action responses and by recovering via "improvisation" or "bricolage" [2,29]. This calls for flexibility and willingness to deal with new threats in novel ways.

5) *Sensitivity to Operations*. HROs, at any moment, collectively comprehend operational detail and develop a holistic picture of their operations [26,29]. In the case of malware focused organizations they need to understand holistically how to address malware detection and risk and understand the effect of specific operations on the organizational outcomes.

In their review of HROs, Levinthal and Rerup [11] point out that most studies have reflected a one-dimensional view of HROs, which neglects complementary interactions between the mindful and the mindless or the rigidity of some of the operations and how to deal with automation in the case of malware operations while at the same time permitting flexibility. Such interactions are often hidden when studies associate mindful behaviors with positive outcomes and mindless behaviors with negative outcomes (e.g., [24,26]). At best, HRO studies merely note the co-existence of the mindful – the flexible – and the mindless – the rigid (e.g., [2]) or describe them as being in sequential alternation (e.g., [12]).

This five-part conceptualization of HRO operations [29] does hint at the necessity of dynamically balancing antagonistic requirements in order to achieve reliable performance [16,20]. For example, quickly creating reliable responses can benefit from centralization and convergent decision-making, but at the same time, coping with uncertainty can benefit from decentralized and divergent decision-making.

2.1. IT and HROs

IS scholars have for some time studied how IT supports the phenomena of distributed cognition (e.g., [3]), which is a medium of HRO based operations. These studies have revealed the polarizing tendencies of centralization and decentralization, or uniformity and diversity, that often accompany IT use [1,21]. Despite such findings, IS research has not explored how IT artifacts relate to or influence HRO like operations. Most relevant is Grabowski and Roberts's [9] discussion of virtuality in HROs but it contains only a few scattered observations about how IT "glues" together increasingly fluid organizational structures. One reason for this oversight may be that IT is often treated as a passive element of an organization's "technical core" [25] and not as a socio-material, evolving and malleable component uniquely appropriated by organizational actors.

Two exceptions can be noted. The first one is the work of Butler and Gray [5], who suggested that reliable IT systems, ironically, can promote the mindless by enabling efficient and routine behaviors, while unreliable IT systems can promote the mindful by encouraging "individuals to seek out multiple information sources and critically evaluate the data upon which they rely" (p. 221). In a recent study Luo et al. [13] studied the impacts of 3D modeling technologies and showed how the same information technology (IT) capabilities were enacted as multiple, contradictory technologies-in-practice which promoted both the mindless – rigid part of the operations as well as the mindful – flexible part of the operations thus managing the tensions inherent in HROs. Our present study develops this initial insight further and describes how digital capabilities are appropriated to automate the mindless while leaving room for human adaptation to support the mindful. In particular, we focus on the specific effects that digital technologies have on HRO operations when the task environment – in contrast to other previous studies – is fully digital and involves malware identification and protection.

3. Anti-virus companies as HROs

Computer viruses form one of the sources of failures in digital operations. Organizations seek protection from such threats using multiple means, the most important one being the use of specific software that detects and protects against such digital threats. Protection is, however, difficult and requires extreme levels of technical competence. Therefore practically all organizations acquire this service from malware protection companies who develop related software and services. By being in the forefront of malicious software developers, these companies face constantly high risks of failure of *not* detecting or *protecting* both their own and their clients' operations.

In this section, we describe some of the operations in malware companies to illustrate the relevance of the HRO concept in digital operations as well as those of their customers from malware. This forms the basis for a development of a concept of digital HROs.

3.1. Malware protection

Malware protection differs drastically from traditional HRO operations in terms of the source and context of threats to cope with. While planes or uranium rods in air traffic control and nuclear power plants, respectively, are not intentionally malicious (unless a plane or a plant is hijacked), this is not the case in malware protection. Cyber criminals develop malware by attempting to use all possible means and their wit to obscure their operations, make the software appear normal, and make use of security vulnerabilities in unexpected and creative ways. The motives of the attacks have also changed during the last years: what used to be hobbyists are now actors driven by money (criminals), personal cause (hactivists), or national interest (nation-state attackers).

Malware protection operations have also grown increasingly difficult to carry out due to constant change in the computing environment. Networked computers and other devices are now easily accessible for attacks. In addition, the proliferation of multiple platforms requires different protection approaches. While in the past practically all viruses were Windows-based, the current array of platforms includes in addition OSX and Android, both of which operate in different ways and present different vulnerabilities. Finally, the number of new viruses has steadily increased due to increased activity in the virus programmer community, related tool automation and their ability to create increasingly complex self-mutating viruses that do not appear in one form only.

Malware companies can react to these challenges both proactively (i.e., through "threat hunting" which involves counter-espionage within the limits of applicable laws) and reactively, by developing detection and malware removal algorithms that their clients can use as part of their infrastructure. Software security companies also share material and knowledge with their competitors on a regular basis. Our focus next is on the reactive operations such as failure mitigation that deal with identification and protection: the key of HRO based digital operations in these companies. Space prohibits us from addressing proactive operations such as failure detection and threat hunting.

Typically, anti-virus software contains three protection layers. The outermost "reputation" layer monitors the Internet sites that the user visits and forms by which the computer retrieves data. If a site's address is unknown, the anti-virus software queries the site's reputation from database run by a malware protection company. If the site is cleared, the retrieved data is next examined in the "detection" layer, often called the computer's 'firewall'. If the data does not appear to contain malicious content, it is allowed to enter the computer. Finally, the "behavior" layer monitors any suspicious actions taking place in the computer and activates a removal mechanism, if any pattern of malicious behavior is detected. All these operations are automated and involve communications between the anti-virus software and the anti-virus company's servers. This way the company receives information about new threats and can also monitor the effectiveness of their anti-virus updates.

Due to the extensive volume of real-time data that needs to be reacted quickly (from seconds to a few minutes), most of the operations in the malware protection are fully automated. This does not only involve responses to queries arriving from antivirus software running in millions of computers, but also the analysis of incoming data samples (on average a million per day) that need to be examined. Improving reputation, detection and behavior layers in clients' software is challenged by the unknown nature of the emerging threats.

We paraphrase the U.S. Secretary of Defense Donald Rumsfeld's famous note on WMD that threats cover both "known unknowns" and "unknown unknowns". By known unknowns we refer to known vulnerabilities in the existing protection software. For example, in some cases the company needs to allow access to some web addresses that it knows are reputable (e.g., banks or search engines), but whose contents it cannot completely analyze due to their dynamically changing nature (e.g., a news feed). The company must accept that it cannot block its customers from these sites, but must monitor them closely and improve protection in the other layers.

By unknown unknowns we refer to those threats that the existing systems are not designed to detect and which the company is not yet aware of. Such threats may exist due to so far unknown security vulnerabilities in commonly used software (e.g., web browsers, security protocols and the like). Typical means to counter such threats are in-house proactive research and use of the three-layer protection hierarchy that is able to notice malware in mutually complementary ways.

3.2. Digital HRO challenges

Malware protection operations are a good example of HROs and meet well the five characteristics of HROs. Malware protection companies are continuously preoccupied (1) with threats of viruses and other malware that may endanger their own and customers' operations. The presence of a creative, rogue community preempts a possibility of ever having a robust protection from the threats, and this makes companies reluctant to simplify their interpretations (2). These threats must be solved typically in a flexible manner through adaptive escalation processes without predefined hierarchical structures (3). Similarly, malware protection companies must be committed to resilience (4) in improving their protection without outside help when being faced with dangerous viruses, and stay vigilant and sensitive to their operations (5) to understand the present state of threats.

Yet, there are new threats that the digital nature of operations presents to the malware protection companies. First, threats are highly heterogeneous, not only because of the creative efforts of the rogue community, but also because dependencies between today's computer system components are complex and evolving, and afford innumerable variations of alternative "infection vectors" (i.e., creative ways in which systems' vulnerabilities are exploited). Second, digital threats are not directly observable and can remain easily unnoticed without extensive tooling and analysis. They may also reside longtime invisible. Third, threats often present themselves in disguised forms, both intentionally and unintentionally. In malware context, a virus's program code is usually "obfuscated", meaning that a programmer has made its interpretation intentionally difficult. By unintentional guises we mean, among others, all the software bugs that first appear to be of one type (e.g., memory problems) while they turn out being something else (e.g., data type problems). Thus, while it is rare that in air traffic control threats would result from clouds appearing as planes, in digital contexts these are common sources of threats. Fourth, because digital information can travel rapidly, threats propagate and replicate themselves at exponential rates that would be impossible in physical systems.

Notably these challenges are not specific to malware protection but apply to digital business in general. The challenges presented above pose risks with catastrophic consequences for many companies (consider, e.g., Target's loss of customer data in 2013 Christmas sales). In the following case study, we will show how digitalization challenges provide a new perspective to the five traditional HRO characteristics.

4. Case study

To understand how a digital HRO functions and how it organizes its operations, we studied the malware protection operations through a case study within a successful, high-quality security software company F-Secure (www.f-secure-com). F-Secure has offered highly reliable security software services for 25 years and is widely acknowledged as one of the technology leaders in the market per AV-Comparatives listing for successful security software providers. F-Secure currently employs around 940 employees in 20 offices around the world and is headquartered in Helsinki, Finland (2014 data). These figures make it the largest malware protection company in Europe.

F-Secure has organized its malware response operations into three 8-hour shifts. All those shifts are

Table 1. Interview participants

Phase	Informant	Experience (years)	Work location
I	Director	10	HQ
II	Senior Researcher	14	HQ
II	Service Owner	7	HQ
II	Service Owner	7	HQ
II	Director	10	HQ
III	Director	10	HQ
III	Service Owner	7	HQ
III	Service Owner	7	HQ
IV	Team Lead	9	Offshore
IV	Team Lead	7	Offshore

working from the Kuala Lumpur office. If the shift (level 1) cannot handle a certain case, then that case is *escalated* to the senior experts first in Kuala Lumpur (level 2) and then in the Helsinki office (level 3). In addition to this response unit, F-Secure operates a research lab, which is responsible for developing new malware protection services and conducts research on different types of software threats.

As a response to the dramatic rise in the amount of malware threats, the role and impact of automation in protection operations has increased in scope and importance during the last decade. F-Secure now uses automation in all of its operations to improve efficiency, support analysis, decrease the risks of human error, and to manage the high volume of incoming samples, among others. Ten years ago most of the samples were processed by human intervention; today, all cases are initially manipulated by a digital platform that uses in its core a rule-based engine to identify different types of threats from a stream of samples. Only a small portion of the samples is escalated to the response team for human intervention.

We collected the data in four phases (see Table 1). Phase I consisted of an interview of the director of research to get an understanding of the structure of operations at F-Secure and to gain access to the informants. In Phase II, we conducted four interviews with the key informants of the response and research units to understand how these units function together to provide security services to the market. In Phase III, we carried out three interviews with the Helsinki office response unit's team leaders. Here, we focused on the five characteristics of HROs and how they are manifested in the operations. In Phase IV, we made two additional interviews with Kuala Lumpur response team leaders. These interviews focused on the nature of digital reliability operations and how the response teams balance between structure of automation and flexibility required by the threats. In all phases, all informants were first asked to present their views in a free format and only in the later stages of the interview we presented the informants with the HRO framework to get their insights on the specific question (e.g., on the five characteristics of HROs). Table 1 depicts the informants and their roles in the company.

The interviews lasted 1–1.5 hours each and were all tape-recorded and transcribed. The interview data consisted of 135 pages of text, which were analyzed using the NVivo software. Two of the authors coded the data following the five characteristics of HROs [29] as well as to find evidence of the operating environment and the role of technology in the operations of F-Secure. Because of a small sample size and our promise of anonymity, we cannot offer detailed information about interviewees in our quotes.

5. Findings

The analysis of the interviews highlighted three aspects that define F-Secure's malware operations and render it a good example of a digital HRO.

5.1. The state of constant anomaly

Experts at F-Secure are continuously aware of the fact that their systems cannot detect all the threats perfectly. At any given point of time, F-Secure's response unit works on improving its detection capability of 1–10 malware families. Often these families are mutated versions of the same virus. When a malware family is analyzed and the company knows that there are weaknesses in that part of the detection layer, F-Secure must monitor related vulnerabilities in more detail and make decisions whether to act conservatively or permissively with the possible failures that these weaknesses may cause. Adopting a conservative strategy makes their system detect also false positives (i.e., samples that are falsely recognized as malware), meaning that users will be blocked from content that would actually be safe. A permissive strategy has the opposite outcome: the system will accept false negatives (i.e., samples that are checked clean when they, in fact, are malicious). As we already noted, permissive strategy is sometimes necessary, because users cannot be blocked from websites that are critical to their business (e.g., online banks) only because there may be a small chance of vulnerability.

These and other kinds of threats create a *constant anomaly* in F-Secure's malware operations. In other words, the system is known to have a potential of failure all the time, and the company must put effort in decreasing the likely failure level to a minimum. However, because new types of malware emerge at a frequent (but unknown) rate, the situation will never stabilize. On a more detailed level, this anomaly involves dynamic balancing between known unknown and unknown unknown threats. The vulnerabilities caused by unanalyzed sample families, as described above, present known unknown threats. In these cases F-Secure is aware of the potential vulnerability and can actively work towards removing it. Unknown unknowns, in turn, are problems that may be introduced into the detection systems in multiple ways: as unintended side-effects when the detection logic is changed, as new kinds of malware that the company is not prepared to detect, human error, and in rare cases, bugs in the underlying components in the operating system on which the malware protection operations rest.

> *“[The operating environment] is terribly instable in a sense that the world around us is changing extremely fast. That you think that there is a need for computer security and there are actors who try to bypass your shields. You have to react all the time, be in a state of emergency, follow what’s going on, follow the feedback around the world and the field, so to say, as fast as possible.”*

> *“In game of cat and mouse, so we develop certain technologies to prevent this kind of infection to certain known malware within this week. Then the bad guys come up with new innovations to come up with that. Sometimes they are able to do it so that we have to really do big restructuring on our structure and process to be able to battle that.”*

The state of constant anomaly is related to the concept of preoccupation with failure and consequent reluctance to simplify interpretations – two of Weick et al.’s [29] five HRO characteristics. In particular, we observed that experts at F-Secure were continuously aware that their system has potential vulnerabilities and that all the changes made to the system are provisional and may therefore need to be changed again once new intelligence about malware has been acquired. Therefore F-Secure sometimes releases improvements in a careful incremental fashion in order to see whether the changes are having the expected effect, thereby avoiding simple interpretations or attempts to develop once-and-for-all solutions:

> *“We trust it [an internal test process] to a certain level definitively. So we need to give up, aahh, maybe give up is not the best word, we need to agree on a certain threshold, that we trust that this much, so that we can actually release this process. And that has gone through a lot of testing, people looking at it. So that we can establish that ok this is good enough for us to use.”*

5.2. Balance of structure vs. flexibility

Due to the presence of constant anomaly, the threat situations are in a continuous move, and F-Secure as a HRO must adapt quickly to these emerging threats. Earlier literature [29] identifies a paradox that dealing reliably with such threats requires orderly procedures (structure) and flexibility in operations (underspecification of structures). Finding a balance between the two emerged as a constant theme in our interviews and characterized F-Secure’s operations. In our case, structures do not connote solely manual routines or strictly prescribed processes, but also and primarily the creation and enactment of an automated workflow for managing incoming samples. Structure is, therefore, largely dictated by the needs of automation to handle the massive scale and variety of incoming malware samples.

> *"So we have expert rules so that we try to define the rules and we try to rate it as malicious if it falls under this category and then automatically rates it."*

> *"... but we have certain things we must do in a structured way. There are some legal issues in treating malware samples and merely ethics that simply oblige us to have rules and structure for doing certain procedures."*

Flexibility or underspecification of structures, on the other hand, is mainly required to stay innovative and agile. This requirement arises from the pace of change in the industry and the diversity of threats.

> *"The pace of change in technology makes us to have temporary structures"*

> *"Flexibility is important as well. Because if you are going to be too structured, then it kills innovation and the possible changes you can do, improvements in the process."*

Due to the massive amount of incoming malware samples (around 200,000 per day), it is crucial to F-Secure to link expertise – whether codified, shared, technical expertise embodied in the rule engine or human expertise that is sought for through escalation – to the problems (treating incoming malware samples) efficiently. Automation treats the majority of the incoming malware samples and a small part of the samples is escalated to the response teams, requiring human intervention. Escalation of the samples from the rule engine onto the response team acts as a "safety valve" for the structured workflow. Therefore, the required "underspecification of structures" and agility is achieved through escalation.

Overall, information technology serves a dual role in balancing between structure and flexibility in the case company. The rule-based engine gives structure to the workflow by enabling the automation of detecting, organizing and reporting a vast majority of the incoming malware samples. At the same time, the rule engine allows quick and flexible changes to the workflow by enabling entry of new rules and conditions when new profiles of malware are detected, thereby supporting swift flexibility. In other words, the rule engine acts as a middle-ground under-specification device: on one hand it is a fixed knowledge infrastructure that can be enacted to identify and operate on threats but on the other hand it is swiftly expandable as the rules are easily adaptable:

> *"Instead of writing code to automate the treatment of incoming samples, we have a rule-based mechanism, which allows us to change our workflow in a flexible manner if needed, even during the day. And we do that often if there is something wrong with the workflow. I simply write new rules, test it, and push it to production. This allows us to do one year's development in one day.*

5.3. Rapid codification and scalability

In addition to the threats posed by the different types of unknowns (see above), another risk for F-Secure comes from the sheer volume of data. Millions of installations of anti-virus software on clients' computers create a constant flow of samples, with a requirement for an instant verdict on their maliciousness. The global coverage adds another challenge, with a need for understanding different modes of operation both among customers and among malware communities. The risk is that even when the emerging malware samples are not challenging as such, their sheer volume may outnumber F-Secure's ability to respond appropriately.

In line with the strategy of aiming for the optimal amount of flexibility, F-Secure's solution to this threat has been to develop a highly rapid knowledge codification process. While traditional knowledge management (KM) models, informed by studies on R&D and innovation teams (e.g. [7,10,15]) have observed a significant role of tacit knowledge in successful KM, the F-Secure's process bypasses the slow codification in the communities by turning new knowledge directly into detection rules whereby related detection operation can be automated:

> *"The main documentation for the rules is actually in the rules. When someone edits the rules he sees in the comments what it does"*

In practice, when an expert starts analyzing a new set of suspicious samples, she or he starts by attempting to classify the threat into specific detection rules from the very start. Such rules will then be able to analyze whole families of samples with a single expression, and output varying verdicts that can be used in further examination of the threat's nature, both automatically and with human inspection. This information can be readily used in implementing an improved automated protection. In typical cases through this fast codification process, F-Secure manages to react to new types of malware within a few hours.

By emphasizing the role of rapid codification to externalized representations of malware (e.g., samples and rules), we do not mean that tacit knowledge does not have significant role in F-Secure's operations. In fact, constant recognition of new "threat vectors" demands intuition and tacit knowledge. Overall, the knowledge about new threats does not need to be fully internalized before expressing it computationally.

6. Discussion

In this paper, we have analyzed HRO operations in a fully digital environment. This is a domain, which has been largely ignored in the existing HRO literature. The lack of attention motivated us to analyze one leading malware protection company so as to learn about the characteristics of "digital HROs". We will next discuss how digital HROs constitute a new category of HROs.

6.1. Digital HROs

The digital nature of the threats creates some particular conditions that are not captured in standard HRO characteristics. We can synthesize the differences to the standard HROs with the following characteristics:

1) *Increased awareness of unknown threats in the system.* While also traditional HROs' actors experience malfunctions and show weaknesses (e.g., human fatigue or machines' wear and tear), such behaviors are more easily predictable than in digital HROs. In the latter, systems unavoidably contain unknown bugs[1]. Because of the possibility of hidden bugs, an organization facing a threat cannot easily assume that

[1] An often stated claim among programmers is that a typical computer software contains one bug for every hundred lines of code, and one exploitable bug per every thousand lines (e.g., security.stackexchange.com/ questions/21137/average-number-of-exploitable-bugs-per-thousand-lines-of-code).

the threat can be straightforwardly resolved. The awareness of such threats is therefore a particular characteristic in which digital HROs portray heightened preoccupation with failure and reluctance to simplify interpretations – two (1 & 2) of the original HRO characteristics.

2) Awareness that *automation can quickly escalate the failures*. If the organization's core operations are automated, its failures can be observed only indirectly and may remain hidden longer. In addition, due to the demands for high computing capacity and network bandwidth, observed failures will fast reinforce themselves (as has been observed several times recently in computer-operated algorithmic stock exchange trading). A digital HROs must prepare for this by being able to modify its operations quickly and, if necessary, return the operations under manual control. Therefore digital HROs must heed attention to identifying a suitable balance between the structure and flexibility in its operations. This is an example of the need for maintaining an underspecification of structures – (3) in the original HRO definition.

3) *Use of specialized personnel for threat management.* While in a traditional HRO the same members of the organization both manage the operations and repair them, in digital HROs highly specialized IT specialists are required for carrying out repairs. This may call for a dedicated unit or rare technological expertise throughout the organization – as is the case with F-Secure. This is an example of the characteristic of the commitment to resilience – (4) in the original HRO definition.

The fifth original characteristic – sensitivity to operations – was not observed in our case study. This does not mean that sensitivity would not be important in other digital HROs. We assume that in our case this characteristic did not emerge because it is related to real-time human management of a continuously changing threat situation. In our case all this real-time management had been delegated to workflow systems while the human component was allocated to offline analyses. The absence of extensive remarks on sensitivity to operations raises a question whether digital HROs in general tend to delegate such operations to computational systems, thus warranting an additional digital HRO characteristic. This would be the negation of the original HRO characteristic. This will be a relevant question for further research.

Yet, our definition of a digital HROs is tentative and in need of improvement. In particular, it needs to be triangulated with other studies of HROs operating in fully digital environments such as financial services. Though our analysis is based on only a single case study, it warrants discussion on the following implications.

6.2. Digital HROs and lean management

During our analysis, we became aware of several similarities between the principles of lean management and the management of systemic risks in digital HROs. First, the careful monitoring of the effectiveness of rule updates resembles the principle of validated learning in lean management [17] that suggests that operations should be changed with carefully planned improvements whose effectiveness can be immediately validated. In lean management this often involves split tests that allow for A vs. B comparisons. In our interviews we did not hear about such procedures at F-Secure, but we contend that they could be applicable in sandboxed testing (i.e., in analyses carried out in a secluded environment that does not pose harm for company's actual operations or clients).

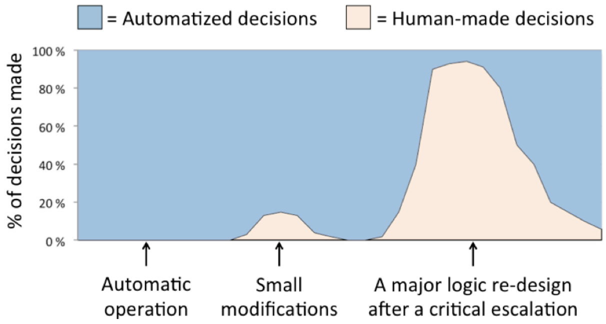
Second, lean management suggests that companies should aim for customer satisfaction and be willing for compromises instead of clinging to strategies whose success can be confirmed only after considerable investments. In F-Secure we observed that the company is sometimes willing to compromises with the benefit of customer satisfaction. This happens particularly when customers cannot be blocked from high-traffic websites even when their safety cannot be ultimately guaranteed. A high-investment strategy would be to always gather extensive intelligence before allowing access to a potentially harmful website.

Third, while lean management's recommendation to take risks in order to learn rapidly appears an antithesis of HROs avoidance of failure, digital HROs may present an exception to this. Digital HROs, by being able to gather and store authentic data from their operations, can develop minimum viable products [17] and test them safely with extensive realistic simulations. F-Secure's detection rules in the rule engine were examples of such minimum viable products. Similarities between lean management and HROs have already been observed in occupational health management [8], but we find it interesting to extend studies also to digital HRO operations.

6.3. Interchangeability of responsibilities

Our findings concerning the balance between flexibility and structure show principally a relationship between human and automated work processes and warrant a closer analysis. At F-Secure the flexibility of the detection logic lets the company divert, shut down, bypass and in other ways manipulate the automated flow of samples so that varying human interventions become possible. Alternatively, the experts may patch

Figure 1. The fluctuation of interchangeable responsibilities



and improve the system's logic tests while it is kept all the time in operation (of course, given that the changes pass the necessary unit, functional and other test runs). We call this *interchangeability of responsibilities*: that the operational logic can be carried out jointly by automated and human efforts, and that this balance may change from one time point to another.

We can consider a continuum that designates the extent by which humans vs. automation take the responsibility for operational decisions. At any given time, the sum of percentages of human-made and automated decisions is 100. In the 0/100 extreme, there is no human intervention and the automatic logic takes the responsibility of 100% of the decisions. The middle range in the continuum (e.g., 30/70) represents the situations where humans are making some percentage of the decisions. At F-Secure this corresponds to ordinary rule modifications, each modification affecting a fraction of a percent of all the active decision logic. This is the state where a constant, but dynamic anomaly prevails and is managed by experts' active involvement. Finally, the 100/0 extreme would represent a rate when a threat or a failure would force the company enter into a completely manual control. Cases that are close to this are the major and unknown escalations: the ones in which a significant sub-component is being re-programmed and carefully monitored. These three different scenarios are presented in Figure 1.

While quantifying the exact percentages for human modification's extent may be complicated, this conceptualization of risk management may help understand the nature of threats faced by a digital HRO. High percentages in human interventions would be a sign that the company's IT infrastructure does not offer enough flexibility and underspecification of structures, because drastic changes to it are frequent. A constant rate of small modifications would be a sign of a suitable balance between flexibility and structure. More data on modifications is needed to investigate these and other assumptions, however.

6.4. Digital HROs as sociotechnical systems

The interchangeable nature of responsibilities, as described above, opens an important issue how we can appreciate the divergent roles of humans and IT in digital HROs. Interchangeability suggests that IT and humans must be understood as complementary parts of the same unit of analysis, though this is not always the case in the past reliability research where the technology is often seen as a subservient tool for humans and only a source of unreliability and mindlessness [5]. Based on our analysis, however, both possess agency and participate in decision-making, which can have both positive and negative reliability consequences [13]. Both humans and technology shoulder dynamically the cognitive workload and neither of them can be seen as a primary actor in the organization's strive for high reliability. In this regard our analysis extends that of Luo et al. [13] as it demonstrates in detail how high reliability operations use digital technologies simultaneously to address the scale and improve flexibility. Our analysis also shows how reliability operations can be distributed digitally geographically in ways that have not been discussed in the past high reliability research, which has mainly focused operations in smaller physical locales like control rooms or cockpits.

Finally, our definition of digital HROs identifies automated decision-making as one of the key characteristics of digital HROs. This feature was a common trait across all digital operations we examined – not only HRO-related ones. However, in a digital HRO context, the ability of technology to act intelligently and make decisions on human's behalf in critical operations makes digital HROs stand out from traditional HROs. The focus on analog, physical operations in the previous HRO literature has, indeed, hindered researchers from noticing that their conceptualizations posit only humans as decision-makers in all central questions of reliability, whereas in our case it is the overall socio-technical system and how it distributes decision rights and capabilities.

In our case, we failed to observe the presence of sensitivity to operations. The reason for this is that when IT systems take responsibility for those operations (i.e., time-critical real-time response) most of the detail of the operations becomes blackboxed after initial specification. This is one example where erecting a boundary between human-based and machine-based operations becomes arbitrary. In fact, our study suggests that the concept of a HRO may be in need of update in digital environments.

7. Acknowledgments

AS acknowledges support by Academy of Finland through funding for Creative use of ICT project.

8. References

[1] Applegate, L., Austin, R., and McFarlan, F. *Corporate Information Strategy and Management: The Challenges of Managing in a Network Economy*. New York: McGraw Hill-Irvin, 2007.

[2] Bigley, G.A. and Roberts, K.H. The Incident Command System: High-Reliability Organizing for Complex and Volatile Task Environments. *Academy of Management Journal 44*, 6 (2001), 1281–1299.

[3] Boland, R.J., Tenkasi, R.V., and Teeni, D. Designing Information Technology to Support Distributed Cognition. *Organization Science 5*, 3 (1994), 456–475.

[4] Bourrier, M. Organizing maintenance work at two American nuclear power plants. *Journal of Contingencies and Crisis Management 4*, (1996), 104–112.

[5] Butler, B.S. and Gray, P.H. Reliability, Mindfulness and Information Systems. *MIS Quarterly 30*, 2 (2006), 211–224.

[6] Cohen, M.D., March, J.D., and Olsen, J.P. A Garbage Can Model of Organizational Choice. *Administrative Science Quarterly 17*, 1 (1972), 1–25.

[7] Davenport, T. and Prusak, L. *Working knowledge*. Boston, MA: Harvard Business School Press, 1998.

[8] Gnoni, M.G., Andriulo, S., Maggio, G., and Nardone, P. ‘Lean occupational’ safety: An application for a Near-miss Management System design. *Safety Science 53*, (2013), 96–104.

[9] Grabowski, M. and Roberts, K.H. Risk Mitigation in Virtual Organizations. *Organization Science 10*, 6 (1999), 704–721.

[10] Leonard, D.A. *Wellsprings of Knowledge: Building and Sustaining the Sources of Innovation*. Harvard Business School Press, 1998.

[11] Levinthal, D. and Rerup, C. Crossing an Apparent Chasm: Bridging Mindful and Less-Mindful Perspectives on Organizational Learning. *Organization Science 17*, 4 (2006), 502–513.

[12] Louis, M.R. and Sutton, R.I. Switching Cognitive Gears: From Habits of Mind to Active Thinking. *Human Relations 44*, 1 (1991), 55–76.

[13] Luo, J., Lyytinen, K., and Boland, R. Dialectics of Collective Minding: Creating Radical Architecture with Information Technology. *MIS Quarterly 36*, 4 (2012), 1081–1108.

[14] Marcus, A. Managing with danger. *Industrial and Environmental Crisis Quarterly 9*, (1995), 139–152.

[15] Nonaka, I. and Takeuchi, H. *The knowledge creating company*. New York, Oxford: Oxford University Press, 1995.

[16] Perrow, C. *Normal Accidents: Living with High-Risk Technologies*. New York: Basic Books, 1984.

[17] Ries, E. *The Lean Startup: How Constant Innovation Creates Radically Successful Businesses*. London, UK: Penguin Books, 2011.

[18] Roberts, K. Cultural characteristics of reliability enhancing organizations. *Journal of Managerial Issues 5*, 2 (1993), 165–181.

[19] Roberts, K.H., Stout, S.K., and Halpern, J.J. Decision Dynamics in Two High Reliability Military Organizations. *Management Science 40*, 5 (1994), 614–624.

[20] Roberts, K.H. Some Characteristics of One Type of High Reliability Organization. *Organization Science 1*, 2 (1990), 160–176.

[21] Robey, D. and Boudreau, M.-C. Accounting for the Contradictory Organizational Consequences of Information Technology. *Information Systems Research 10*, 2 (1999), 167–185.

[22] Rochlin, G.I., La Porte, T.R., and Roberts, K.H. The self-designing high-reliability organization: aircraft carrier flight operations at sea. *Naval War College Review 40*, (1987), 76–90.

[23] Sagan, S.D. *The Limits of Safety: Organizations, Accidents, and Nuclear Weapons*. Princeton, N.J.: Princeton University Press, 1993.

[24] Swanson, E.B. and Ramiller, N.. Innovating Mindfully with Information Technology. *MIS Quarterly 28*, 4 (2004), 553–583.

[25] Thompson, J.D. *Organizations in Action: Social Science Bases of Administrative Theory*. New York: McGraw-Hill.

[26] Vogus, T.J. and Welbourne, T.M. Structuring for high reliability: HR practices and mindful processes in reliability-seeking organizations. *Journal of Organizational Behavior 24*, 7 (2003), 877–903.

[27] Waller, M.J. and Roberts, K.H. High reliability and organizational behavior: finally the twain must meet. *Journal of Organizational Behavior 24*, 7 (2003), 813–814.

[28] Weick, K.E. and Roberts, K.H. Collective Mind in Organizations: Heedful Interrelating on Flight Decks. *Administrative Science Quarterly 38*, 3 (1993), 357–381.

[29] Weick, K.E., Sutcliffe, K.M., and Obstfeld, D. Organizing for High Reliability: Processes of Collective Mindfulness. *1*, (1999).

[30] Weick, K.E. and Sutcliffe, K.M. *Managing the Unexpected: Assuring High Performance in an Age of Complexity*. San Francisco: Jossey-Bass, 2001.

[31] Weick, K.E. and Sutcliffe, K.M. Mindfulness and the Quality of Organizational Attention. *Organization Science 17*, 4 (2006), 514–524.