Department of Computer Science

The Relative Effectiveness of widely used Risk Controls and the Real Value of Compliance

Ioannis Agrafiotis, Sadie Creese, Michael Goldsmith,
Jason R.C. Nurse and David Upton

CS-RR-17-01

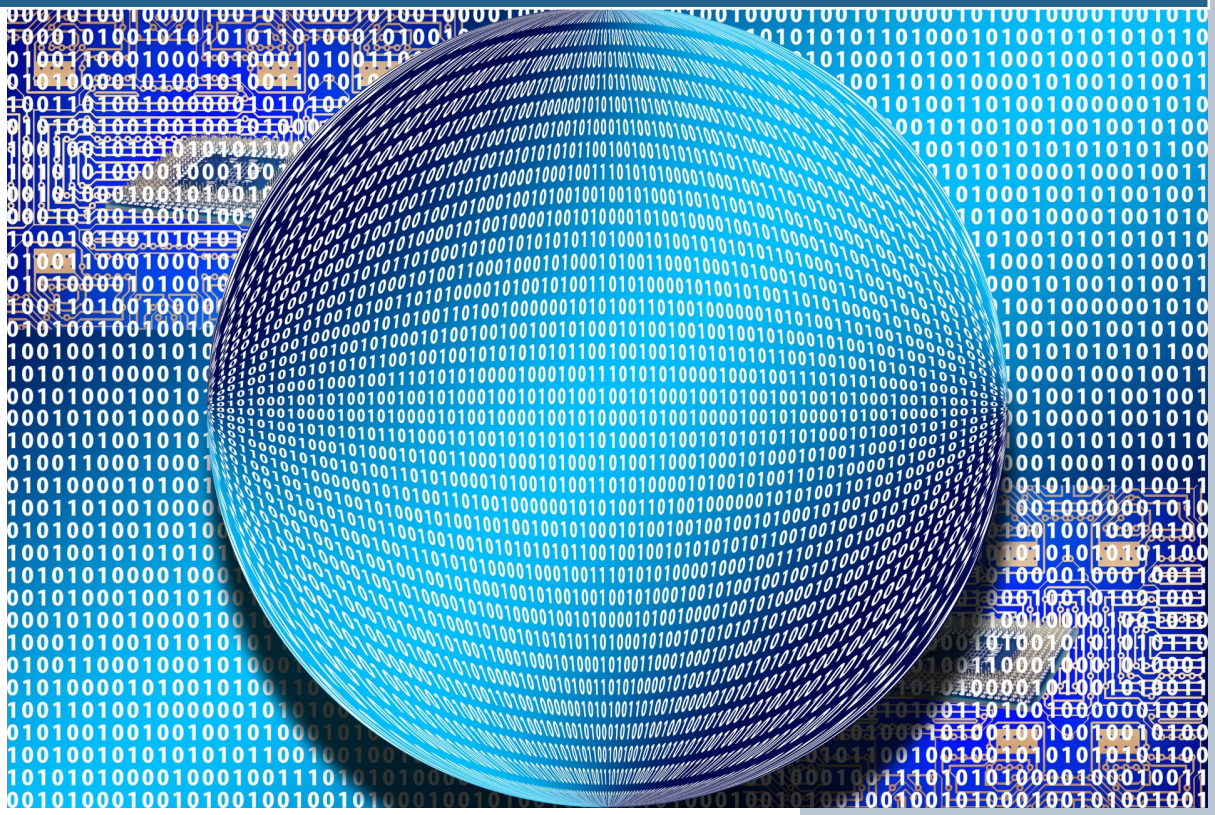
Department of Computer Science, University of Oxford
Wolfson Building, Parks Road, Oxford, OX1 3QD

UNIVERSITY OF OXFORD

DEPARTMENT OF COMPUTER SCIENCE

SAID BUSINESS SCHOOL

The Relative Effectiveness of widely used Risk Controls and the Real Value of Compliance



Authors (alphabetical ordering):

Ioannis Agrafiotis
Sadie Creese
Michael Goldsmith
Jason R.C. Nurse
David Upton

University of Oxford

30/11/2016

This research was sponsored by

Novae Group plc

Novæ
LOOKING FORWARD

Executive Summary

As the volume of cyber-attacks continues to rise and also the levels of harm suffered from them, it is becoming critical that organisations can demonstrate that reasonable efforts are being undertaken to reduce cyber-risk. However, the risk responses and controls typically viewed as necessary, and even essential, by the professional and expert community are generally not underpinned by any framework that facilitates rigorous reasoning, qualification or quantification of the benefits resulting from their deployment. This means that the real value of compliance, or the variability of compliance, to risk-control standards is not well reasoned or measurable in any scientific, unambiguous or verifiable sense.

This further means that methods used to manage, mitigate or transfer the risk by stakeholders across the information security and cyber risk landscape` are existing increasingly in isolation. Our study here has shown that a more rigorous risk valuation and risk management environment can only fully exists where transparent and effective collaboration between stakeholders exists. Only through such collaboration can risk be fully understood, modelled, valued and thus managed.

In this project we explore the value of risk controls to security posture, and so the value of compliance to standards and frameworks prescribing these controls. Our approach in this report centres on the effectiveness of controls, with particular emphasis on determining the residual risk for each control. The residual risk may occur due to inherent vulnerabilities that controls have against specific attack vectors or due to implementation practices. We provide details on how residual risk, a critical factor for determining the effectiveness of controls, may be assessed for each risk control. This report examines the effectiveness through a number of lenses: assets and their attack-surface (which are determined through a number of risk dimensions, with the most important being the Bespoke or Common dimension) and the ability of controls to protect them; controls in the context of inherent and known vulnerabilities; the capability of threats in compromising the effectiveness of controls; and the role controls play in how cyber-harm manifests and propagates. These lenses offer a unique insight into how key organisational contexts can be impacted by the effectiveness of controls. This knowledge is also used to extend our work and the model that we have created to map the relationship between risk controls and organisational concerns such as assets, cyber-value-at-risk (cyber-VaR) and cyber-harm.

Our main contributions in this report can be summarised as follows:

- An extension of our model that defines the associations between risk controls on the one side and assets, cyber-VaR and the different types of cyber-harm which may occur in a typical organisation on the other. In particular, we have significantly extended our research on the relationships between these components, which incorporates general relationships, relationships across the three main levels, inter-dependencies between levels, links to controls and control effectiveness. This has also involved the provision of a reasoning approach and syntax whereby the associations between the various enterprise components could be defined. This would allow organisations to have a better understanding of what harms may be caused to which assets (or set of assets) and how those harms could trigger other harms, and eventually increase the cyber-VaR. Using our proposed reasoning method, we also considered how controls as implemented could protect assets, reduce or limit cyber-harm, and impact cyber-VaR. This links directly to the control's effectiveness, as different controls have different influences on these variables.

- An iniial validation of our model through interviews and focus groups with industry professionals in the areas of cybersecurity and cyber-insurance. To comment briefly on the findings from this stakeholder engagement, we found that:
    1. There appears to be a clear understanding of what critical assets to organisations are and these assets are all successfully covered in our model. Furthermore, organisations determine critical assets based on their importance to key business processes, requirements from legislation and regulators, and harms that may result to the organisation if the asset is compromised.
    2. The model was able to capture a vast majority of the harms that professionals identified as potential impacts from cyber-attacks. However, professionals generally had not considered the full range of cascading harms that can result from attacks, nor did they have well-defined metrics for estimating or measuring harms.
    3. While organisations do apply controls to address specific risks that they face, other key motivators for selecting certain controls include the requirements placed by regulatory bodies, legislation and broad concerns facing the industry (e.g., new types of attackers). Most importantly, many organisations do not have a good understanding of how to measure the effectiveness of controls especially on a real-time basis.
- Focusing specifically on the topic of the relative effectiveness of risk controls, our analysis of the literature (academic and industry-related) and the interaction we have had with industry professionals has provided little scientific evidence to suggest that there exist clear ways to measure effectiveness. To validate the utility of our model therefore and the effectiveness of controls more broadly, we have outlined several data requirements at each level of the model. Data is crucial as it will enable users of the model to reason about numerous aspects including the links between certain assets and cyber-harms (e.g., typical harms that result from certain assets), the likely propagation paths of harms (e.g., specific harms that are likely to result due to other harms), the probability distributions that allude to the likelihood of particular losses, and effectiveness of risk controls. These requirements are outlined and followed by an explanation of how they would be applied in the use of the model.
- The following actions should be considered in order to further this line of research:
    1. Implement a prototype software tool of the model proposed, which would be capable of determining the potential range of impacts of a risk control upon exposure to harm. This might be developed with a selection of estimate probability distributions based upon knowledge in the community, and with which it would be possible to test the sensitivity of results in a range of scenarios.
    2. Design a methodology for learning the impacts of risk controls within an organisation using software sensors with an organisation's infrastructure. This would enable the collection of data to establish the probability distributions required by the Model (in 1 above) through aggregation of results across multiple organisations and identification of general patterns. This approach would have the added benefit of allowing organisations to consider results tailored to their specific operations.
    3. Additionally, further consideration is required of how different datasets may be linked to provide quantitative evidence on how effective controls are. This new approach should take into consideration the interdependency of controls how

effectiveness of the ecosystem of controls may change if certain controls are not present. Further exploration of historic data is required to identify features that will be abstract enough to provide useful information regarding the likelihood of an attack, even when the data is considered obsolete (i.e. on systems which are not used any more).

4. The approach should be extended to address other classes of harm, from natural disaster or accidental insider actions (for example, where our research in other projects leads us to believe that current risk controls are often inadequate).
5. We should also consider expanding the findings from the qualitative research. A possible next step could be to conduct large scale questionnaires, focusing on which controls are widely accepted in the industry and what metrics are used to determine their effectiveness. Additionally, interviews and focus groups could take place to emphasise on how the interconnection of assets may change the way organisations perceive assets, as well as identifying how cascading harm may occur and which types of harm are triggered. Interviewing lawyers will shed light into how recent developments in legislation may influence the way organisations reason about controls and whether cyber-insurance will become a norm, enabling insurers to suggest a set of desirable controls to hedge risk
6. Specific research should be conducted into the relationship between harm, and an assets level of digitisation. We need to know if it is the case that the level of digitisation has a consequence for the likelihood of susceptibility to successful cyber-attack, and the potential for harm and cascading harms within an organisation. New controls might be suggested.
7. Specific research should be conducted into the value of unpredictability in control usage as a mechanism for improving cyber-defenses to reduce cyber-harm.
8. We still need to consider how aggregation and systemic risk affect propagation of harm and how potential for such affects the decisions that organisations make. In particular we should focus on harm propagation across organisations who share common technologies, harm from unavailability of web-services and impact on business interruption.

Table of Contents

Abbreviation List

CIS Center for Internet Security

CSC Critical Security Controls

DPA Data Protection Act

GDPR General Data Protection Regulation

ISO International Organisation for Standardisation

NIST National Institute of Standards and Technology

SMEs Small and medium-sized enterprises

VaR Value-at-Risk

1. Introduction

1.1 Motivation

As the volume of cyber-attacks continues to rise and also the levels of harm suffered from them, it is becoming critical that organisations can demonstrate that reasonable efforts are being undertaken to reduce cyber-risk. However, the risk responses and controls typically viewed as necessary, and even essential, by the professional and expert community are generally not underpinned by any framework that facilitates rigorous reasoning, qualification or quantification of the benefits resulting from their deployment. This means that the real value of compliance, or the variability of compliance, to risk-control standards is not well reasoned or measurable in any scientific, unambiguous or verifiable sense. This is further complicated by the variety of control-sets being promoted by a wide variety of stakeholders, who clearly have overlapping but different perspectives.

The consequence of this gap in knowledge is that it is very difficult to argue for budgetary allocation based on quantifiable benefits. This vacuum is often filled by considering worst-case and common-case scenarios for similar organisations and businesses operating in the sector, which may lead to fear, uncertainty and doubt. This can usually persuade organisations of need for *some* action, but it leaves the requisite case for investment built solely on the potential risk of inaction, as opposed to the measurable benefit to security posture of the possible responses advocated. Unsurprisingly then, we observe an acute need to develop a model within which the relative benefits of risk controls and responses can be compared and considered. This will allow the governance functions of businesses to take better-informed decisions around security investments and budget, and furthermore, insurance products might be better tailored to take account of the risk controls being used.

In this project we explore the value of risk controls to security posture, and so the value of compliance to standards and frameworks prescribing these controls. Our approach to this study involves the design of a model with which to reason about the effectiveness of controls. Specifically, we seek to test the hypothesis that a model relating risk controls to assets and value-at-risk can be created which:

(a) considers the potential harm resulting to an organisation from successful cyber-attacks; and
(b) can also be used to assess the benefit of compliance to security standards and frameworks (in the context of reducing harm and protecting value).

Through this work, we will lay the foundations for an approach which will ultimately be able to assess the relative effectiveness of controls, thereby providing organisations with the insight they need to make better decisions regarding which controls are best to adopt.

1.2 Overview of our approach

In the context of this work we define effectiveness of a control to mean that an organisation is exposed to reduced cyber-risk as a result of deploying it, and by this we mean that less harm will be suffered should a cyber-attack be levelled at the organisation. The specific impact of the control will depend on each control specifically and its context. This effectiveness is clearly quantifiable to precisely the same extent as that harm can be quantified, but even where that is problematic it will generally be possible to judge qualitative improvements and compare the effects of different controls. We observe that the majority of current risk controls are oriented towards malicious acts, mostly by external actors, and so our analysis concentrates mainly in this area. The approach can, though, and should, be extended in the future to address other classes of classes of cyber threat-based harm or accidental insider actions (for example, where our research in other projects leads us to believe that current risk controls are often inadequate).

We consider effectiveness through a number of lenses:

- Assets and their attack-surface and the ability of risk controls to protect them: what it is that organisations typically need to protect in the face of cyber-risk, and how the risk controls under consideration might be deployed with the aim of protecting them; what the expected nature of the residual risk is, based on documented scientific experimentation and evidence. This aspect is covered in Section 2.
- Controls in the context of known vulnerabilities: given an understanding of common attack-surfaces and of the nature of the controls under consideration, how well the latter address the former, assuming they are deployed and configured to their maximum effect. This is a best-case analysis, as controls will typically not be configured for optimum effect in practice, so we also need to consider common use-cases and the likely results in terms of residual risk. This aspect is also covered in Section 2.
- The threat perspective: anecdotal or evidenced reports on the impact of current threats; what risks are being realised and what harms are being suffered despite the deployment of controls according to industry best-practice. This aspect is discussed in Section 3.
- The perspective of organisational cyber-harm: developing a wider view of the negative consequences of realising cyber-risk and considering the knock-on effects in terms of harm. This aspect is examined in Section 4.

The bulk of this report is in Section 5, and addresses our approach to modelling and analysing the effectiveness of risk controls. In it, we introduce our model which hypothesises about the relationships between risk controls on the one hand, and assets (in the broadest sense), cyber-VaR and cyber-harm on the other. The model is created to allow analysis into areas where value and harm are unaddressed by the controls in place, and it enables further understanding on topics such as control effectiveness and residual risk and data needed to evidence these factors. Finally, in Section 6, we draw our conclusions and produce an assessment of the issues driving the project, the relative effectiveness of risk controls, and the value of compliance to standards and frameworks mandating them.

2. Assets, Attack Surfaces and Controls by Design

In this section, we seek to consider assets and attack surfaces, and the controls that are commonly put in place to protect and address them. Specifically, we first concentrate on broadening our understanding of the organisational assets that require protection and the characteristics of those assets. We then introduce risk controls and reflect on how they affect the attack surface and protect organisational assets. Here we emphasise the importance of examining the effectiveness of controls and residual risk, i.e., where controls, by design, may not address the entire risk. With controls introduced, we move to examine the interdependencies present within common risk controls, and discuss the impact that this has when considering how to protect assets and the residual risk maintained across the organisation.

2.1 Broadening our understanding of the organisational assets which need protecting

To understand the effectiveness of risk controls, it is important to first consider in detail the assets that controls are attempting to protect. We adopt the definition of National Institute of Standards and Technology (NIST) and define an asset as 'anything of value to the organisation'[1]. This section aims to build on this definition and clearly define assets within organisations. Our approach to this task seeks to be complete, and thus, to outline a comprehensive list of core assets at various levels. We believe there are numerous benefits to this, particularly as it would allow modelling across a set of assets, and from various perspectives. For instance, this would be useful for situations where different sets of risk controls target different levels of assets.

To assist in the creation of this asset list, we have reflected on several sources including articles in business studies and computer science, and more practitioner-focused documents[2]. For presentation, we use a high-level categorisation of assets based on: Physical, Informational / Systems, People, Routine, and Enterprise. These categories are useful later as they provide some insight into the asset dimensions mentioned in our previous report; for instance, Devices and Buildings are Physical assets.

Below we present the main types of assets identified, and also the key sub-types that may be found in organisations.

Physical

Circuit, User device (Desktops, Laptops, Mobile phones, Telephones, Tablets, Peripheral devices), Server (aspects of hardware), Network (Local Area Networks, Wireless Networks, Telephone Networks, and Network systems such as Routers/switches, Firewalls, and Intrusion Detection Systems), Media (Hard disk Drives, Flash Drives, Disk media, Back-up tapes, Documents, Smartcards, Payment cards), Public terminal (Kiosks, Detached PIN pad or card reader), Power Supply, Interconnection point, Transmission node, Land, Buildings, Machinery, Equipment (Printer, Copier, Fax machine, Telephone equipment, Office supplies), Furniture (Chairs, Desks, Filing cabinets, Bookcases), Fixtures (Awnings, Lighting, Plumbing), Vehicle, Cash.

Informational / Systems

Data (Files, Logs, Network traffic), Information (Customer information, Sales information, Human Resources information, Company information, Financial information, Production information,

[1] NIST (2011) "Specification for asset identification" http://csrc.nist.gov/publications/nistir/ir7693/NISTIR-7693.pdf [Accessed online 24 May 2016]

[2] Legg, P.A., Moffat, N., Nurse, J.R., Happa, J., Agrafiotis, I., Goldsmith, M. and Creese, S., 2013. Towards a Conceptual Model and Reasoning Structure for Insider Threat Detection. JoWUA, 4(4), pp.20-37; Investopedia. (n.d.) Balance Sheet Components - Assets http://www.investopedia.com/exam-guide/cfa-level-1/financial-statements/balance-sheet-components-assets.asp [Accessed online 12 July 2016]; Alberts, C.J. and Dorofee, A., 2002. Managing information security risks: the OCTAVE approach. Addison-Wesley Longman Publishing Co., Inc.; VERIS Community (n.d.) http://veriscommunity.net/enums.html#section-asset [Accessed online 12 May 2016]; Spurga, R.C. (2004) Balance sheet basics: financial management for non-financial managers. Penguin Books Ltd.; WebFinance, Inc. http://www.businessdictionary.com/; Wikipedia, https://en.wikipedia.org/wiki/List_of_software_categories

Intellectual Property Information), Software (Application software, Database, System software, Programming software, Directory, Logging, Remote Access support), Advanced autonomous systems, User accounts (or Identity credentials).

People
Employee, Contractor, Temporary staff, Business partner, Customer.

Routine (mirror main typical business functions)
Procurement routine, Production routine, Sales routine, Human resources routine, Accounting routine, Payroll routine, Information Technology routine, Research and development routine, Marketing and Acquisitions routine.

Enterprise
Goodwill, Reputation, Patents, Copyright, Wisdom, Knowledge, Culture, Policy, Governance, Trust, Profitability.

With the set of assets outlined, we next move onto the dimensions of relevance to our desired analysis. Dimensions here may be viewed as characteristics or properties of the specific asset under consideration. These dimensions focus on risk factors and capture characteristics of assets that may enable attacks. From our initial work (in Phase 1 of the project), we noted that at very least, asset dimensions include: the scale from Connected to Isolated, Digital to Analogue/Non-Electronic, Human to Non-human, Intelligent to Unintelligent, Portable to Fixed, Novel to Established, Persistent to Transient, Physical to Non-Physical, System to Component and Tacit to Explicit.

Our more recent reflections have cemented our belief that these dimensions are key and that they all provide insight into understanding whether assets are more or less attackable, and their potential for being harmed. Moreover, we decided to add another dimension, namely, Common to Bespoke (discussed below). In what follows, we examine each dimension in more detail and consider how it is useful for our analysis into attacks and harm.

Connected ↔ Isolated
This dimension defines a scale on which an asset may be judged to be linked to other assets, and spans from Connected (i.e., it is fully linked) to Isolated (i.e., it is not linked and is completely detached). If an asset is *highly connected* (e.g., connected to the Internet and connected to internal network systems) it is arguably more at risk of an attack (given that it is reachable from the public space) and if breached can result in more significant harm for an organisation (given it may be used as a platform to attack other internal systems, i.e. harm propagation). Isolated or air-gapped systems are harder to access and therefore attack, and if they are compromised there may be little opportunity for onward propagation.

Digital ↔ Analogue/Non-electronic
This dimension defines a scale on which an asset may be judged to be in digital form, and ranges from Digital (i.e., fully digitised) to Analogue/Non-electronic (i.e., containing no digital components). If an asset is in digital form, generally speaking it is directly targetable in cyber-space and possibly more susceptible (although this is yet to be proven) to cyber-attacks than if it is in analogue/non-electronic form. Attacks on such assets may also be launched remotely depending on how connected it is. Some digital assets (e.g., a database) may also be easier to completely destroy in an instant; therefore resulting in a significantly higher level of harm over a shorter period of time than with typical non-electronic assets. (Hence why redundancy is considered an essential control.)

Human ↔ Non-human
This dimension defines a scale on which an asset may be judged to be in human form. It can range from Human (i.e., a 'being' composed fully of flesh and bones) to Non-Human (i.e., an object with no flesh or bones); within the spectrum there is also cyborg, which is a human embedded with mechanical/digital parts. An attack on the human body directly requires some form of physical access, and harm can range from minor pain to loss of life (a permanent state); this is important to note as humans are irreplaceable. Cyborgs would be susceptible to other forms of attack (e.g., large magnets), and depending on their connectivity may be remotely accessed. Non-humans can be attacked in numerous ways depending on their connectivity, but can usually be replaced.

Portable ↔ Fixed
This dimension defines a scale on which an asset can be moved, and ranges from Portable (i.e., can be easily moved from one place to another) to Fixed (i.e., impossible or extremely difficult to move). In terms of an attack, one observation is that the more portable an asset is, the easier for it to be stolen or lost. Fixed assets avoid these types of attacks to some extent, but would suffer in cases where there is an imminent attack on a site and their lack of portability make it difficult to move them from harm's way. From our analysis, portability has a limited influence on the degree of harm that may result from an attack.

Novel ↔ Extant
This dimension defines a scale on which an asset may be judged based on its age, and spans from Novel (i.e., newly created) to Extant (i.e., existing before). One assumption that is often made here is that newer systems are less susceptible to attacks than extant systems. This is based on the argument that extant systems have been available for some time and are vulnerable to attacks or exploits that have been discovered since their release. Moreover, if those systems have not been patched or indeed, manufacturers have ceased support, they are even more at risk. This does not necessarily mean that new systems are free from vulnerabilities, but in general, extant systems may be more open to attack. The harm to be associated with novel or extant assets is dependent on various aspects including how connected the assets are, how important it is to the organisation.

Persistent ↔ Transient
This dimension defines a scale on which an asset may be regarded to continue or be transient. It ranges from Persistent (i.e., can last or exist for a long period of time) to Transient (i.e., only lasts for a short period and is very ephemeral). The persistence of an asset influences the time period (or, window of opportunity) over which it can be attacked. This has an impact on the efforts also invested in protecting the asset on a continuous basis.

Common ↔ Bespoke
This dimension defines a scale on which an asset may be regarded as common or commonly used as opposed to one that is bespoke and more unique. It ranges from Common (i.e., is widely used and available) to Bespoke (i.e., not widely used and potentially custom-built). There are two ways to consider assets in terms of attack likelihood according to this scale. Firstly, we could assume that assets that are more used are more attractive to attackers, as there is a larger pool of potential targets. Conversely, it may be the case that following on from the first point, there are more security vendors and professionals aiming to constantly examine such asset attacks and produce patches (be they technical or otherwise). In the case of bespoke assets, while these may be less appealing to attack in general (depending on what they are, and their value, of course), their custom-built nature may mean that they are not critically constructed or reviewed – thus, they may be more prone to weaknesses.

Physical ↔ Non-Physical
This dimension defines a scale on which an asset may be judged to be physical, and ranges from Physical (i.e., is a material object) to Non-Physical (i.e., is immaterial and largely intangible). Assets which are more physical are prone to physical theft, whereas assets which are non-physical (i.e. information) can potentially be stolen with cyberattacks. When damage of assets is considered both physical and non-physical may be damaged with cyberattacks. No assumption can be made for which assets may be more attractive to attackers, and in some cases physical assets may contain with some processing non-physical.

Intelligent ↔ Unintelligent
This dimension defines a scale on which an asset may be judged as considered to be intelligent and spans from Intelligent (i.e., adaptive, ability to learn) to Unintelligent (i.e., just following rules). Intelligent systems are potentially more valuable to organisations, thus more attractive to attackers and probably more difficult to detect attacks against (since their behaviour is subject to change). Unintelligent systems are potentially an easier target since their functionality is stable and may exhibit weaknesses. Attacks against these systems should be easier to detect since there will be deviations from their routine which can be identified with simple analytic tools.

Autonomous ↔ Non-Autonomous
This dimension defines a scale on which an asset may be judged to be autonomous, and ranges from Autonomous (i.e., capable of acting independently) to Non-Autonomous. Autonomous systems are potentially more valuable to organisations, thus more attractive to attackers, and probably more difficult to detect attacks against (since their behaviour is subject to change). Non-autonomous systems are potentially a target for insider threat activity.

System of systems ↔ Component
This dimension defines a scale on which an asset may be composed of other assets. It can range from System (i.e., the asset is composed of other assets) to Component (i.e., an individual object which is used to make a system). In terms of attacks, the threat surface of a system of systems is much greater than a component and may not be a simple addition of the threat surface of each constituent component, since we need to consider how interactions between components may be subject to attacks.

Tacit ↔ Explicit
This dimension defines a scale on which an asset, particularly one like knowledge, is judged to be explicit or known. It can range from Tacit (i.e., understood without being openly expressed) to Explicit (i.e., needs to be formally defined). In terms of attacks, tacit knowledge is extremely difficult to target since it is not explicitly captured or defined, so extremely difficult to replicate, and thus identifying an attack surfaces for stealing it also difficult. Explicit knowledge is more vulnerable to attacks focusing on stealing such assets. On the other hand, tacit knowledge is probably more valuable than explicit, and attacks focused on destroying or damaging such knowledge by damaging/removing from service those who possess the knowledge are likely to be possible.

From our complete list of asset dimensions, we now have a core part of the foundation through which assets can be analysed. The next key aspect to be considered is the dimension scales and the process of rating assets along each of these dimensions. There are many ways in which this rating can be achieved, and while we leave the final decision to organisations choosing to implement our approach, we provide some guidance here on potentially appropriate levels.

The first option that may be useful is a three level rating scheme, here we use: Full, Partial or None. This especially suited to dimensions such as Connected, Digital, Human, Portable, Physical, Intelligent, and Autonomous. Therefore, a particular asset could be Fully Connected, Partially Connected or Not Connected. Similarly, it also could be Fully Portable, Partially Portable (i.e., can be moved but only though a notable amount of effort) or Fixed.

The benefit of a three-level scheme is that it is simple to use. The disadvantage is that it may not be granular enough to identify important distinctions in assets which may impact attack likelihood or impact, and this lack of granularity may have consequences in a lack of nuance in risks mitigation. For instance, if we consider the example above, there are a significant number of assets that would fall within the Partially Connected category, as connections may be mediated by time or those that connect via other devices (e.g., devices such as smart watches that pair with mobile phones). From an attack likelihood perspective, it would be ideal to better distinguish between such devices as they may be more or less likely to be attacked.

The second option is a five level rating scheme, here: Full, Almost Full, Partial, Almost None, or None. This scheme attempts to allow for a more granular rating of assets, particularly to provide better insight when attempting to relate stated asset dimensions to attack likelihood and impact assessments. There may also be instances where this scheme can be used in concert with the three-level scheme. This could be because a particular scheme relates better to the dimension under focus or simply due to organisational preference (e.g., they may only care to know that an asset is Partially Digital as this means that it could be susceptible to any digital attack).

A third option, which may be useful in very basic situations, is a binary system where we are interested in whether an asset has the potential or exhibits a characteristic, which the dimension describes. The advantage of this approach is its simplicity in reasoning about characteristics of assets, whereas the disadvantage is that it does not follow a wide spectrum and thus, lacks in granularity. We should note that not all dimensions are relevant for every asset hence the N/A value in the examples given below for those dimensions which are not applicable for the asset under consideration.

Having discussed assets, dimensions and potential rating scales, now we present examples of a few devices of the types mentioned at the start of this section, with assessments in terms of the dimensions. Notice that the dimensions apply to the assets and not to the respective controls that may be in place.

Device	**Connected**	**Digital**	**Human**	**Portable**	**Novel**	**Common**	**…**
Customer personal information	Partially Connected (it is stored on specific servers)	Fully Digital (data is kept in digital form)	None / Non-human	Fully Portable (can easily be moved, transferred or transported)	*N/A*	*N/A*	
Dell-PC-Desktop-1	Fully Connected (connected to organisational systems and networks and to the Internet)	Partially Digital (hardware and software)	None / Non-human	Partially Portable (can be moved but not built for portability or use on the go)	Almost None (desktop that is 4-5 years old)	Fully (Dell PCs is a main provider of enterprise workstations)	
Dell-PC-Desktop-2	None / Isolated (workstation	Partially Digital (hardware	None / Non-human	Partially Portable (can be moved but	Almost Full (desktop was only	Fully (Dell PCs is a main provider of	

	is isolated from other network workstations and kept behind an airgap)	and software)		not built for portability or use on the go)	purchased and installed recently, and has mostly up-to-date specs)	enterprise workstations)	
Wileyfox Swift-Mobile-phone	Fully Connected (connected to organisational systems and networks and to the Internet)	Partially Digital (hardware and software)	None / Non-human	Fully Portable (can easily be moved, transferred or transported)	Novel / Full (mobile was purchased within the last few days and has the most up-to-date specs)	Partially Common (while it has entered the market, it is not at all common or mainstream)	
Jane Goodman-HR-Employee	Fully Connected (employee is connected to and has access to several other assets)	None / Non-electronic (at least in the traditional sense)	Human	Fully Portable (can easily move from one place to another)	*N/A*	*N/A*	
Organisation-X-Reputation	*N/A*	None / Non-electronic	None / Non-human	Partially Portable (can be transferred but not built for portability and it may diminish if moved)	*N/A*	*N/A*	

Table 1: Examples of specific assets and their dimension ratings

Using the information available in Table 1, an example of the type of inference that may be made is that *Dell-PC-Desktop-1* is more at risk of being attacked than *Dell-PC-Desktop-2*. The argument here is that it is fully connected (therefore more open to attacks from a host of threats across cyberspace) and a very dated device (representing an increased likelihood that vulnerabilities exist in it) therefore is more at risk. Such inferences can inform an initial analysis and be added to the many other factors that can affect attack likelihood.

One of these factors is captured by the relationships between assets; this is a point we consider in detail later in this document. Relationship is important because *Dell-PC-Desktop-2* may, in fact, be more likely to be attacked given that it contains data of high interest to an attacker, such as customer personal information (e.g., credit cards, addresses) or Intellectual property (e.g., sensitive patents in progress). This is also relevant to our analyses on Cyber-Harm, as only through knowing how assets are connected or related can one understand the full impact on the organisation's set of assets (both direct and indirect impacts) of an attack. Practically speaking, it may be possible to create an approach, based on this asset taxonomy and dimensionality, to help an organisation better determine its assets, and then the subset of assets most exposed to cyber-attacks, and most likely to become onward attack platforms if compromised.

With the outline of assets now complete, we move on to consider how controls act to protect assets.

2.2 Reflecting on how controls affect the attack surface and protect assets

A control is a security mechanism put in place to reduce the attack-surface and protect an asset. Controls may exhibit one or both of two main classes of aim. The first aim is to reduce the likelihood of a successful attack by completely avoiding a risk and removing the attack surface in consideration; or diminishing the risk by reducing the relevant attack surface; or increasing the work-factor or effort involved in conducting an attack thereby making it much less likely a motivated threat will attempt to conduct it. The second aim, without excluding the first, is to reduce the impact of a successful attack, thus restricting the loss which may occur. In all cases this might be seen through the lens of protecting the value-at-risk by reducing the associated risk from cyber-attacks, or preventing harm to the organisation by limiting or removing some of the risks that might have resulted in said harm.

There are numerous security controls available today. To aid in their adoption and implementation, many of these controls have been organised into control sets and standards. Examples of well-known control sets and standards include the International Organization for Standardization (ISO) / International Electrotechnical Commission (IEC) 27000 series[3] of security standards, the Center for Internet Security's (CIS) Top 20 Critical Security Controls (CSC20)[4], and the National Institute of Standards and Technology (NIST) Cybersecurity Framework[5]. Each of these sets provides its own unique approach and guidance on how organisations may use the controls included to best protect themselves, their data, systems and processes.

To allow us to engage in a detailed analysis of how controls can reduce the attack surface and protect assets, we have chosen to concentrate on the CSC20 control set as an example. The CSCs are a set of 20 prioritised and well-vetted actions, activities and tools – which are defined in more detail as sub-controls – that have been put forward to assist organisations in improving their state of security. They have been derived from an understanding of the threat environment (including the main types and vectors of attack) and current technologies used within organisations. We present further detail and descriptions of the controls within Appendix 1.

Our approach to considering how these controls act involved listing controls and sub-controls, and then analysing each of these separately in terms of assets that they guard and how they protect them. The outcome of our analysis is presented in tabular form in the attached spreadsheet file, but next we provide a few examples of our analysis and findings. This focuses on four main controls. These are *CSC 1: Inventory of Authorized and Unauthorized Devices, CSC 8: CSC 8: Malware defences, CSC 10: Data Recovery Capability* and *CSC 17: Security Skills Assessment and Appropriate Training to Fill Gaps*.

The first control we consider is CSC 1: Inventory of Authorized and Unauthorized Devices. This control encourages organisations to "Actively manage (inventory, track, and correct) all hardware devices on the network so that only authorized devices are given access, and unauthorized and unmanaged devices are found and prevented from gaining access". From the control title, we can immediately notice its wide asset scope, that is, to protect physical User devices, Servers, Networks, and Media. This protection is targeted towards reducing the likelihood of attacks by maintaining a knowledge of the legitimate (and thus, illegitimate) devices on the corporate network. By managing these devices and networks carefully, the attack surface available to threats might be reduced.

[3] ISO/IEC. ISO/IEC 27001 - Information security management http://www.iso.org/iso/home/standards/management-standards/iso27001.htm
[4] CIS. The CIS Critical Security Controls (CIS Controls) https://www.cisecurity.org/critical-controls.cfm
[5] NIST. Cybersecurity Framework. https://www.nist.gov/cyberframework

The next control we examine is that of *CSC 8: Malware defences*. This control is intended to "Control the installation, spread, and execution of malicious code at multiple points in the enterprise, while optimising the use of automation to enable rapid updating of defence, data gathering, and corrective action". As the description states, this control is dedicated to preventing malware execution and spread in an organisation. To map this control onto the set of assets which it protects, the control could directly be mapped to: User devices, Servers, Networks, Media, Information, and Software. It is the aim of the control to protect these assets by reducing the likelihood of successful attacks that could compromise them; these attacks could result in impacts particularly relating to unauthorised access (e.g., of data) or denial-of-services (e.g., ransomware making a system unavailable). This control achieves its aims and reduces the attack surface by anti-malware software, limiting the use of untrusted devices, or automated tools that continuously scan for malware, amongst other approaches.

CSC 10: Data Recovery Capability is the control we focus on next. This control emphasises "The processes and tools used to properly back up critical information with a proven methodology for timely recovery of it". The assets that are generally protected by this control are Information and Software (particularly when Software is the developed within the organisation). The nature of this control is to protect assets by reducing the impact of a successful attack. Therefore, if an attack occurs that compromises systems or data, it ensures that the organisation has a back-up which can be quickly restored to minimise harms such as compromised data, or service disruption.

The last control we discuss here is *CSC 17: Security Skills Assessment and Appropriate Training to Fill Gaps*. This control is described as follows: "For all functional roles in the organisation (prioritising those mission---critical to the business and its security), identify the specific knowledge, skills, and abilities needed to support defence of the enterprise; develop and execute an integrated plan to assess, identify gaps, and remediate through policy, organisational planning, training, and awareness programs." This control seeks to protect and inform the 'People' asset of an organisation, and also influences the organisation's culture. This is an important goal as the better informed and trained individuals are, the better chance they stand at not becoming a victim of an attack (and, for instance, leaking data or inadvertently installing malware). Or, if they are a security professional, the better the chance of them quickly detecting an attack. Increasing these chances directly reduces the likelihood of a successful attack and minimises parts of the attack surface that would be targeted by attackers.

From this brief reflection on some of the CSC20 controls, we have witnessed how each control seeks to protect an asset or a set of assets within the asset sub-levels that were introduced earlier in this report. Furthermore, we briefly considered the aim of the control in addressing a particular risk by reducing the likelihood of a successful attack and/or the negative impact the attack would cause. An interesting point that can be seen here is that there are often multiple controls protecting single assets. This suggests the prevalence of a system of controls which are often layered and inter-dependent.

Two other observations that can be made about the range of controls reviewed is that firstly, controls are not without inherent design vulnerabilities, and secondly, there is also little real data on their effectiveness. If we consider the first point as it relates to *CSC 1: Inventory of Authorized and Unauthorized Devices* for example, virtual machines may be much harder to account for (e.g., falsified MAC addresses), while Virtual Private Network (VPN) access may hinder the accurate identification of a device. It is also critical that this control and the supporting list of whitelisted devices remain up-to-date and protected, as a compromise in either could severely impact the control. A similar situation can be seen with *CSC: 8 Malware Defences*. An inherent vulnerability of this control is its reliance on known vulnerability and attack listings – these listings are used to

identify malicious or suspicious files. Therefore, without them, the control's utility is extremely limited.

The second observation relates to having a good understanding of the effectiveness of controls. In our assessment of the range of CSC20 controls, we were only able to find a small amount of specific data about the effectiveness of controls. One example of such data is a university study on the effectiveness of the Cyber Essentials risk control scheme[6]. From their analysis, researchers found that Patch Management (which is essentially CSC 4: Continuous Vulnerability Assessment and Remediation) was the most effective at addressing a majority of attacks that the report focused on. Conversely, Anti-malware tools (similar to our *CSC: 8 Malware Defences*) performed poorly in terms of effectiveness, only mitigating a small number of the vulnerabilities it was tested against.

Whilst we uncovered other evidence about controls it was either too generic or arguably, too tool specific. For instance, in the case of the former, the Australian Signals Directorate (ASD) Top 4 controls said to mitigate at least 85% of intrusion techniques (thereby being very effective) generally include: Application Whitelisting (similar to CSC2), Patching Applications (similar to CSC4), Patching Operating Systems (similar to CSC4), and Minimising Administrative Privileges (similar to CSC5). For the latter, on certain vendor pages[7], we could find tools such as email security systems being regarded by customers (via testimonials) as being very effective at addressing the related threats.

Reflecting briefly on these findings therefore, it is clear that there is a gap in the literature and research regarding exactly what controls are most effective at protecting against attacks, and to some extent, how should the effectiveness of controls be judged. This lack of scientific experimentation and effectiveness data means that decisions regarding the implementation of controls are not being based on facts about the true performance of controls. As we will discuss next, this problem is exacerbated when considering the inter-dependencies of controls and that a lack of effectiveness of one control can impact numerous others and severely compromise an organisation.

2.3 The interdependency of controls

While an individual security control can protect assets and act to mitigate risks, in reality controls must be orchestrated together in an organisation's architecture. This ensures that controls do not conflict with each other, but instead, allow for layers of security and defence-in-depth. We have been studying this topic of security controls and their interdependencies within key standard sets (e.g., CSC20, ISO27000), and below present our findings with special focus on CSC20.

In our analysis we focused on three areas regarding the dependency of the controls. Firstly, we considered the nature of the dependencies of the sub-controls, for instance we analysed whether the sub-controls within a control are sequential. Secondly, we assessed dependencies at the control level – this could highlight fundamental controls within the set. Thirdly, we were keen on examining the controls to determine whether there were any implicit dependencies, for example, sub-controls that depended upon other security mechanisms that were not in the core sub-control list.

The first area examined the nature of dependencies of the sub-controls. Our analysis indicates that, in some cases, sub-controls provide atomic steps towards a bigger aim, which is the implementation of the control. These atomic steps are not only essential for the effectiveness of the control, but they also imply a sequence in terms of implementation. In other cases, sub-controls were independent and provided all the necessary features for a complete control. There is no sequence implied and in

[6] Cyber security controls effectiveness: a qualitative assessment of cyber essentials http://eprints.lancs.ac.uk/74598/4/SCC_2015_02_CS_Controls_Effectiveness.pdf
[7] FireEye. Email Security. https://www.fireeye.com/products/ex-email-security-products.html

some rare cases, sub-controls describe the same notion, providing alternatives to the organisation for implementation. Generally, therefore, there is little generalisation in approach that can be witnessed at the sub-control level, which is a notable finding for anyone attempting to apply or further research these controls.

The second area focused on the dependencies of the controls on other controls. We examined every sub-control and indicated whether its success or not will depend on the existence of another control. We aggregated the identified controls supporting all the sub-controls for the control into question and assumed that the successful implementation of the control depends on this aggregate set of controls.

Considering that the numbering in CSC 20 implies prioritisation or sequence in implementation, it is expected that the higher in priority the control is, the more dependencies there are on it. Figure 1 presents the number of controls supported by the control in question.

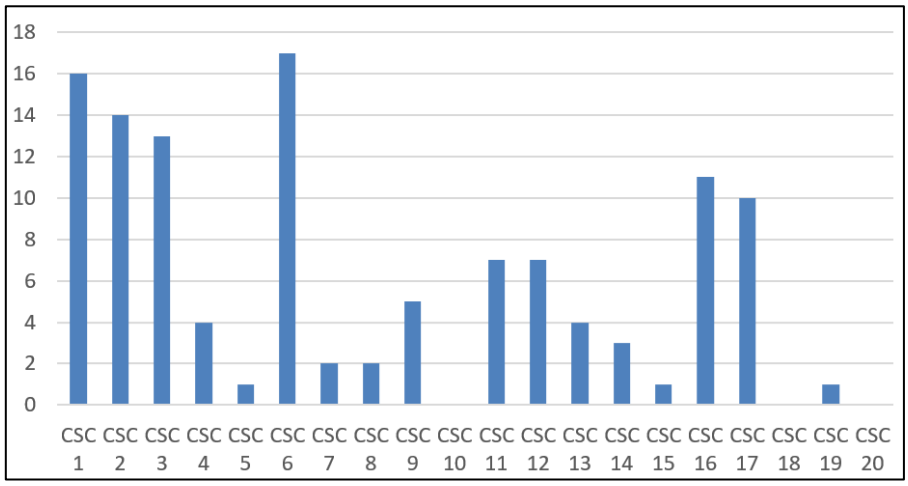


Figure 1: Number of controls dependent on the control in question

Our analysis found that the first three CSC controls, namely 'CSC 1: Inventory of Authorized and Unauthorized Devices', 'CSC 2: Inventory of Authorized and Unauthorized Software', 'CSC 3: Secure Configurations for Hardware and Software on Mobile Device Laptops, Workstations, and Servers', facilitate the implementation of more than twelve of the other controls. This therefore supported the notion that these controls were high priority and definitely to be considered as foundational. Outside of the CSC's own top 5, we found that 'CSC 6: Maintenance, Monitoring, and Analysis of Audit Logs' was a crucial control in that, it had the most dependencies on it. This is undoubtedly because it involves the collection, management, and analysis of event audit logs that could help detect, understand, or recover from attacks. The next control with over ten dependencies on it was that of 'CSC 16: Account Monitoring and Control'. The criticality of this control, albeit low in the CSC20 listing, is due to the growing necessity of stricter controls on individuals allowed to use corporate systems.

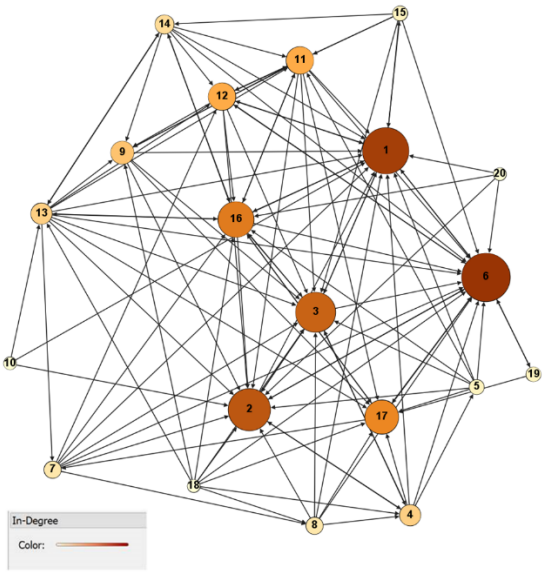


Figure 2: CSC20 control dependencies graph (the direction of the arrow highlights which control is the dependent control)

Another way in which we can consider the dependency of controls is as a connected graph, as presented in Figure 2. In this figure, each numbered node depicts the respective CSC control and the edges denote dependencies. The size and colouring of the node indicates a degree of connectivity, where larger nodes are more connected and node colouring ranges from a cream colour (less connective) to red (most connective). In this graph, we can again see the five controls mentioned above along with 'CSC 17: Security Skills Assessment and Appropriate Training to Fill Gaps' as controls that provide the most support for others.

To reflect on control dependencies more broadly, our findings in this section have highlighted the fact that the extent to which an organisation is effectively protected from risks is not only dependent upon controls separately, but also the system of controls that are implemented. Therefore, inadequacies in basic controls (e.g., maintaining an inventory of authorised devices) could impact the ability of an organisation to properly implement more complex controls (e.g., audit log and event analysis). The key point to note is that controls often will overlap, and more so, while the dependencies can benefit organisations, they may also be areas of weakness. An example of this weakness can be seen in the context of residual risk – i.e., the amount of risk remaining after a control has been implemented. As controls depend on other controls, if one of these controls is not effective – whether by design or implementation – and results in a higher residual risk, the controls that are dependent on that control also carry that risk. This propagation of risk has the potential to expand drastically as the network of controls, such as those depicted in Figure 2, is considered, and has a broader impact on the accumulated risk maintained by the organisation.

Another important point worth discussing is that of the value of a control. While it is difficult to reason about *value* abstractly, from the findings of the dependency analysis above we might conclude that the six controls identified are particularly valuable for setting the foundation for layered security. While some parts of our findings may have broader implications, for now we limit them to CSC20 given that other control sets, such as ISO 27001, may lead to other findings depending on how they outline and describe their specific controls.

The third aim of this section was to identify controls that the CSC20 implicitly depends upon. This would allow us to attain a more comprehensive understanding of the dependencies within this

control set, and is crucial as we move to consider effectiveness of systems of controls. We have documented our findings in detail in the attached spreadsheet in the CSC Controls column. In Table 2 below we present a few examples of these implicit dependencies and the respective controls.

Control	Description	Implicit dependency
CSC 4.1	Run automated vulnerability scanning tools against all systems on the network on a weekly or more frequent basis and deliver prioritized lists of the most critical vulnerabilities to each responsible system administrator along with risk scores that compare the effectiveness of system administrators and departments in reducing risk. Use a SCAP-validated vulnerability scanner that looks for both code-based vulnerabilities (such as those described by Common Vulnerabilities and Exposures entries) and configuration--based vulnerabilities (as enumerated by the Common Configuration Enumeration Project).	Automated vulnerability scanning tools that are up-to-date and reliable at detecting vulnerabilities
CSC 4.5	Deploy automated patch management tools and software update tools for operating system and software/applications on all systems for which such tools are available and safe. Patches should be applied to all systems, even systems that are properly air gapped.	Automated patch management tools and software update tools that are able to comprehensively implement and deploy patches and updates across all types of systems
CSC 7.6	The organization shall maintain and enforce network based URL filters that limit a system's ability to connect to websites not approved by the organization. The organization shall subscribe to URL categorization services to ensure that they are up-to-date with the most recent website category definitions available. Uncategorized sites shall be blocked by default. This filtering shall be enforced for each of the organization's systems, whether they are physically at an organization's facilities or not.	URL categorization services capable of up-to-date and accurately classified URLs
CSC 8.1	Employ automated tools to continuously monitor workstations, servers, and mobile devices with anti-virus, anti-spyware, personal firewalls, and host-based IPS functionality. All malware detection events should be sent to enterprise anti-malware administration tools and event log servers.	Full sets of tools with anti-virus, anti-spyware, personal firewalls, and host-based IPS functionality. These will need to be kept updated and ideally, are high quality security systems
CSC 11.1	Compare firewall, router, and switch configuration against standard secure configurations defined for each type of network device in use in the organization. The security configuration of such devices should be documented, reviewed, and approved by an organization change control board. Any deviations from the standard configuration or updates to the standard configuration should be documented and approved in a change control system.	Standard secure configurations for each type of network device are required. This may not be an issue for an established company but for a new organisation or an individual in a new role, the best configuration may not be known.
CSC 12.1	Deny communications with (or limit data flow to) known malicious IP addresses (black lists), or limit access only to trusted sites (whitelists). Tests can be periodically carried out by sending packets from bogon source IP addresses (non-routable or otherwise unused IP addresses) into the network to verify that they are not transmitted through network perimeters. Lists of bogon addresses are publicly available on the Internet from various sources, and indicate a series of IP addresses that should not be used for legitimate traffic traversing the Internet.	Known malicious IP addresses (black lists) that are up-to-date and complete
CSC 15.2	Configure network vulnerability scanning tools to detect wireless access points connected to the wired network. Identified devices should be reconciled against a list of authorized wireless access points. Unauthorized (i.e., rogue) access points should be deactivated.	Access to reliable network vulnerability scanning tools

CSC 18.4	Test in-house-developed and third-party-procured web applications for common security weaknesses using automated remote web application scanners prior to deployment, whenever updates are made to the application, and on a regular recurring basis. In particular, input validation and output encoding routines of application software should be reviewed and tested.	Automated remote web application scanners that are up-to-date and reliable at detecting vulnerabilities

Table 2: Examples of implicit control dependencies

From Table 2, we can appreciate the reliance on several other controls as a part of the main CSCs. These include automated vulnerability scanning tools, known standard secure configurations for devices and blacklists, and automated remote web application scanners. A key point to note here which relates to control effectiveness is how crucial it is that these controls are of a high-quality and are up-to-date – this, too, is a difficult task in itself.

For instance, CSC-4.1 depends on automated vulnerability scanning tools to detect vulnerabilities in the organisation's computer network. However, firstly, there is no guarantee that all vulnerabilities will be discovered as some are 0-days and are completely new. Secondly, depending on the quality of the tools and completeness of their vulnerability catalogues, they may miss vulnerabilities. Finally, if these scanners are not configured or implemented correctly and if regularly updates are not made by organisational personnel, then again, these controls could fail. Some of these factors that impact control effectiveness are caused by the external party that provides the control, but others can be caused by the organisation itself. These are points to be considered as we move to further analyse control effectiveness and residual risk.

3. Threat Orientation

Attack surface and organisation vulnerability is not the only lens through which we should consider the effectiveness of cyber-risk controls in protecting against malign attack. Clearly there is also an attacker in the equation which must provide a context to the operation of the risk-control. The threat actor, or attacker, is selecting and controlling the attack vector, the tools to be used, and the various steps to be taken. They are acting with intent and are likely have a target effect in mind (whether to steal, sabotage, or simply gain access and persist etc.) The question we consider here is whether the performance or effectiveness of a risk control is variable in relation to the threat faced.

3.1 Relationships between threat and control effectiveness

Risk controls and how they perform in relation to particular threats can be characterised in the following way:

- **Prevent:** the risk control removes the attack surface that the threat is targeting entirely, rendering the threat unable to successfully conduct an attack. This is ultimately the objective of any threat vulnerability management programme, however, whether it is possible to know that all attack surface has been removed it an interesting question. Clearly, this will be easier to achieve for some surfaces that others; it is theoretically possible to design and implement software which is free from the kinds of software behaviours that could allow a threat to invoke a run-time error and take control of the machine hosting the software. But in reality the software stacks present on a particular machine can be so varied that in practice this might be difficult to achieve, although the trusted computing platform seeks to provide aspects which can be relied upon to this degree. Any analytical approach to taking account of such measures must somehow resolve the question of the likelihood that prevention has been achieved completely. Many organisations rely on testing to provide

some kind of confidence that attack surface has been removed, but the question of knowing whether one has tested enough remains unsolved in general.

Alternatively, a risk control which seeks to remove the likelihood that a threat can reach an exploitable attack surface is also an option. This is what a firewall technology seeks to do. Again, one would need to be confident that the firewall is able to catch all possible attacks that are aimed at an exploitable attack surface. This could be abstracted to all possible attacks (since if all attacks are prevented the question of whether they were targeting an exploiting vulnerability becomes moot). As above, the question is whether we can be confident that we can anticipate all types of attack, otherwise how would we know that we had been successful?

A practical route forward might be to develop a measure of likelihood for each of the above cases. Whereby the probability assigned represents the likelihood of success. At present no data exists which could underpin the assignment of probability – so at best we might explore the sensitivity of the system to a particular estimate of probability, so providing a range of possible outcomes based on a scenario. Interestingly, it is possible to make progress in this line of reasoning without concerning oneself with the powers or capability of a particular attacker; one can just assume that the threat faced has the necessary capability and in so doing would be estimating a possible worst case scenario namely that the prevention has not worked and therefore attack surface is present and exploitable. Current risk assessment techniques will typically attempt to categorise the likelihood of a threat successfully exploiting a vulnerability by considering a range of threat actor types. The likelihood of this threat will depend not only on the ability of a threat actor to execute an attack but also on their motivation to use such resources. However, this may be an unnecessary activity given that in truth the ecosystem of suppliers offering capabilities to orchestrate attacks has become so developed that ultimately the question may simplify to whether or not a threat has the resource to purchase the necessary capability. This then becomes a binary question – either they do or they do not, and either the prevention is working in the face of the attack vector or it is not..

We must surely always need to consider the possibility that the prevention has failed if we believe we face a resourced threat. Therefore, in conclusion, all prevent controls may vary in this binary sense in the face of the threat – since the threat may be resourced enough to exploit failures in the prevention. This scenario could only be ignored if there were very compelling reason to assume that the prevention type of control is completely effective.

- **Detect and limit:** risk controls which assume that some aspect of an attack is successful, but that seek to detect the presence of the threat within the system and respond in a manner which seeks to mitigate the risk and limit the harm faced. Considered entirely necessary for organisations that face a large amount of threat, on a frequent basis, and for which the risk is considered substantial enough that simply allowing the threat to continue unaddressed would be unacceptable. Here the effectiveness of the risk control can be considered by its ability to:
    - Detect threat quickly enough to allow time for a response which can limit the harm, and,
    - Respond quickly enough to limit the harm posed by the threat.

  The intuition is, of course, that the earlier the detection takes place then the less risk exposure, or realisation of harm. However, that very much depends upon the nature of the risk. It is entirely possible that the presence of a threat on a system does not actually result

in any harm at all. A good example of this might be the accessing of a system, but no other action being taken. For some organisations the very presence of an unauthorised party on a system is risk enough – but in reality no harm may be realised. This is akin to trespass, whereby the rights of the owner have been violated in law. However, there are many examples of attacks where the threat has persisted on the system for quite some time, and it has transpired that this is very likely to facilitate reconnaissance which allows for further attack steps to be crafted and deployed, eventually resulting in harm[8]. In which case, the effectiveness of a risk control in deploying a detect and limit capability for a threat is a complicated quantity to define. Arguably, in this situation, we might need to account for the possible harms that might have been realised had the detection not taken place. Here we might very much need to reason about the likely capability and intent of a threat in order for this account to be realistic. We might then attenuate this estimate given the rapidity and nature of the response. If the attacker has been removed, then we might be able to estimate the possible harms that would have been achievable in the time period that they persisted. If the attacker has been quarantined the same is true, less the harms that the attacker can still achieve within the quarantine zone. So it is clear that detect and limit controls must vary in effectiveness in respects to the nature of the threat faced, and it may be possible to more accurately measure and then predict the degree of harm exposed to for a given threat capability (independent of the intent). It also may be possible to develop metrics for establishing thresholds on the necessary performance characteristics and effects of detect and limit controls in order to limit potential harms for given threat capabilities. This would then allow the consideration of capability requirements given assumptions on threat faced. This is really a topic for further research.

3.2 Value of predictive threat analytics in control orchestration

Most large organisations will stress the importance of a threat intelligence approach to driving operations, although it should be noted that no scientific data has been collected that would underpin this position, at this time we rely on anecdotal evidence. It feels intuitively correct - know thy enemy helps you predict their moves and so focus your resources towards this. Of course, it is also the case that security resources are bounded, and therefore a reasonable response might be to try and direct such limited budget towards preventing the most harmful attacks. For many, this means trying to direct defence towards the threat that this likely to result in the most harm. So the pertinent question in relation to a consideration of risk control effectiveness is whether or not we can determine such a threat directed programme as being more of less effective at reducing harm from cyber-attacks.

A simplistic consideration might conclude that any attempt to predict threat and therefore configure risk controls to match it must result in less exposure to harm, and therefore be satisfied with the approach outlined above in 3.1. However, the situation is not as simple. In reality we face a creative threat, a threat that is capable of reason and innovation, and which can seek to adapt to take account of risk controls being deployed. The basic example of this being a threat which seeks to determine the nature of the intrusion detection deployed and to either switch off or evade the sensors that would alert the presence of the threat.

However, it is also true that thy enemy can trick you. They might conduct multiple attacks on multiple and varied vectors in order to create some *noise*, make the security operations team busy, and attempt to hide the *real* attack from view. Several examples of this have been seen in practice,

[8] Symantec. "W32.Stuxnet Dossier". 2010. http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/w32_stuxnet_dossier.pdf

and particularly using denial-of-service attacks as the noise or 'smoke screen'[9]. This might result in a large amount of security resource being put into predicting the intent and capability of a threat which in fact is not actually going to do more harm that directing attention away from a different attack being leverage in the same timeframe. A more sophisticated attack might seek to exploit the fact that predictive threat analytics are in use, in order to determine the defence strategy and playbook by stimulating the defence response and over time mapping out the likely behaviours. This could then provide an attacker with insight with which to develop attack methods that are specifically designed to invoke a defence response by creating a pivot towards a particular attacker (a pivot that can be predicted), and in so doing create the *noise* which is specifically designed to draw attention in a particular direction and create the opportunity for an attack emanating from elsewhere. Both of these examples working on hiding in amongst other attacks, the former relying simply on the volume of attacks making the chance of the actual attack being addressed less likely, the latter relying on developing a model of the defence strategy and so being able to create one or more decoy attacks that are certain to occupy the defence team. So arguably, being threat driven could actually introduce weakness to your security posture as you predictably pivot to face a fake enemy.

We might conclude from this that predictability in detect and limit, or defensive risk controls orchestration in general, is a bad thing as it could render some controls much less effective in operation. Research is required into this question since it could stipulate additional requirements of a control system and how it is managed, in order to mitigate or avoid cyber-harm. We do not address this issue further in our analysis here since it really is an open research question and further detailed consideration is required (beyond the scope of this paper). Any analysis would have to look at the level of sophistication needed to exploit such an attack model and thus the likely targets to such sophistication and the further mitigation such targets would have in place.

4. Controls and the Manifestation and Propagation of Cyber-Harm

4.1 What cyber-harm is and why it matters

Cyber-harm is considered to be the set of detrimental impacts resulting from cyber-attacks and related incidents in an organisation. This also includes those emanating from an organisation as well; this is important if internal assets are compromised and then used as platforms for attacking other third-party organisations (e.g., as members within a botnet, for example). Harm may be localised in business assets at all levels of the organisation from computer hardware and applications, through to people, services, business units and corporate reputation. In this report, we scope our work to malign attacks and the results of coercion and social engineering of employees; at this stage therefore, we do not consider broader unintentional attacks or those linked to natural disasters.

Cyber-harm has become an increasingly important consideration for organisations because of the wide range of impacts that can now result from attacks. One of the main reasons for this is that technology is a core part of today's society and is set to become even more central as society moves towards being more connected (e.g., smart cities). In our previous research work[10], we investigated the various types of cyber-harm that might result from cyber-attacks at present and in the future. The outcome of that work has been the development of a detailed taxonomy of cyber-harms. This

[9] Incisive Media (V3). "Hackers using DDoS attacks as a 'cyber smoke screen' to mask wider threats". 2015. http://www.v3.co.uk/v3-uk/feature/2428971/hackers-using-ddos-attacks-as-a-cyber-smoke-screen-to-mask-wider-threats

[10] Agrafiotis, I., Nurse, J.R.C., Goldsmith, M., Creese, S. and Upton, D. (2016) "Understanding cyber-harm for organisations", Paper in preparation – contact point: Louise Williams (University of Oxford).

taxonomy is based on a comprehensive survey of known cyber-incidents (from typical phishing attacks to attacks targeting critical national infrastructure) in combination with a critical review of academic articles, news reports, and categorisations of attack impact and loss; as an example, we outline a selection of the cases analysed in Appendix 2. In Figure 4, we present our taxonomy, and this is followed by a discussion of its main components.

Figure 4: A taxonomy of organisational cyber-harms

Our taxonomy was structured to emphasise the notion of harm, and therefore, used well-known harm tenets as the main harm types. These included Physical / Digital Harm (i.e., harm describing a physical or digital negative effect on someone or something), Economic Harm (i.e., harm that relates to negative financial or economic consequences), Psychological Harm (i.e., harm which focuses on an individual and their mental well-being and psyche), Reputational Harm (i.e., harm pertaining to the

general opinion held about an entity), and Social / Societal Harm (i.e., a capture of harms that may result in a social context or society more broadly).

For each of these types, we have also identified several sub-types that characterise that harm in further detail. Therefore, examples of Physical / Digital Harm from a cyber-attack that can affect the organisation include compromised or exposed customer records, unavailable web services, or bodily injury of employees or customers. The types of Reputational Harm that an organisation may suffer span from a damaged public perception (e.g., they may be regarded as insecure or incapable of protecting customer data) to reduced corporate goodwill (i.e., the business becomes one that others are not keen on interacting or trading with). Harms in the Social / Societal space include negative changes in corporate culture (e.g., after a cyber-attack, the staff may view a certain type of technology as unreliable or insecure and refrain from using it), and the disruptions in the daily lives of the public or the employees of the company.

The benefit of this taxonomy therefore is its ability to outline cyber-harms, and characterise the main types and sub-types of harm which organisations need to consider as potential outcomes of cyber-attacks.

4.2 Propagation of cyber-harm, and consequences for risk and controls

While many direct cyber-harms can occur as a result of a cyber-attack, as highlighted in Appendix 2 there are also several subsequent (or, indirect) harms that may result. In what follows, we examine how harm propagates after a cyber-attack and the consequences for increasing risk. To facilitate our discussion, we reflect on a set of case studies including the attacks on Sony Pictures, JP Morgan, and Ashley Madison (more detail can be found in Appendix 2), and use these to identify examples of how different types of cyber-harm emerge and cascade. Specifically, we draw on the harms in our taxonomy in Figure 4, we identify the assets which were targeted in the case studies, which types of harm occurred first and how these types triggered different types of harms.

We start with one of the most common types of harms today, i.e., data breaches and the exposure of customer or personal data. In the case where personal data is leaked (the Sony Pictures case is a good example of this), the direct harm according to our taxonomy is *data breach or leakage of information*. In all of the case studies examined, various harms occurred that affected different entities (e.g., organisation under attack, its employees, customers and suppliers). Adopting an organisation's perspective, which is the first entity to witness the types of harm, one of the first and most prominent harms is that of *reputational damage,* which can further result in *damaged relationships with employees and customers* (e.g., in the case of Ashley Madison losing clients). At the same time, *economic harms* occurred because once the cyber-attack was announced publicly, for some of the businesses, there were *falls in stock prices, downgrading of debt, reduced numbers of customers* and *reduced growth.*

Changing the perspective and focusing on employees and customers, *psychological* harm is the most common type of harm following *leakage of digital information.* In the cases we examined, people felt *discomfort, frustration* and *worry.* Where there were instances of blackmail, resorting to *extortion payments* has also been an option. In some cases, people felt so *shameful* and *embarrassed* that some news reports claimed that it resulted in *loss of life* (this was in the Ashley Madison case).

Furthermore, where *identity theft* was evident, compensation payments are often required. Finally, if the situation is not resolved, *social harm* may occur, as it happened with the Sony case, where there was *disruption of daily lives* and a *drop in internal organisation morale.* A similar sequence of harms is repeated in all case studies where information was leaked. The impact and the length of the

propagation chain, however, depended on how well and timely stakeholders who were responsible for addressing harmful situations responded. Thus, there is a temporal element which is critical to the propagation of harm that is related to the quality of controls which organisations have in place to mitigate harms.

Similar observations occur when the assets under attack are *destroyed.* From an organisation's perspective, emerging harms were *disrupted operations, deteriorating sales* and *loss of key staff* (in cases where they are forced to resign). It is evident that the types of harm occurring depend on the assets exploited by the attacks and the remediation measures which organisations have in place. As a pattern, *physical* harms lead to *economic* harms, which if not resolved may lead to *reputational* harms for organisations. When *psychological* harms for employees occur after *physical harms,* then *economic* and *physical* harms may follow for employees and customers, whereas *economic, reputational* and more scarcely *social* harms may result for organisations.

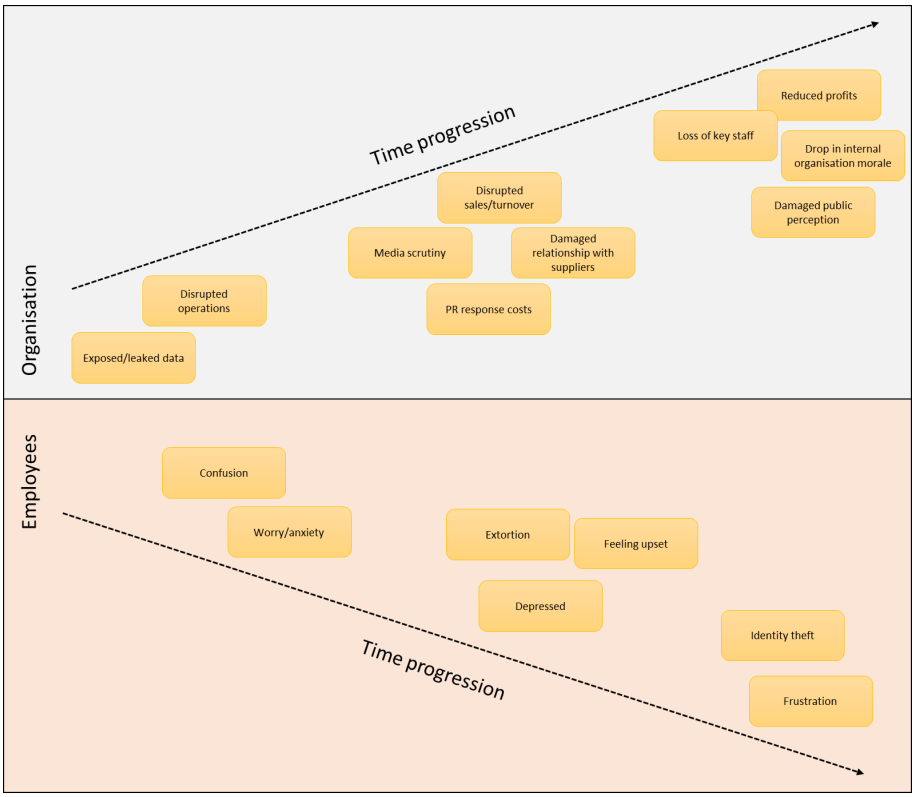


Figure 5: Propagation of harm after the Sony Pictures cyber-attack in 2014

As an example to consolidate our discussions above, in Figure 5 we illustrate how we might visualise the various cyber-harms that resulted from the Sony Pictures attack. Here we focus on two domains, the organisation and its employees, however the attack also impacted external parties such as the public (via their perception of Sony Pictures) and the government (via the impact that hackers, potentially linked to nation states, can have on large organisations within their country). To focus on the organisation, we can see that the first harm was the leak of information and as time progressed there were additional harms regarding costs, damaged relationships and reduced profits. At the employee level, there was first confusion about what was happening (e.g., why they were unable to use their systems), then as their personal data was leaked, attempts to extort them and steal their identities. This highlights the breadth of harms possible and how they progress over time.

One of the key findings from our research as it pertains to harms was that current risk assessment and control approaches focus almost solely on the direct impact of an attack and not on the range of

subsequent consequences. For instance, a traditional risk analysis within Sony Pictures prior to the breach is likely to have predominately focused on the exposure of data and subsequent corporate harms, such as disrupted operations, response costs and profit impacts. Harms unlikely to be considered or modelled by the organisation include ripple effects of attacks such as the company's employees and their psychological states (e.g., depressed employees) and issues such as theft of employee identities (and the resulting frustration felt by the individual). Such subsequent harms are increasingly important factors as they can be very costly and may not be planned for, therefore having other impacts on the operations of the company. A class action lawsuit filed by employees (present and past) against a breached organisation is one example that could result in significant costs and damage the organisation's reputation – this was one of the issues faced by Sony Pictures.

Another disadvantage of not considering the full extent of harms that could result from cyber-attacks is that controls may not have been put in place to mitigate and reduce these harms. As mentioned earlier in this report, one of the main aims of controls is to reduce the impact of a successful attack, thus restricting the loss to the organisation which may occur. If, however, the organisation does not have controls in place to reduce the range of harms that can occur, or, that those controls are not effective at their task, the harms will occur and could propagate.

A good example of these issues and also the interdependencies of controls can be witnessed in the Target breach in 2013 (see Appendix 2). In this case, Target had an incident response control (i.e., security operations centre and incident detection tools) set up but it was not effective at adequately escalating the risk across security teams nor, responding to that risk (i.e., the infection of systems and exfiltration of data)[11]. This failure to address this immediate harm of system infection, led to numerous subsequent harms including loss of customer data, angry customers, a ~50% drop in profits compared to the previous year, loss of key employees especially the CEO and CIO, and class-action lawsuits against the company. This case perfectly exemplifies how important are effective security controls at reducing both harm and likelihood of successful attacks.

Building on the knowledge regarding assets, harms and controls therefore, we can begin to better appreciate how cyber-risk may be understood within organisations. If such impacts of an attack, especially in the early stages, can be managed effectively and in a timely manner, organisations would have a chance at protecting against the various types of subsequent harm that have the potential to aggravate the overall harm significantly.

5. Modelling Control Effectiveness

5.1 Aim and method

The aim of this report is to examine the relative effectiveness of risk controls and the real value of compliance. Thus far we have laid the foundation for this analysis and explained the different lenses through which we might consider effectiveness, i.e., attack surface and control designs, threats and cyber-harm. In this section, we introduce our model which hypothesises about the relationships between risk controls on the one hand, and assets (in the broadest sense), cyber-VaR and cyber-harm on the other. Our model is created to allow analysis into areas where value and harm are unaddressed by the controls in place, and it enables further understanding on topics such as control effectiveness and residual risk and data needed to evidence these factors. The approach we have adopted to creating our model is composed of two key stages.

[11] Bloomberg. "Missed Alarms and 40 Million Stolen Credit Card Numbers: How Target Blew It". 2014. http://www.bloomberg.com/news/articles/2014-03-13/target-missed-warnings-in-epic-hack-of-credit-card-data

The first stage frames the model, and the requirements for scope, structure, and analytics that it must support. This in turn requires us to elucidate the core model components, and the characteristics that they might possess. It also requires us to reflect again on the concepts of assets, cyber-VaR and cyber-harm, and the possible relationships that these may have with each other. Our approach to this task broadly involved a critical reflection on literature covering assets, and cyber-harm[10,12,13,14] and cyber-VaR[12,14] both in the academic and wider space. From our reflection, we then sought to distil this knowledge and define and detail these concepts for the context of this report.

The second stage is to understand the ways in which risk controls and responses may have an impact on the model. For instance, we contemplate how we might map controls to the assets that they protect (directly and indirectly), the nature of residual risk, as well as the effectiveness of the controls and the inherent vulnerabilities of their implementation.

5.2 Analytical requirements of the model

The analytical requirements frame the questions that we want the model to help us answer, capturing the types of analyses that we want to be able to conduct using the model. These are very close to the overall objectives of the research, although recognising that the model may not be our only method within the research for realising our aims. The analytical requirements are as follows:

1. **Identify and predict where value and harm are unaddressed by the controls and responses** – controls are widely implemented in the expectation that they protect assets and the organisation against risks. A key aim of the model is to be able to clearly identify where there may be a misalignment in controls, in that controls do not actually mitigate the harm and protect VaR as expected.
2. **Elucidate and refine our understanding of residual risk within our systems after deployment of controls** – the topic of residual risk is widely discussed, but there is little formal understanding of it and of the true residual risk that exists in organisational systems after controls have been implemented. We seek to use the model to explore this concept in detail and to enable enhanced reasoning about the actual residual risk that may exist.
3. **Identify where we urgently need to collect data in order to quantify and refine our understanding of the real risk from cyber-attacks, and the impact of adopting certain risk controls or responses** – a core challenge in reasoning about risk from cyber-attacks and the impact of deployed controls is a lack of data. By this, we mean data on all of the specific asset types that are impacted by a cyber-attack (and how they relate to each other), data on the effectiveness of controls, and also how controls impact each other. The goal for the model, in this regard, is to elucidate the areas where we, as a community, need to collect more data to be able to better understand the range of impacts.

These analytical requirements provide the foundation for the initial model for reasoning about cyber-risk which incorporates technology, business and security processes, people and information. Addressing these requirements will allow us to describe how key assets of organisations interact with cyberspace, and how risk might propagate across them impacting ham and cyber-VaR; thus, allowing us to reason about the effectiveness of controls, being the core aim of this project.

[12] World Economic Forum (2015) "Partnering for Cyber Resilience: Towards the Quantification of Cyber Threats" http://www3.weforum.org/docs/WEFUSA_QuantificationofCyberThreats_Report2015.pdf [Accessed online 15 August 2016]
[13] NIST. "Specification for asset identification". 2011. http://csrc.nist.gov/publications/nistir/ir7693/NISTIR-7693.pdf [Accessed online 24 September 2016]
[14] Rod Beckstrom Group (RBG) (2015) "CyberVaR: Quantifying the risk of loss from cyber-attacks" http://www.beckstrom.com/uncategorized/cybervar-quantifying-risk-loss-cyber-attacks/ [Accessed online 15 August 2016]

5.3 Model overview

To define our model, we analysed the three main concepts of Assets, Cyber-Harm, Cyber-VaR and how they were related, and also how controls could be applied across these levels. This allowed us to more clearly understand a control's effectiveness in the context of how much it protected assets, reduced harm and the value-at-risk. In what follows, we present the outcome of that analysis and the final version of the model. To begin, we present the model visually in Figure 6.

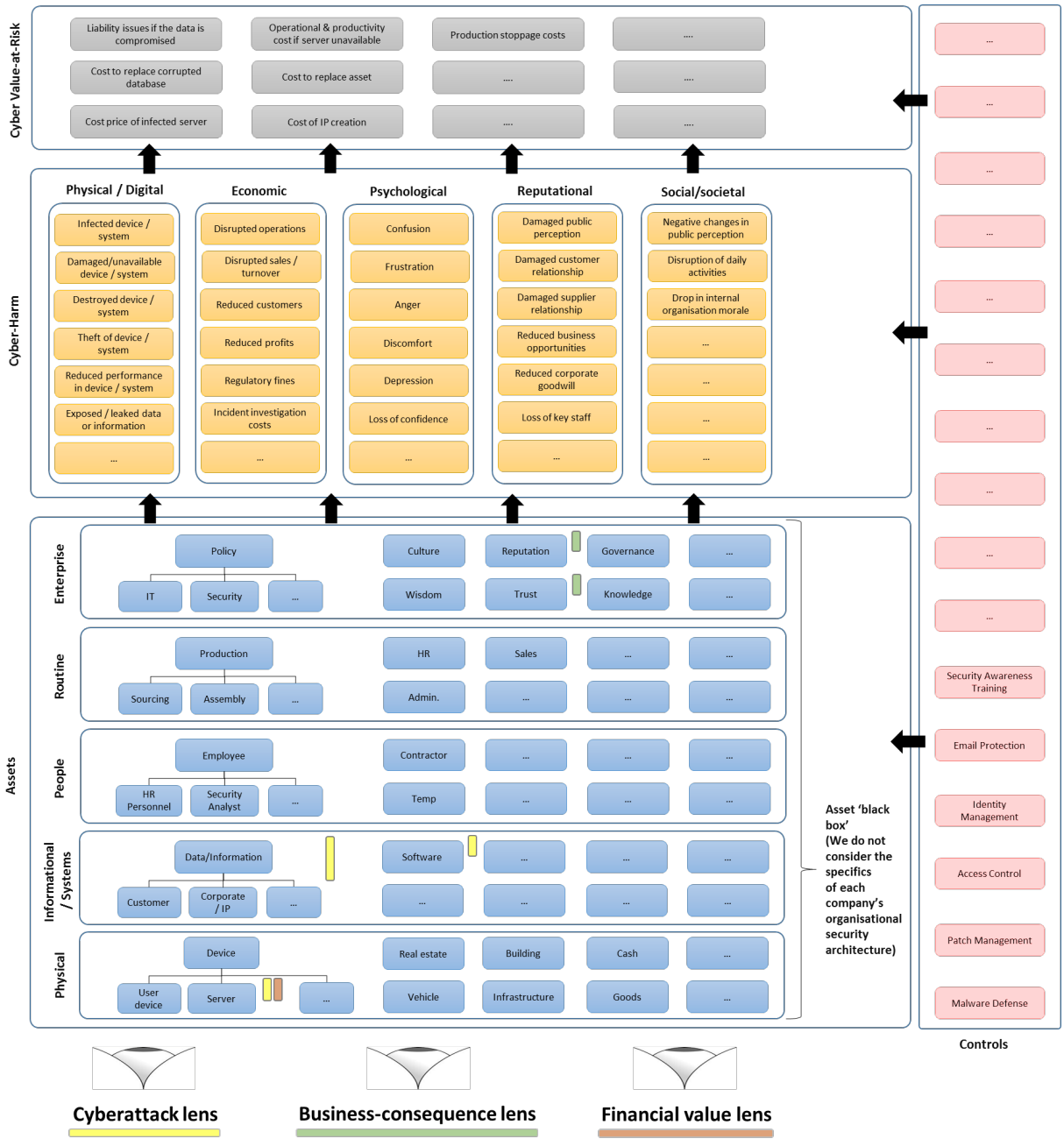


Figure 6: Initial model in detail

As can be seen in Figure 6, the model is split into three levels. The first level is the Asset level, and this covers all of the assets which an organisation may have (this utilises the work from Section 2). We have further sub-divided this level into the traditional categories of Physical, Informational / Systems, People, and Enterprise[2]; Routines have been added to capture the key procedural activities conducted within businesses. These categories are useful as they provide some insight into the asset

dimensions mentioned earlier; for instance, Devices and Buildings are Physical assets (both in the visual below and the dimensions in the section above).

Next, the Cyber-Harm level is as presented in Section 4 with the main categories of Physical / Digital, Economic, Psychological, Reputational and Social/Societal. These map closely with assets, such that if an attack occurs on an asset, it has respective harms. For instance, an attack on a file server (asset) may result in exposed or leaked information and disrupted operations. This harm may further propagate and lead to damaged employee relations or loss in key staff, as with the Sony Pictures case.

Cyber Value-at-Risk (cyber-VaR) is the topmost level and draws on both of the levels below it. Cyber-VaR is regarded as an estimation of the likely loss from cyber-attacks over a given period of time, and the overall goal with this measure is to standardise and unify various pertinent factors into a single distribution[15] that can quantify the value-at-risk in case of a cyber-attack[12]. As VaR is regarded as a monetary amount, Figure 6 simply presents examples of some of the main losses that would feature in calculating the VaR – this is intended to complement the harm level that is more textual.

In addition to the three main levels, to the right of Figure 6, we present the range of controls that may exist. In our model, these controls can be projected across the various levels with the general goal of mitigating risks. Most notably, as controls target the Asset level (e.g., Patch Management), we can model consequential impacts on higher levels including Cyber-Harm and Cyber-VaR. It is through these associations that we later attempt to reason about how controls are mapped to assets, and their effectiveness at protecting these assets by examining if there are mitigating harm and reducing cyber-VaR.

Note that there are three lenses in the initial model: the Cyber-attack lens, the Business-consequence lens and the Financial-value lens. These are described as follows:

- **Cyber-attack lens**: Focuses on physical asset entities that can be attacked in or via their interaction with cyberspace (e.g., devices, networks, infrastructure and people);
- **Business-consequence lens**: Emphasises asset items that would have a business consequence for an organisation if damaged (e.g., reputation, organisational culture, innovation and competitive advantage); and
- **Financial value lens**: Asset items that have a financial value on an organisation's balance sheet (e.g., inventory, goods, good will, infrastructure and vehicles).

These allow different types of stakeholders to understand the model and identify the most relevant components based on their focus. We have depicted these lenses at the bottom of the model looking across it from assets to cyber-VaR. As an example, we have used colour coding to depict assets that may relate to certain lenses more than others. For instance, a server is an asset that may be attacked in cyberspace (so maybe of interest to an IT administrator), whilst reputation is a larger business concern (and typically the focus of managers and C-suite employees).

5.4 Model Detail and Reasoning

In the section above, we presented a high-level model that outlined the association between assets, Cyber-Harm and Cyber-VaR, and controls. Our goal now is to understand these relationships in further detail and in particular, with respect to controls and their effectiveness. In this section therefore, we examine the relationships between these components, which incorporates general

[15] Not that the occurrence or effect of cyber-attacks are actually stochastically distributed.

relationships, relationships across the three main levels, inter-dependencies between levels, links to controls and control effectiveness.

5.4.1 Reasoning within Levels

Asset Level

The first level that we consider is the Assets level – shown in Figure 7. The first point of note about our model, as it pertains to assets, is that it aims to be abstract and thus to provide a foundational approach which companies can apply in their specific contexts. One of the tasks that companies will need to engage in as they use our model is making associations between assets themselves and assets and controls. Providing a method for organisations to describe such links unambiguously (e.g., Asset A relies on Asset B, or Asset C passes information to Asset D, or Control X protects Asset E) will enable them to obtain a better understanding of how secondary assets may be affected by a cyber-attack or, indeed, may be indirectly protected by a risk control.

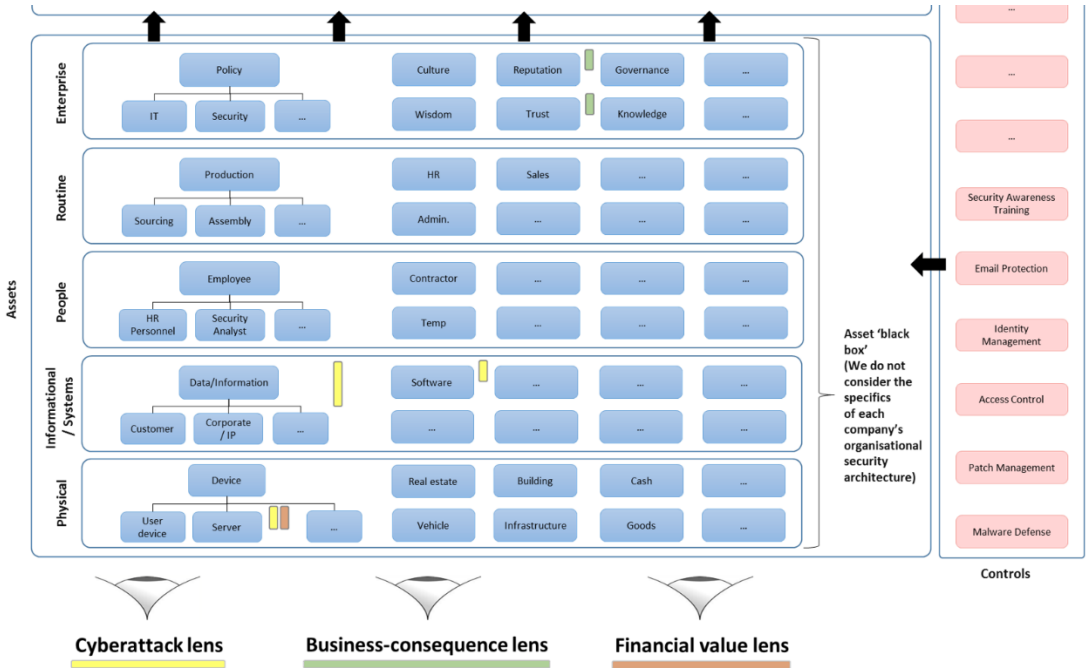


Figure 7: Asset layer and interaction with controls

We have reflected on the associations between assets and controls, and outlined a basic syntax to capture these links. Our syntax is based on the five asset sub-levels, given their similarities. Our syntax requires a subject and an object, an action and characteristics about the action. There is a set of actions which create links between the subject and the object based on their definitions. The subject and the object are assets or groups of assets. Characteristics are tailored to the actions and parameterise the type of relationship of the two assets. Examples of the syntax, outlined in terms of asset types, are presented below:

Physical ***CONTAINS*** Information / Systems
Physical ***CONTAINS*** Physical
Person ***CONTAINS*** Information
Routine ***CONTAINS*** Physical, Information / Systems, People
Person ***INTERACTS WITH*** Information
Person ***INTERACTS WITH*** Physical

Person ***INTERACTS WITH*** Person
Physical ***INTERACTS WITH*** Physical
Routine ***INTERACTS WITH*** Routine

CONTAINS – is on or is a part of, either physical or logically. For instance, Asset A (e.g., a set of customer records) is on Asset B (e.g., a computer server). For people, a person could contain or hold information related to a customer record or company file, for example. An important point about this relationship is that it is (a) directional and (b) transitive. Therefore, in the example above, it should be noted that Asset A is within (→) Asset B, as opposed to Asset B being on or a part of Asset A. Transitivity is relevant as if Asset C (e.g., data) is within Asset D (e.g., a server) and Asset D is within Asset E (e.g., a physical server room), then Asset C is also within Asset E.

INTERACTS WITH – can access, use or reference. For instance, Asset A (e.g., a sales system) can interact with Asset B (e.g., a warehouse stock system), or Asset C (e.g., an employee work station) can access Asset D (e.g., a network file server). This relationship is generally regarded as bi-directional (←→) as interaction is usually two-way. There may be instances however where interaction is only one-way (e.g., information is passed from one asset to another within requesting any information returned). In these cases, the relationship would be modelled as follows: Asset A (the 'user' or initiator) *INTERACTS WITH (→)* Asset B (the 'accessed' or 'referenced').

To give an example of these relationships within the model on a simple organisation structure – here, around an organisation's sales routine – we present the Figure 8 below.

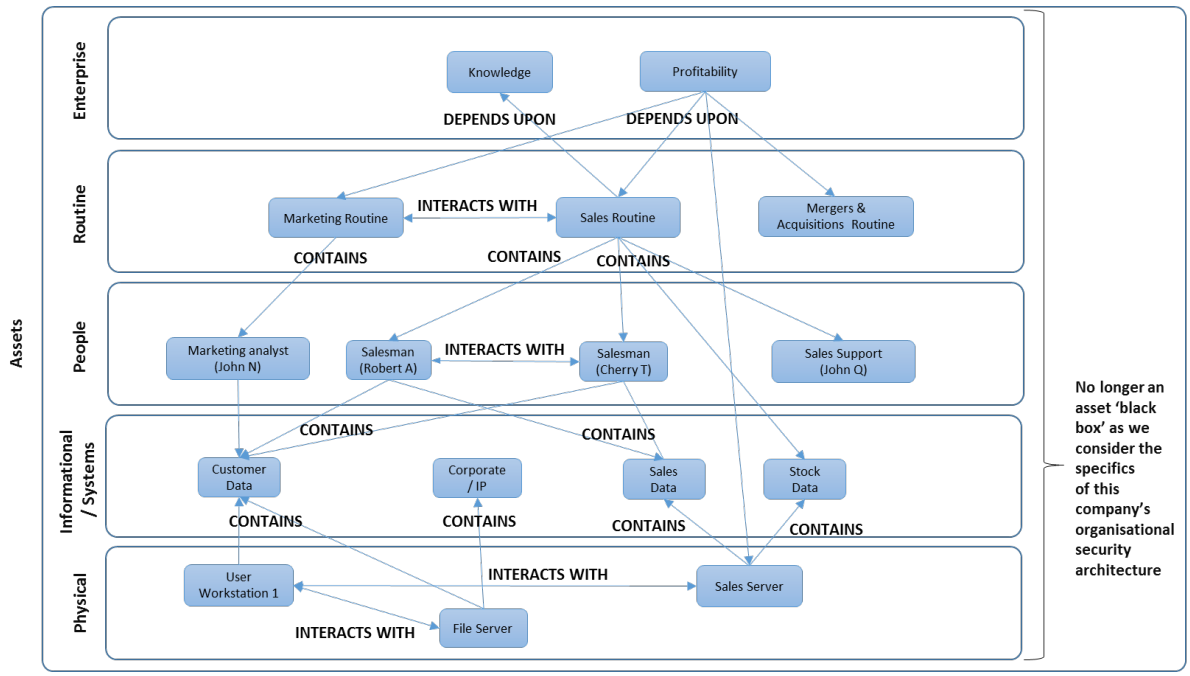


Figure 8: A set of potential relationships between assets using the defined syntax

From Figure 8 we can see the five sub-levels within the Asset level, all connected by sets of relationships. These relationships are directional and the arrows in the figure represent this. To give an example of the relationships: a User Workstation *INTERACTS WITH* a File Server, and File Server *CONTAINS* Corporate Data / Intellectual Property. Direct connections between higher and lower sub-levels are also depicted (e.g., the fact that a company's Profitability relies on its Sales Server), and

these can further be inferred via the intermediate sub-levels (e.g., Marketing Routine *CONTAINS* Customer Data).

Some relationship definitions can also be parameterised to allow more insight into the relationship. For instance, an asset may *INTERACT WITH* another asset, as a part of its core functionality or interaction may be possible but not required. We can also consider groups of assets within the asset level. Such groups may represent systems that are much more interconnected or completely dependent. In Figure 8, we present two assets within such a group; in this case, there is a File Server that contains Customer Data. To model the characteristics of groups of assets, we use a combination of the dimensions of the assets themselves (see Section 2). Here, for instance, Customer Data may have the following asset dimension values: for the Digital dimension – Fully Digital; and for the Portable dimension – Fully Portable. Similarly for the File Server, it could have dimension values as follows: for the Digital dimension – Partially Digital; and for the Portable dimension – Partially Portable. If we considered the group overall, it would have values which are a composite of both of these individual asset dimensions. We would therefore be able to analyse the group in terms of harm given it has both physical and digital, and portable and partially portable properties.

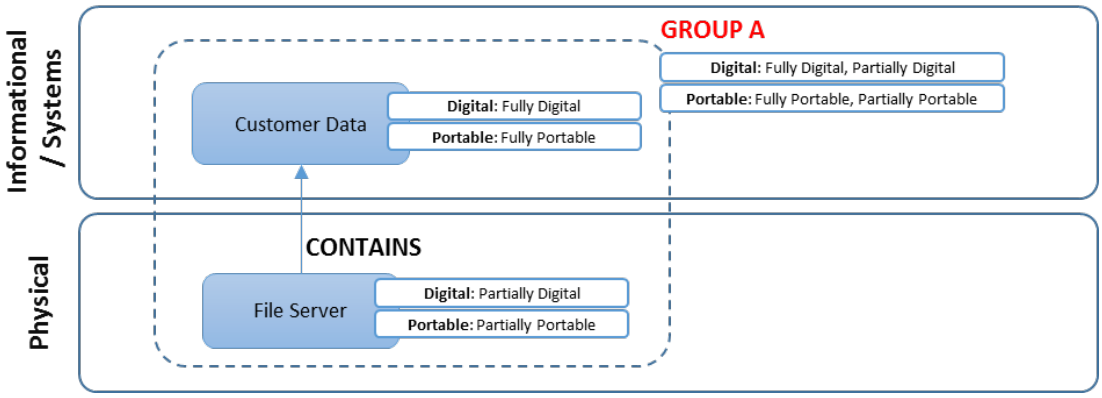


Figure 9: Modelling the properties of an asset group

Another important factor to be considered in this level is the importance of the asset for the organisation. In Phase 1 of this project we described a three-lens model, namely Financial value lens, Business consequence lens and Cyberattack lens. These lenses provide a framework for evaluating the importance of each asset. Therefore, every asset will be assigned with a value pertaining to any lens. We define a function, *EVALUATE(x)*, to assign the importance value for every lens to a particular asset *x*. When relationships are formed between assets, we can consider as parameters of the relationships the values of importance that will characterise newly formed group of assets.

Consider the example in Figure 3. We can apply the function *EVALUATE*(File Server) = £1000, None, None – this would assign a financial value of £1000 (which is the cost of buying the server) and no further value to the remaining two lenses. In a similar vein, with the action *EVALUATE*(Customer Data) = (0, Customer Data, None) we denote that the data in Figure 3 pertain to customer data, which is a value assigned when the business-consequence lens is considered.

Harm level

The next level we need to consider is the Cyber-Harm level. In Section 4 above, we presented our taxonomy of different types of harm and also considered how harm propagates. In a similar vein to the syntax presented in the asset level, here we seek to establish relationships between harms. The subject and object will be a type of harm and different actions will describe relations between these

types of harm. All the actions will be parameterised with characteristics, which will relate to the controls that organisations have in place to mitigate harm. Examples are provided below:

Data loss ***TRIGGERS*** Financial loss
Financial loss ***AMPLIFIES*** Reputational Damage

TRIGGERS – may cause another harm to be realised. For instance, loss of customers' data may lead to reputational damage.
AMPLIFIES – may exaggerate the loss from a type of harm, which has already been realised. For instance, regulator fines as a result of a cyber-attack lead to financial losses, and these financial losses may be further exacerbated due to the loss in customers.

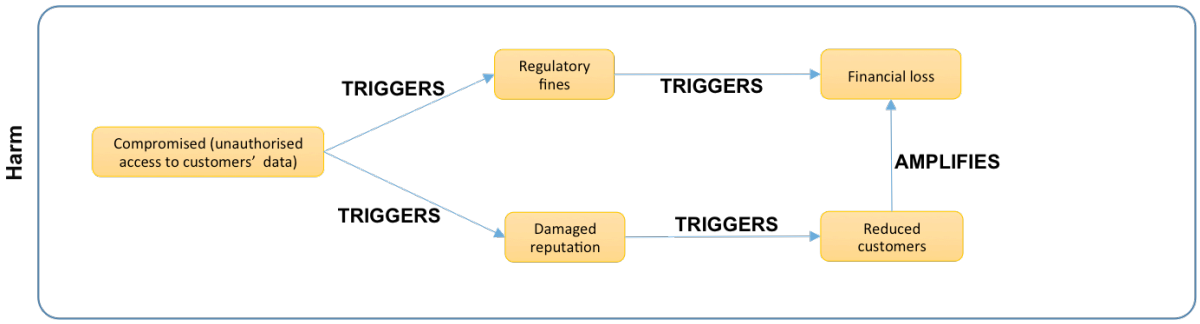


Figure 10: An example of harm propagation being modelled

Figure 10 illustrates how we may be able to reason about different types of harm occurring from unauthorised access to customers' data. Note that here we assume a single original harm for simplicity, but there may well be multiple. This therefore allows us to model cascading harms, and potentially also consider time factors, how harms may evolve, what other types of harm may be triggered and how likely is for a type of harm to be amplified further when another type of harm is present. At this level, it is important to have a clear understanding of cyber-harms but especially, appropriate methods for quantifying or qualifying those harms. It is only through some degree of measurement (whether guided by academic or industry perspectives[16,17]) that realistic estimations for the next level, which is *value* at risk, will be defined. The outcome will ideally be a more well-defined method to measure the various types of harms.

Cyber-VaR level

The final level to consider relations within is the Cyber-VaR level. Here different types of harm that occur to different assets and groups of assets, will determine the Cyber-VaR outcome. We also need to reason about the likelihood of this particular harm being realised. Our initial understanding indicates that there are at least two main factors to be considered regarding the likelihood of a harm being realised; the likelihood of an attack being successful and the motivation of the attacker. We describe our current approach in the next section.

[16] Kannan, K., Rees, J., & Sridhar, S. (2007). "Market reactions to information security breach announcements: An empirical analysis." International Journal of Electronic Commerce, 12(1), 69- 91.
[17] Deloitte (2016) Beneath the surface of a cyberattack: A deeper look at business impacts. http://www2.deloitte.com/content/dam/Deloitte/us/Documents/risk/us-risk-beneath-the-surface-of-a-cyber-attack.pdf

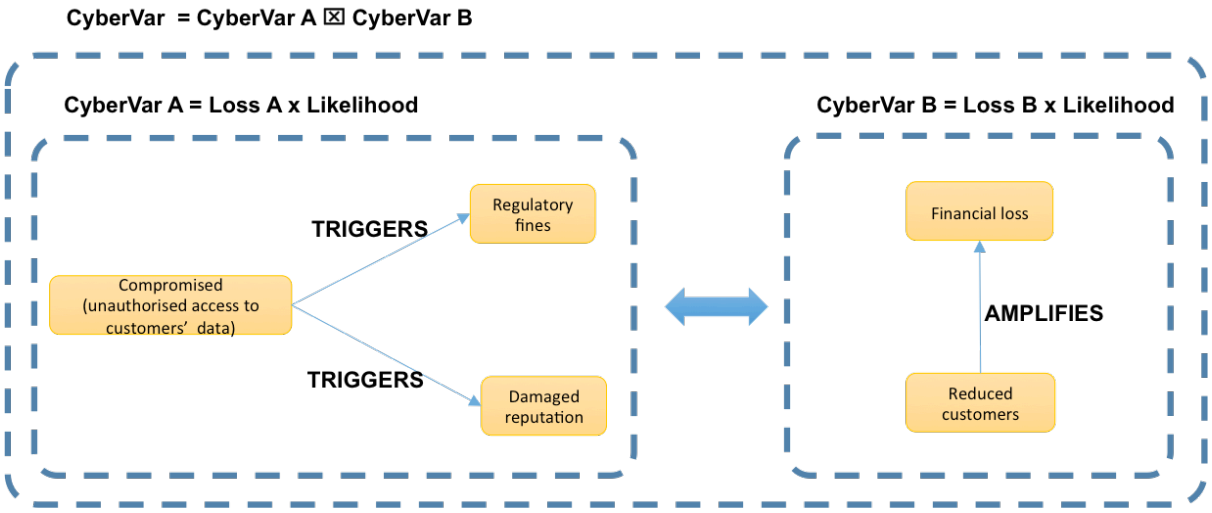


Figure 11: An example of reasoning about Cyber-VaR when multiple types of harm occur

Different types of harm (emerging from different attack-surfaces) will give rise to different values of cyber-VaR. There might be cases where more than one Group of harms must be considered on the Cyber-VaR level. Thus, we need to reason about how multiple Cyber-VaR values may be combined. Literature suggests that VaR is not sub-additive, thus we will need to characterise how one Cyber VaR value may influence another and what would the overall cyber-VaR value be.

An initial step towards this goal is to define a method to combine individual or propagating harms. Take Figure 11 as an example. The delineated area to the left captures Loss A and the one to the right is Loss B. We define Loss to be a group of harms. We first need to measure the losses associated with each of these areas by measuring their individual loss (e.g., for Loss B, Reduced Customer and Financial Loss) and then combining them in some way to determine the aggregate. Once we have the aggregate for that grouping, we would then calculate the group's cyber-VaR.

As mentioned above, once we have cyber-VaR values for each loss set, these will need to be combined in some way. The simplest case is an addition of the cyber-VaR values. Figure 11 demonstrates this process and shows how different losses, which may occur from a cyberattack, can be combined to calculate the cyber-VaR. The large arrow in the middle denotes that for the overall loss we need to combine, in some mathematical fashion (denoted by ☒), the two components of the overall loss, namely Loss A and Loss B.

A point worth mentioning at this stage is that all types of harm may not necessarily be quantifiable (i.e., capable of being *accurately* stated in a quantitative manner) and thus, on occasion qualitative metrics may be applied to better describe the situation. In fact, we have defined harm be a heterogeneous concept regarding measuring its components. When considering the cyber-VaR value though, we will need to compensate for the "uncertainty" in the numeric value which the qualitative measurement imposes. This introduces a new dynamic to our aggregation, which we reserve to explain in the next section.

5.4.2 Reasoning between levels

Having presented options for reasoning within the three model levels, this subsection focuses on reasoning *between* levels. More specifically, we will outline how information regarding the assets may be used to better determine loss, cyber-harm types and cyber-VaR, and how information from the harm level may be used to further reason about cyber-VaR.

Asset to Harm

Firstly, to reason between assets and types of cyber-harm, we use information regarding the importance of the assets, as well as the dimensions describing the features of the assets that may be exploited (i.e. digital vs physical). There are specific characteristics in the importance of assets that provide insights into which types of harm may occur. Additionally, information about the dimensions of the assets may be used when considering which attack-surfaces may be responsible for such a type of harm. Therefore, there is a relationship between the asset or group of assets and types of harm, defined by the action HARM. The action can be parameterised with the importance of the asset (i.e. Financial Value, Customer Data, Intellectual Property) and the susceptibility dimension (i.e. Digital, Mobile etc.).

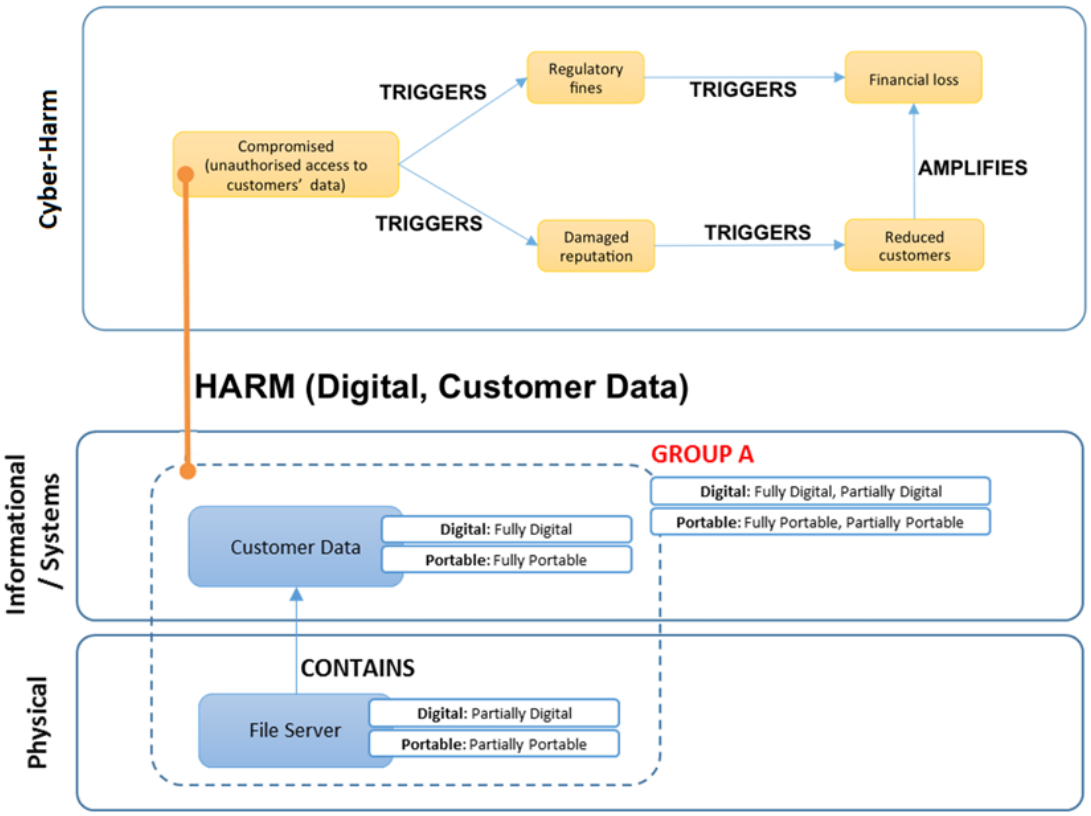


Figure 12: Example of a parameterised relation between a group of assets and a type of harm which cascades further

Figure 12 illustrates how we can reason about harms occurring to a group of assets (Group A). The action HARM is parameterised with Digital being the value for the susceptibility dimension and Customer Data being the value for the importance of asset component. Thus, we consider the direct source of harm to be the theft of customers' data via compromising digitally the file server; an alternative exploiting the physical dimension of Group A and resulting in a similar harm would be to steal the physical server; for such a scenario, the harm action would require not only the customer data component as a parameter but also the financial value of the server, triggering an additional harm named financial loss due to server being stolen. It is important to establish how an asset or a group of assets may be directly harm and insights on this subject will be sought during the focus groups and interviews with experts.

Harm to Cyber-VaR

Having established the link between the asset level and the harm level, we now examine the link between the harm level and the Cyber-VaR level. Before considering this relationship it is worth reflecting on the literature on VaR. The notion of VaR was first being established in the field of economics as a metric to estimate the risk of losing an investment on a given portfolio for a specific period of time[18,19]. Metrics developed based on the mark-market variable. There exist various models on how to calculate VaR, all of them sharing three main characteristics. The first is that every model provides the factors which are considered to calculate the mark-market variable. Most of the factors which are included in the equations are established through various economic theories. There might be parameters in these variables which need to be estimated. The second step is to provide a procedure to estimate the unknown parameters. There are two main avenues for achieving this: (a) through historical data and (b) running Monte Carlo simulations. The third step is to reason over the loss function.

We will follow the same rationale regarding reasoning for cyber-VaR, thus the first step is to establish what a loss in this context is and which factors are relevant for this loss. We begin by considering how to approximately derive a quantitative value for loss. Once a numerical value is established, the likelihood of this value occurring should be considered next.

In a simplistic scenario, it would suffice to examine data describing losses (numeric approximations) for the types of harm we are interested for a specific asset and estimate a continuous probability density function based on this data. For example, if there is a dataset regarding the cost from various data breaches occurring over a period of time, then a continuous probability density function can be designed based on this dataset. In this scenario there is only one factor influencing loss which we already have a value for. This is a rather naïve approach because it does not take into account any information provided in the harm level about the number of records held by the organisation or any other importance value occurring from the three-lens approach.

We could, however, parameterise this variable with factors such as records lost per breach, the size of an organisation etc. A better estimate can be obtained by dividing the overall loss per case with the number of records lost, thus obtaining the average record per loss per case. We can further refine this equation, if information regarding the proportion of records breached to the overall number of records held by organisation is available. Since we will have a good estimate of the number of records held by the organisation we are interested in calculating the loss, we can multiply the average loss record with the number of records the organisation has to project the loss which the organisation would have encountered if they were breached for every case. Then the likelihood of a loss exceeding a specific value would be derived by considering the appropriate integral of the continuous function.

In a similar vein, we may calculate the loss occurring from business interruption. In many cases measurements consider the value of stock in the stock-market exchange. We could identify the daily changes (the percentage) in the cases where a cyber-attack caused a business interruption, as well as changes in other factors influencing stock prices and estimate the loss for the organisation we are interested in, by projecting these changes into the value of the stock in the present time. However, business interruption value more likely results in a loss of profits or in extra costs and expenses over the period of restoration following a loss for an organisation. These are more specific issues but should be included in accurately measuring business interruption loss.

[18] Linsmeier, Thomas J., and Neil D. Pearson. "Value at risk." *Financial Analysts Journal* 56, no. 2 (2000): 47-67.
[19] Rockafellar, R. Tyrrell, and Stanislav Uryasev. "Optimization of conditional value-at-risk." *Journal of risk* 2 (2000): 21-42.

In this scenario we use data already obtained from organisations, thus we need to identify how these measurements occur and which types of harms they concern. We believe it is possible tol gain further insight from focus groups on how organisations measure loss and be able to reason about the types of harm often neglected in these measurements as well as how often these types occur. Once we have obtained this information, we could further parameterise the loss probability distribution function to cater for the neglected harm, exacerbating the loss based on a probability distribution of the neglected harm. Obviously this could then be refined over time, assuming an ongoing activity aimed at collecting this type of information (possibly an activity that a consortium within the insurance sector could undertake).

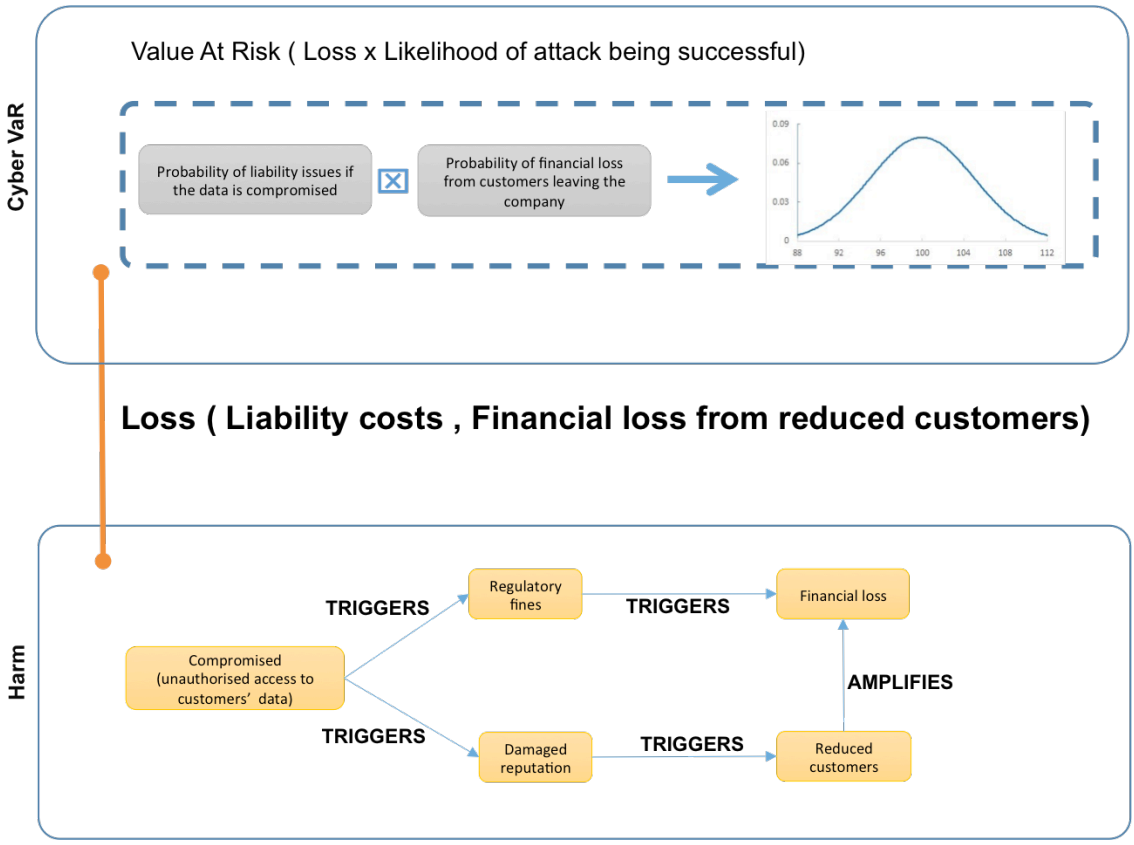


Figure 13: Example of a parameterised relation between aggregate losses from the harm level and probability of these losses occurring

Figure 13 illustrates different types of harm which may occur in a data breach scenario and how we may better reason for a cyber-VaR. Most of the numerical values when organisations consider data breaches focus on the liability loss for each customer record (upper branch of the harm propagation scenario). They often neglect though the financial loss occurring from customers leaving the company which suffered the breach. In this example, we produce the probability function of the total cyber-VaR taking into account distributions from both losses. A Poisson distribution could then provide further insight for the likelihood of a number of losses happening within a given time-framework. It is important to note that the approach described considers the likelihood of a loss occurring *given* a successful cyberattack.

A more sophisticated cyber-VaR model should make use of the other types of information available to the model, pertaining to the susceptibility of the asset to diverse attack surfaces. Since the dimensions of an asset or a group of assets determine how the harm may occur, then these dimensions can also be used to estimate the likelihood of a successful attack, guiding our reasoning on which attack surfaces are relevant for which assets and for which type of harm. Thus, we will be able to answer questions such as what is the likelihood of a loss occurring if a cyberattack takes place?

To reason about the likelihood of a successful attack, we need to understand which factors are important. We initially consider the motivation of an attacker and more specifically, how capable the motivated attackers are as well as the success rate of specific types of attacks. The loss function could be parameterised with these variables and Monte Carlo simulations could be run. We might also consider taking account of the issues outlined in Section 3 above – the variability of the controls set effectiveness in the face of a resourced attacker, and an attacker that can predict the control set configuration Returning to the example provided in Figure 13, it is as if we have additional information for every single breach about the type of the attack and the capabilities of the attackers. In addition, it would be ideal to have information on all the unsuccessful attacks which took place over the same period of time (although in reality we could not be certain that we had all of this information since we may not have actually detected all attacks prevented).

One important factor which will influence the Cyber-VaR that we have not mentioned yet is the effectiveness of the controls which organisations have in place. The effectiveness will be another parameter influencing either the occurring losses or the likelihood of a successful attack. Further details on how the effectiveness will be calculated are provided in the next section.

5.4.3 Applying controls across model levels

Model

Having detailed the three levels of the model and the relationships between them, we now move to consider controls. A control is a security mechanism put in place to serve a dual purpose: to protect an asset in some fashion or to mitigate the impact of an attack. Therefore, in our model, it may influence all three layers (assets, cyber-harm and cyber-VaR). It is evident that determining the effectiveness of a control fulfilling its purpos,e is of paramount importance. There are scant sources in the literature that provide evidence for or reason about, how to understand how effective a control is. Our intention is to shed light on how we can approach the problem for assessing how effectively controls protect assets and mitigate harm.

The first step is to understand what the purpose of a control is. We need to define for each control or set of controls (based on our reasoning about dependency of controls piece) actions to denote what asset a control is protecting, what the nature of risk to be treated is and what the nature of the harm being mitigated is. We define an action named *PROTECTS* which maps a specific control (or a set of controls) to an asset or a set of assets, parameterised by the argument *risk treatment*. We define a second action *MITIGATES* which maps a specific control (or a set of controls) to a type of harm, parameterised by the *asset* the control is applied to.

In Section 2.3, we demonstrated how controls may depend on other controls for enhanced functionality. Therefore, the presence or not of specific controls may have an immediate effect on the effectiveness of the dependent controls. It is crucial to capture these relationships between the controls and being able to reason about the extent to which the overall functionality of the ecosystem of controls is affected either positively or negatively. We define a function named *DEPENDS ON* to capture the dependency of a control (or a set of controls) on another control.

In terms of our syntax, we outline the relationship that a control may have as follows:

Control ***DEPENDS ON*** Control
Control ***PROTECTS*** Physical, Information / Systems, People, Enterprise
Control ***MITIGATES(customer data)*** Reputational harm

DEPENDS ON – relies on for functionality. For instance, Control A (e.g., Incident Response) relies on Control B (e.g., Computer and Network Monitoring); here it is only possible to respond to an incident once the incident has been discovered. This definition is used primarily related to risk controls though it can also be applied to assets, e.g., a Sales routine could depend on a payment processing routine.
PROTECTS – treats the risk. For instance, Control A (e.g., Anti-virus software) could treat the risk of malware infecting Asset B (e.g., an employee workstation or server). This definition describes the relationship between a control and an asset.
MITIGATES(asset) – constrains harmful situations in some fashion. For instance, Control A (e.g., back up data) may reduce the impact of an attack aiming at reducing the availability of data. This definition describes the relationship between a control and a harm over a specific asset.

The effectiveness of controls based on which assets they protect, how they mitigate harm and what dependencies exist amongst them will inform the cyber-VaR layer. We note that residual risk resides either on the way the controls are implemented (which implies dependencies on other controls) or on the design of the control (which implies that there are some inherent vulnerabilities). Based on evidence from datasets, which we will discuss later in this report, we will need to create probabilistic distributions of effectiveness of controls which will be taken into account in the cyber-VaR model. Having linked controls with assets and harm, we should be able to parameterise the Cyber-VaR level, with the controls which are relevant and their estimated effectiveness.

As we discussed previously, controls may exhibit one or both of two classes of aim. The first aim is to reduce the likelihood of a successful attack by completely avoiding a risk and removing the attack surface in consideration; or to diminish the risk by reducing the relevant attack surface; or increasing the work-factor or effort involved in conducting an attack, thereby making it less likely that a motivated threat will attempt to conduct it. The second aim, without excluding the first, is to reduce the impact of a successful attack, thus restricting the harm which may be suffered. We depict this visually below in Figure 14 focusing on one physical asset.

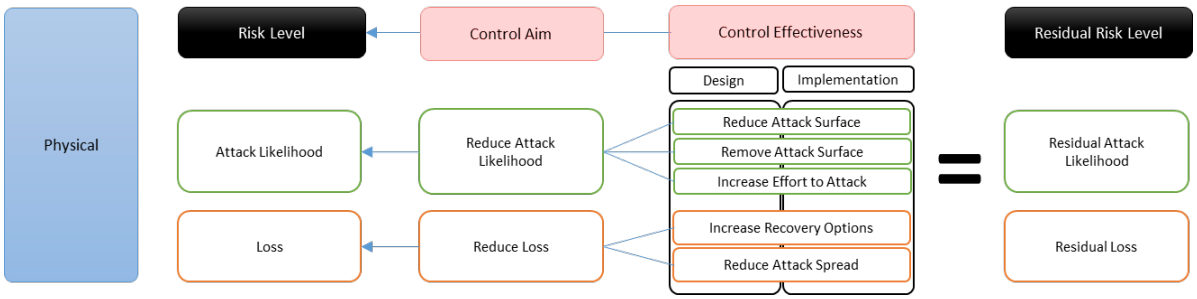


Figure 14: Depiction of residual risks for controls

When we run the models on the Cyber-VaR level, based on the type of control, we will provide parameters which will affect either the likelihood of an attack or the monetised value of the impact. We will also need to consider the dependency between controls and how this impacts residual risk.

We therefore need to establish an understanding on how the parameters for the effectiveness of controls and their specific probability distributions may change with the presence of other controls.

Figure 15 below illustrates the basic concepts of the model in all levels and the interactions across levels. Interconnected assets form a set characterised by business lens as customer data.

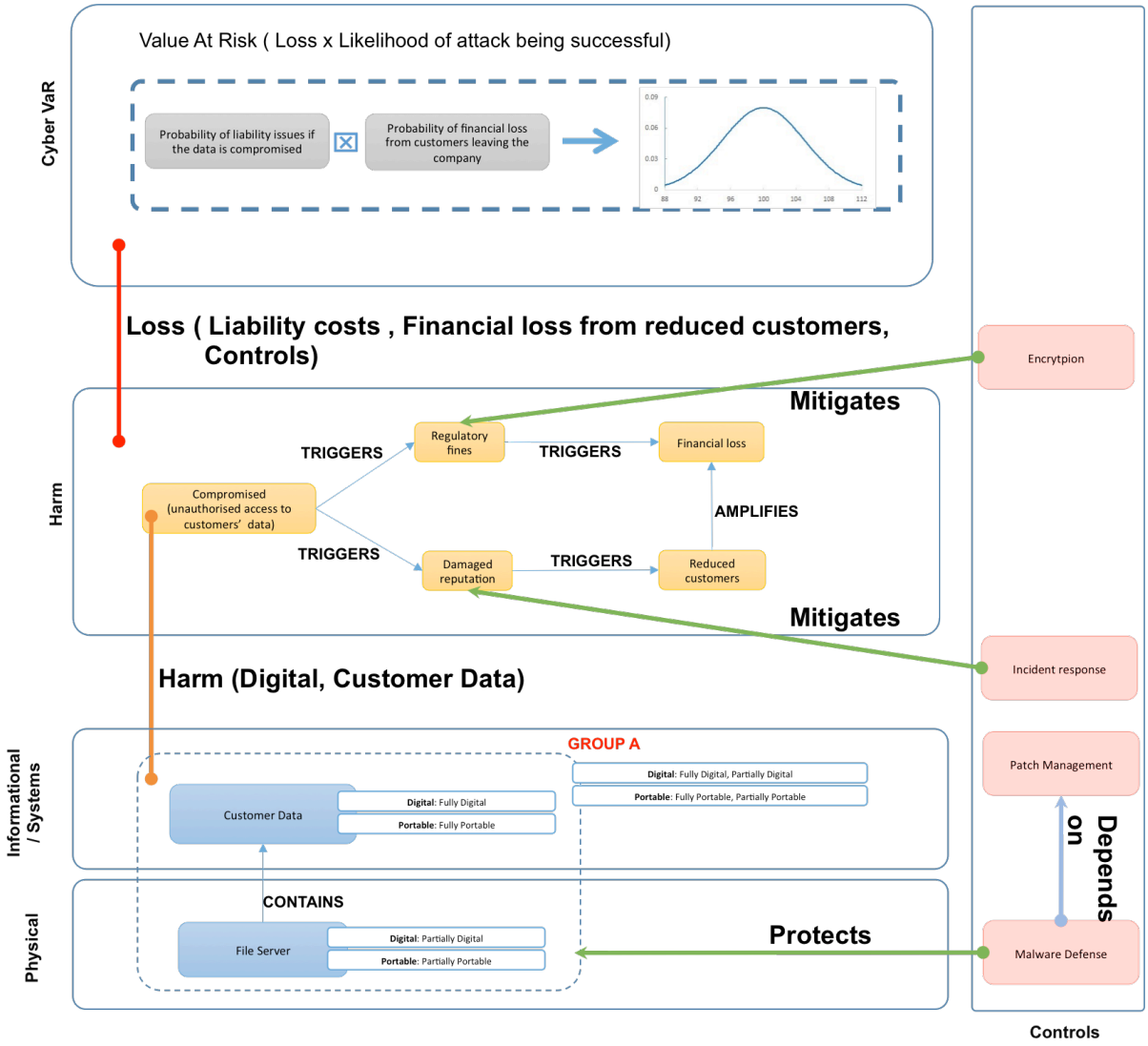


Figure 15: Describing the relationships within and across the model

There is malware defence, which depends on a patch management control and protects the set of assets. The harm which may be suffered in the second level is a compromised server, allowing unauthorised access to data about customers. This situation may trigger other types of harm and have a cascading effect resulting in financial losses, which are amplified by a scenario where there is a reduction in the organisation's customer base. Incident response and encryption are two controls which aim to mitigate the harmful situations in terms of reputation damage and regulatory fines respectively. The last step involves estimating the overall value at risk based on the different cascading harms and calculating a probability distribution for the likelihood of a successful attack resulting in such harms.

5.5 Validating the Model using Interviews and Focus Groups

In order to validate the model presented in Sections 5.3 and 5.4, we engaged in qualitative and quantitative research with security practitioners and experts from the corporate environment. The focus of our research was to understand how industry perceives and reasons about concepts and notions such as assets and harm, how decisions to deploy controls are made, and whether there is a procedure in place for assessing the effectiveness of those controls. To achieve our aims we conducted a small online survey, followed by interviews and a focus group. In total nineteen people completed the online survey and thirteen of these were further interviewed or participated in the focus group.

Focusing on the qualitative research, we held individual interviews with security professionals who had technical and/or business expertise. In total we interviewed six people from five different organisations. Our decision to interview people individually was informed by the fact that most security professionals would be influenced by the presence of other experts and would not disclose as much information or elaborate on past experiences. We decided, however, to hold a focus group with underwriters and brokers since they collaborate on a daily basis and they have complementary experiences. Thus, the interaction within the focus group provided much richer data for analysis. Seven people partook in the focus group (four underwriters and three brokers).

All interviews and the focus group were recorded resulting in eight hours of data. The data was analysed using content analysis. Content analysis is a systematic research methodology applied to analyse and describe phenomena[20,21] by "designing replicable and valid inferences from texts to the context of their use"[22]. Thus, it is a scientific technique that offers insights and in-depth understanding of the concept under study. In addition, content analysis facilitates the testing of conceptual models in order to verify theories and hypotheses[21,22]. Unlike quantitative research where the researcher compares scientific hypotheses with observed evidence, in content analysis these hypothesis are compared with inferences from the available text[22].

More specifically, we adopted a mixed approach of content analysis which requires starting with a deductive approach and in the second iteration obtaining an inductive approach. Deductive content analysis requires the existence of a theory or model to underpin the classification process. This approach is more structured than the inductive method and the initial coding is crafted by the key features and variables of the adopted theory or model. In the process of coding, excerpts are ascribed to categories and the findings are dictated by the theory or model developed prior to the research. Inductive content analysis is based on "open coding" and the categories are freely created by the researcher. In the open coding, headings and notes are written in the transcripts while reading them and different categories are created to include similar notes that capture the same aspect of the phenomenon under study[22,23]. The process is repeated and the notes and headings are read again. The next step is to classify the categories into groups. The aim is to merge possible categories that share the same meaning[22].

The initial themes used in the first iteration where deduced from our model. These were assets, harm, effectiveness of controls, control dependencies, cyber-VaR and cyber-insurance. During the inductive process a number of themes emerged, such as asset interactions, harm propagation, residual risk, metrics for effectiveness, likelihood of attack, motivation of the attacker and market maturity.

[20] K. Krippendorff. Content analysis: An introduction to its methodology. Sage Publications, Inc, 2004

[21] K.A. Neuendorf. The content analysis guidebook. Sage Publications, Inc, 2002.

[22] S. Elo and H. Kyng¨as. The qualitative content analysis process. Journal of advanced nursing, 62(1):107–115, 2008.

[23] H.F. Hsieh and S.E. Shannon. Three approaches to qualitative content analysis. Qualitative health research, 15(9):1277–1288, 2005.

5.5.1. Assets

Regarding assets, the vast majority of the participants noted that there are two-fold, data and business processes. Regarding data, interviewees suggested that the emphasis is on either personal information of customers and clients or information about the business. Underwriters and brokers concurred since data is "the primary source of most claims on cyber policies". Focusing on business processes, participants deemed as assets systems and processes that hold information or are fundamental for the function of the organisation, since there are "data on one side and systems on the other". There is a growing realisation that people should be considered as part of core assets "as each person has an idea of what's critical to them". Other examples frequently mentioned by the participants include IT infrastructure, any type of hardware, devices, as well as "things that may connect remotely to the network of an organisation". It is worth noting that more abstract notions, such as "culture", "reputation", "policies, direction and strategies" were being considered as assets.

Identifying sensitive and core assets is a process which participants claimed organisations excel in since "organisations are quite sophisticated in classifying what they have". Key business functions, legislation and harms which may occur over time are the drive factors for characterising assets. A temporal element on the sensitivity of assets was acknowledged as well. Participants deemed that assigning criticality is a process transient in nature because "some data is more critical and valuable to us in certain times of the year". A classification system often used by organisations classifies assets in terms of secret, sensitive and non-sensitive. We need to emphasise that the interviewees had several years of experience in the security space that may explain their confidence in identifying critical assets.

Focusing on how assets are interconnected, most participants reasoned about assets by starting with data and concluding that IT infrastructure is viewed as an asset because it either involves data or facilitates key processes. This rationale implies that participants consider how assets are connected before deciding what it is critical; especially when harmful situations are used to determine which assets are core.

There is a good understanding in the community about how main IT infrastructure supports processes and where sensitive data is stored; this understanding is limited to "IT or very structured business parts" only and depends on the type of the organisation due to regulatory frameworks. Participants recognised that "quite a lot has been forced upon organisations in certain sectors" and that they have tried to use "very granular methodologies which worked fine in a technology environment or a very controlled production environment". These granular approaches do not "translate well to people who run the business". As a consequence, participants with technical background may be confident on assessing criticality of assets regarding IT equipment, however, estimating the value of business processes is a more complex issue and a range of people holding different roles must be involved.

The use of remote devices and cloud services, as well as the presence of legacy systems create complex network infrastructures and raise the difficulty in identifying how assets are interacting with each other. It was suggested that "most organisations do not know what is connected to what. They just typically add more devices, software etc. as the company evolves". Establishing a better understanding of how assets are connected may change the way organisations perceive the criticality of assets. We believe that even for the well-established processes in IT environments, the criticality of an asset may be reconsidered once the interconnectivity aspect is taken into account.

Many participants emphasised on the role of legacy systems and the lack of motivation from organisations to replace these with modernised versions. Interviewees deemed that innovative

solutions will enhance visibility of interconnected assets. They noted that "always the legacy systems are the most problematic" but companies focus on profits and if "the cost [of an incident] is less than the cost to upgrade the system then they will not do it. Many times people in organisations care about the cost efficiency (convenience/business requirement) instead of understanding exact dependencies between assets, and how this is managed". These problems result in a very low maturity level in understanding interactions between assets.

Reflecting on how the findings from the interviews validate our model, participants mentioned assets from all the categories we presented in Section 2.1. This validates our approach on the asset level of the model, where we distinguish assets in different categories based on their nature (i.e. physical, processes, people, enterprise). No further additions were made in the list of assets presented. Considering the classification scheme which organisation use, this information may inform the business value obtained through the different lenses and provide a more accurate estimation of the overall value of a set of assets. Finally, the low maturity in understanding how assets interact with other assets highlights the importance of providing a language in our model to facilitate this reasoning process for organisations.

5.5.2 Cyber-Harm

Considering the notion of harm, the vast majority of the participants suggested that financial loss and reputational damage are the most prominent harms that organisations face. Financial loss is suggested to be experienced either via the cost of responding to an incident or through the theft of funds. Reputational damage revolves around negative publicity and the impact in "ability to maintain or attract customers"; as one participant pondered, "it is reputational damage really. It is not good press, we do not care about the data but the fact that it happens".

Other harms identified during the interviews are business interruption, loss of customer trust, regulatory fines, loss of personal data, and publication of sensitive plans. The overall effect of these harms may place organisation in dire situations and there is a shift recently from limiting the reflection of cyber-harm just in IT systems to considering more fundamental business risks. To quote one participant "for many years we, as an industry, focused on the damage to IT and small components of infrastructure or systems. With the breaches in recent years and the coverage, people see that the impact is much more than destruction of IT. It's like any other business risk, which can take down the business".

It is worth noting that some participants believe that we have not experienced yet the full spectrum of harms which may occur from cyber-attacks, since we do not know yet the effect of a second successful attack in an organisation or how the landscape of data breach may change once a successful class action occurs. As one participant stated for the scenario of another successful attack "if you have one breach you can get away with this but if you have two? There was a guy who had to change 27 different payment forms. If you keep getting that you will get problems, you don't get too many chances". Regarding the class action, this is mainly a concern in the healthcare environment in terms of a data breach and thus far "none has seen the tail of it".

The main incidents via which harms may be realised were human negligence, which may lead to loss of data or to unavailability of services (ransomware), and Distributed-Denial-of-service (DDoS) which results in business interruption. An interesting finding is that organisations do not attempt to "measure the degree of loss or harm from incidents", rather they focus on corresponding in an ad-hoc fashion to every incident, since "it's more a case of dealing with incidents when they happen, worrying about any fall out and dealing with it quickly, and then going back to normal operations". The rational being that "there will be a successful attack at some point. We can survive if we say we

have been doing everything we should have done by now and yes they went through. It is how you respond, that is our focus and the next level of maturity". This is in contrast to the aspect above whereby the mitigation need and the preparedness are poorly understood.

Measuring the impact from cyber-attacks is "a very difficult problem" and sometimes is considered irrelevant to the IT department since "measurement of loss is for the risk department and the operational risk team [and] it should be a business-driven exercise". Further challenges in estimating losses rise from the need to differentiate these from other "general churn of customers and other issues". Even in cases where participants reported recent attacks which their organisations experienced (i.e. email servers being off service for two days) they found it challenging to reason about metrics and suggested that the "person-time spent dealing with incidents" may be appropriate, "for e.g., if it took five days to resolve and it was five days of a technical resource and two days of a project manager, then we can roughly work out the cost". An interesting exercise would be to understand how much of this cost is additional and how much is `business as usual'.

Participants suggested that more granular approaches which focus on the infrastructure level may provide some reasoning with quantifiable approaches via the use of Key Performance Indicators (KPIs) and Key Risk Indicators (KRIs). How representative these indicators may be depend on the maturity of the organisation, since more experienced organisation are "accustomed to dealing with risks on a day-to-day basis" as well as on the type of harm. Underwriters suggested that the "insurable loss is easy to quantify because you have paid it". There are however, harms such as reputational damage which is still often deemed uninsurable due to difficulties in estimating losses. As one participant noted "these are all things that we probably don't have a way of measuring, and one could probably only give a very high-level idea". A potential suggested measure is "the amount of negative information that reaches major papers or national news". Another frequently mentioned example of a type of harm which may be problematic to estimate is business interruption, which is trivially calculated based on profits, but "if there is no profit it does not make much sense". In a similar vein, IP issues, especially for businesses in their infancy, are impossible to calculate since "the sky is the limit". We should note that the concept of propagation of harm as described in our model should shed further light on how to better assess business interruption and IP issues.

Assessing harms individually may be a challenging task; considering secondary and aggregate losses or harms seems an impossible one. People deem that "the pace which the companies work at, it's really a case of dealing with the incidents as they come up and security than understanding such issues such as aggregation". All participants concurred that organisations have a very low level of understanding of aggregate and cascading harms with the exception of data breaches. The insurance community have identified key areas to be considered such as "forensics, notify people, call centres, start defending it, class actions and lawyers' fees, PCI, fines and assessments costs" and "a lot of the understanding of these losses may be based on historical data on security". However, this is a "rough estimation" and context and legislative frameworks are critical in determining costs, which may fluctuate significantly.

A possible approach to obtaining a better understanding of cascading harm is through the use of workshops with the participation of "many people at the director level who know what the key assets and harms are". The result of such workshops would be harm tables where by focusing on assets and how harms may be realised on these, further harms are identified by individuals by "considering what would result next if that harm were to materialise". Once these harm tables are established and the risk appetite of an organisation is decided "then the aim would be to have controls that meet that appetite level". It is common for determining the appetite level to consult "likelihood vs impact type matrices".

Reflecting on the findings from the interviews, there are no further harms identified to enrich our taxonomy. The rationale, however, suggested for reasoning about harms and their cascading effects concurs with our approach as presented in Section 4.2. Historic information may provide a better understanding on how harms are realised and which harms may give rise to other types of harm. We have started analysing different case studies to establish patterns of how harm cascades. Harm is a heterogeneous notion and having just "a quantitative approach doesn't work". "There is not enough information to make it a really effective quantitative method, thus we need to be able to have discussions about orders of magnitude". In our model we do not attempt to necessarily quantify harm and cascading effects and provide the opportunity to determine harm in more qualitative terms. Once a better understanding of harm is established, the quantitative processes are taking place in the Cyber-VaR level.

5.5.3 Deciding on controls, their effectiveness and dependencies

There is a range of factors determining how organisations decide on which risk controls to implement. Participants noted that an important element is incidents occurring in the past and how these relate to the risk appetite of the organisation. There is "risk discussion" and people based on what have seen most reflect on the "inherent risk (untreated) and how various controls treat this risk" as well as on the residual risk "assuming these controls are in place". Some participants note that for such an approach to be effective, organisations have to focus on the value of the assets first, as "for some they will go on the call to what they need to do to protect the assets". Then next step requires to understand how these assets may be harmed, the "organisation's risk appetite and tolerance" and finally choose those "controls which would best fit those needs".

Legislation and industry regulators are important, albeit contextual, factors that are considered "things that you have to do". A common example frequently mentioned by participants was the increase in significance of monitoring controls "because regulators and legislation [such as DPA and the new GDPR] are stating that organisations need to declare breaches/incidents within a certain time". These controls are increasingly attracting the interest of board members. Participants also mentioned that organisations focus on things they know they must do as best practice, which eventually will render their security posture compliant with the regulatory restrictions. As one participant claimed, "the better, more sophisticated organisations say, let's try to make us more secure rather than compliant to frameworks".

There are several standards/frameworks which organisations follow, with ISO 27000 being reported as the most widely adopted. The reason for this is that ISO 27001 is frequently requested by the public sector and considers physical, technical and personal aspects. As one participant noted, "getting it is good, it makes you feel like you are doing good stuff. What it doesn't do is give you a list of the things you must do. It just gives you the principles you should follow." Additionally, ISO is known amongst board members and "they know it can be a tick in the box". Other frameworks mentioned during the interviews are CSC SANS 20, which are considered as a "sanity check", highly technical controls and difficult to comply with. The main drawback is considered to be the lack of administrative controls.

Adopting a standard is informed by the location and the context within which an organisation operates. In the UK Cyber Essentials and the Ten Steps from the UK Government are considered a "good means of on-going communication", providing "a nice block". HIPPAA is widely considered in healthcare organisations while NIST is increasing in popularity, becoming (as suggested by brokers) together with ISO "the bulk of the discussion people now have when they go into board rooms". An interesting finding is that organisations experience a lock-in effect when adopting a framework. It is rather difficult to change to another framework. Organisations try rather to adopt the next version of the framework they currently adhere to. This behavior may become troublesome due to the fact

that controls from a specific framework may not be adequate to ensure a certain level of security. The lock-in effect, however, implies that organisations are not willing to adopt different Key Performance Indicators (KPI); thus compliance to a specific standard may provide a false sense of security. A problematic area identified by brokers and underwriters is the assessment process both in self-assessed frameworks and when external vendors are involved. Some organisations "may have the same mate [friend] to come in to assess them, resulting in compliance becoming a tick-box exercise".

Participants also mentioned that organisations are "looking at what peers are doing" via industry forums or other means, before deciding on which controls to implement. Obtaining a secure posture may result in a competitive advantage, especially when organisations operate in environments where data security is deemed of paramount importance by their clients. The argument here is that the most critical factor, which shapes decisions on controls, is the budget available for security. Therefore, "if you could surpass your peers [in terms of implementing risk controls] it could benefit the company". As the majority of the participants acknowledged "it gets down to almost a pure business decision for most commercial organisations". Therefore security professionals have to convince the board that the "cost of the control and its operation come to a lesser amount than the value of the original system or the profits they would lose". Such reasoning, however, requires facts and figures about the effectiveness of controls.

Reflecting on the effectiveness of controls, participants concurred that it is a big and outstanding challenge, emphasising that it is "one of the holy grails that people in cyber are looking for". Establishing a benchmark for controls' effectiveness is crucial not only to inform decision on budget allocation but also for better understanding of what it means to reduce risk from one level of the harm table to the next, exhibiting signs of maturity towards a more sophisticated security posture.

A common practice which organisations follow, is to establish yearly control reviews using KPIs and KRIs as metrics. These metrics are based on regulators' feedback, recurring incidents and the progression of the industry sector. It is important for a successful control review to establish baselines on acceptable performance and seek input from risk teams on emerging risks. It is evident that this practice is tailored for each organisation since "the metrics selected need to be meaningful for the organisation. Gathering many metrics (especially on low-level technical controls) that can't be translated usefully for the company may not be that useful". Data is gathered by monitoring systems or from control data, e.g., "how much we are hit by viruses now, how much malware is on systems".

A complementary technique in measuring effectiveness is penetration testing on an annual basis, which sheds insight into whether the organisations are improving and whether controls are effective against vulnerabilities. Measuring availability of services is another metric often adopted by organisations, as well as training programmes and using tests (before and after) for employees to see whether they have improved. The latter is a rather important metric emphasising on how well individuals comply with organisational policies. As participants noted, "people make mistakes; and a lot of companies rely on policies but not have controls in place". Therefore relying on a person to comprehend and comply with a policy without the presence of controls requires appropriate training and education.

There is a general consensus amongst the participants that the interdependency of controls underpins a successful framework for assessing their effectiveness. There is an ecosystem of controls which relies on how well controls function in tandem. Participants also acknowledged that the presence of specific controls may boost the performance of other controls. Overall, participants suggested that the interdependency of controls is considered "via the onion model of security" and

the “principle of defence in depth”, e.g., designing controls that overlap each other, resulting in not having dependence on a single control. An example mentioned to highlight this rationale is that of the traditional AV, which is becoming less and less effective. However, the presence of a control for authorised software and whitelisting helps to mitigate a lot of the deficiencies in traditional anti-malware products.

Participants suggested that there is a reasonable understanding of dependencies when technical controls are considered and more “people take a look across all the controls and not focusing on one”. However, a number of interviewees suggested that focusing on technology controls only has an impact on how the risk is reduced or what level of protection is provided. This is due to the fact that “technologists haven’t been the best at communicating what these controls do” and the board members deem them to be complex. Furthermore, although dependencies are considered, being able to assign a value to how much more effective a control is when it functions in tandem with other controls is still an elusive task. In some cases, organisations may employ “vendors to come and do the analysis of dependencies as well as decide what updates to the controls are needed”.

A slightly different approach on deciding how to apply controls, and one worth elaborating further, was mentioned by two participants who are employed by a finance organisation. The organisation decided to invest on a SIEM system due to the fact that they lack the ability to correlate events. The other perimeter control system was coming to its end of life, which forced the organisation to replace it. The SIEM tool which the organisation purchased encompassed a list of recommended metrics to record its performance. These metrics relied on building a baseline of network activity and recording anomalies which were not false-positive events (i.e. events indicating anomalous activity without a reported incident).

The recommended metrics, despite being suggested by the SIEM vendor, were not deemed to be part of their marketing strategy. The organisation spent a year filtering false-positive events, trying to understand how the network is segmented and which assets required closer monitoring. They soon realised that their network infrastructure hindered the performance of the SIEM tool because it bombarded it with voice and conference equipment data, rendering the anomaly-detection and correlation of events useless. Also they noticed that current practices such as logging administration password changes created further noise, since their management system required daily changes.

Instead of changing the metrics for the SIEM system, they decided to change the network configuration to facilitate the use of the SIEM tool. At present, after engaging in a lengthy restructuring process, they are now able to correlate events, describe their baseline and log events they consider relevant. A criterion for adopting a new control is how it may fit and enhance the SIEM tool. In this example, an organisation, due to the fact that it was forced to replace the majority of their ecosystem of controls, modernised their security posture. They decided to change policies and the network configuration to facilitate the use of the new controls and the metrics for assessing their effectiveness. However radical this approach is, participants claimed that they “have a very good understanding of the dependency of controls. Because we built it from the ground up and incrementally added services.”

Despite the methods mentioned in this section for determining the effectiveness of controls, there was a general consensus amongst the interviewees that it is very difficult to report on effectiveness because “you do not know what is going on in the network”. Starting looking for what may constitute an acceptable behaviour of the network is the first step, while the second is the enforcement of organisational policies. However, “organisations don’t generally think through the extent of connected assets and aggregated harms in control selection”. To make matters more

complicated, they focus on IT security while neglecting information security, rendering the whole process of control implementation "piecemeal and a bit reactive".

Reflecting on the effectiveness of controls, our hypothesis to assign controls to assets and harms is validated. Most approaches consider how controls protect specific assets and how risks acknowledged through risk assessments may be mitigated with the presence of controls. It is also important to highlight that the dependency of controls is widely accepted as a critical factor for their effectiveness by the community. We believe that our model caters for all the cases mentioned in this section.

5.5.4 Cyber-VaR

Participants were not familiar with the concept of cyber-VaR. They were, however, familiar with concepts such as *likelihood of an attack occurring* and *motivation of an attacker*. When reasoning about estimating Cyber-VaR, the majority of participants suggested that it must involve "looking at assets and their value, how vulnerable they are, and the probability of that vulnerability being exploited" (either via a third party or inadequate security controls). Once those are established then it is feasible to argue about the risk level of the organisation and whether this risk is acceptable or further actions are required. If the latter is the case, then it "boils down to what controls we can put in place to make the risk acceptable".

Participants pondered that organisations have processes in place for assessing the likelihood of attacks. They suggested that being aware of the business environment and your competitors as well as monitoring reports about attack patterns and recent vulnerability scan reports are fundamental to the success of this process. Additionally, observing how the threat-actor landscape is changing, either based on current experiences or historical data on attacks, may enable a model for making predictions. Other sources include the Open Threat exchange, CERT online forums, reports from GCHQ and alerts from international organisations of "security stand point". Trending threats at the moment are considered spear-phishing and DDoS attacks.

Building on this understanding, underwriters suggested that organisations need to "start with who is motivated to attack them, where the threats are coming from" and then move on to consider "how well they protect themselves from these attacks". They believe that reasoning in terms of probabilities is challenging due to the volatile nature of the cyberspace. As new vulnerabilities emerge, criminals move lateral to attack in novel ways and businesses operating in different environments, rendering historical data regarding past attacks irrelevant. As participants suggested "the criminals will have different methods and go out for different business and companies will have a different set of controls". It is hard to predict what target will be deemed profitable for the attackers. To quote one participant, "one year is credit cards, information sold to embarrass a particular set of individuals, next year is healthcare information, attacks on Linux. What is the next thing? It is difficult to derive to".

Participants also elaborated on issues due to poor maturity of the market in cyber-insurance. Cyber insurance is a relative new market and data may "take another ten years to become rich enough. At the moment you try to maybe buy some data that exist and try to say I have this system in place for fifteen years so I may be talking about a *one-in-fifteen year* event". However, insurance companies base their decision-making on events whose likelihood is estimated in *one-in-two hundred* years. There was a consensus though that understanding and analysing claim forms from past incidents will provide additional sources for predicting events.

The findings from the interviews reinforce our understanding of the cyber-VaR term. Motivation of the attackers, the likelihood of attacks and the value of assets are key requirements for a model able

to estimate a cyber-VaR value. All these characteristics are included in our reasoning as described in Section 5.4. Historical data may shed light into how the threat-landscape may change or how motivated attackers may be to launch an attack on a particular organisation. The challenges recognised by the underwriters, however, require further analysis of historical data and the creation of datasets based on data from claim forms. Different approaches for analysing such data should be considered to cater for the volatile nature. Therefore, novel approaches should focus on identifying abstract characteristics which remain constant over a larger period of time (i.e. instead of focusing on a particular type of attack, observe how often new types of attack are instigated in the historic data).

5.5.5 Cyber-insurance

The majority of the participants reported that cyber-insurance holds an instrumental role in enhancing the security posture of an organisation. As interviewees noted "a lot of people are taking more cyber insurance out. We are seeing that cyber clauses are being added on to all types of insurance policies, property, office content, etc.". Participants deemed that cyber insurance will become the norm at least in more mature industries. As on interviewee acknowledged, "in the last few months people have asked me where are you in terms of cyber insurance. It will be a tick-the-box exercise because it will kill the conversation".

Other interviewees focused on how recent developments in legislation may add value in the cyber insurance market, in particular when "more regulation is coming down the line, and more fines, and increased breach notification costs, supporting customers who have been negatively affected". As one participant stated "things will change with the new GDPR. When you are threaten with a fine of 25% of your operating cost or profit for a data breach I am pretty sure that there will be organisations who want to cover that". This argument is reinforced by the recent developments in the UK legislation where participants claimed that "it was suggested that they would like to see company directors being personal liable for the PII that is held in the organisation. Our directorship if I left my organisation tomorrow and put a request to purge my data and they don't do it then there is a possible fine there".

In some cases participants claimed that cyber-insurance products are promoted by board members when a number of cyber incidents has been reported in the organisation. This concurs with the opinion of brokers who deem that "it is easy to sell when there are loads of really visible claims". Another interesting finding is the claim of cybersecurity that insurers will probably be in a key position to suggest some "absolutely fundamental controls which have to be in place". They believe that insurers will oblige organisations to undertake a proper risk assessment on their systems, demonstrating that they can cope with the risk and essentially proving "that organisations do what they need to do to protect their information". Additionally, interviewees suggested that cyber-insurance would be rather significant for the board members since "they want to make sure that they are not liable, that their shareholders will not remove them from their jobs".

At the same time though, participants deemed that the cyber-insurance market is very nascent. They suggested that the main obstacle hindering the boost of the market was the lack of a comprehensive offering. Participants view cyber-insurance as "all-or-nothing" products, which contradicts with the fact that large organisations possess insurance for different situations and what really need is a "pick and mix menu-driven approach". Interviewees suggested that there is scope for insurers to work with cyber specialists to help define covers and better understand the needs of the companies regarding cybersecurity.

The same views are shared by brokers and underwriters, who believe that there is a "lack of transparency and understanding of the product". Brokers may try to explain the product to their

point of contact in organisations but cyber-insurance may be so complex that "you explain it [cyber-insurance] to the guy, says it is fantastic and when a week after he tries to talk to the board about it he cannot remember half of the things you told him". There is also a lack in brokers who have the knowledge and training to understand the product and sell it. As on participant explained "if someone does not understand the product they will not talk about it; education is a massive piece".

Other reasons which hinder growth in the cyber-insurance market are lack of legislation (especially in the US) and lack of visibility of impacts that cyber-attacks have in organisations. Brokers stated that due to the lack of publicity in cyber incidents they have only anecdotal evidence to discuss. They believe that only the presence of one big event in the public sphere may boost the market. To emphasise on this point, brokers mentioned examples in the US where "big [breach incidents] hit the press and, even in the middle market level, every single client who was buying $5,000,000 before, they started all of a sudden buying $10,000,000 and $20,000,000. Even though they cannot relay themselves to target etc. it is in the back of their mind". What is noteworthy is the fact that organisations may not experience a breach before seeking cyber-insurance. It suffices if their peers experience an attack.

Underwriters suggested that from their point of view it is difficult to make transparent to the IT departments of the organisations that they do not necessarily dictate which controls are right to be implemented and which are not. In addition, many reports use fancy words and complicate policies, contributing to the problem of complex products.

Finally, a number of participants deemed that the premium of cyber-insurance policies was too expensive for the risk these policies were covering. Only one participant suggested that insurance companies may not have the appropriate internal mechanisms to accurately assess the risk which may suggest that they lack the experience to provide prices in premiums that will be realistic. A project however, that attempts to provide a model to accurately assess cyber harm is of great importance and may increase the confidence of organisations regarding the premium values of the cyber insurance policies. As it was stated "insurance companies have a vested interest to build an assessment process and probably that is something you can sell more than an insurance policy".

5.6 Data requirements to operationalise the model

Thus far, we have presented and detailed our model which documents our current thinking on the relationships that hold between security-risk controls, the assets which these controls seek to protect, the value-at-risk and the different types of harm which may occur in a typical organisation. To use this model as well as to validate its effectiveness, there are several data requirements at each level. Data is crucial as it will enable users of the model to reason about numerous aspects including the links between certain assets and cyber-harms (e.g., typical harms that result from certain assets), the likely propagation paths of harms (e.g., specific harms that are likely to result due to other harms), the probability distributions that allude to the likelihood of particular losses, and effectiveness of risk controls. For some of these points, for instance the propagation of harms, the model may be able to initially rely on historical data and trends therein to make inferences. Indeed, we have already identified and modelled some of these trends based on our focus groups and interviews discussions with industry professionals.

In what follows, we define these requirements more clearly and explain how they would be applied in the use of the model.

Asset to Cyber-harm

The association of assets to harms (or negative impacts) that might occur to them is one of the traditional components of security risk assessment. As such, the immediate harms that would result

to most assets if attacked are generally well-known. For example, if a server is taken over by a malicious party, then it might be unavailable or its data may become compromised. What is currently lacking in the literature and in practice, however, is a *clear structure of the direct and indirect harms* that result from cyber-attacks on assets.

We have scoped out these harms in the attached spreadsheet and in our examination of real cyber-attack scenarios (in Appendix 2) but this information could be enriched if there was *an extensive dataset with a record of assets (and interacting sets of assets) and how they were harmed, and how those harms cascaded and resulted in other harms*. *Severity and timing of each of these harms* could also prove very insightful in identifying trends based on that data, which would allow us to further inform our model. This would allow organisations to use the model to infer exactly which harms might occur to assets that they have, and potentially even consider the context of their organisation (assuming the organisation had data for this or we could identify such industry trends from the historical data gathered).

The data might be gathered from sets of organisations which have had their assets attacked and incurred losses (including how those losses were measured), or in partnership with third-party organisations that gather and correlate such data (e.g., Advisen Cyber Dataset[24]). In the first case, cyber insurers may be best placed to gather data from their customer base using claims data (e.g., what assets were attacked, how they were attacked, and what loss may be claimed), though a wider partnership across either cyber insurers or clients (e.g., via industry forums or sharing platforms) would be more informative. If relying on third-party reports, the limitations in that data should be clearly understood including the extent to which they capture all related assets and attacks, and how exactly loss types and loss values have been defined. Only this way can users of the model be confident that once such information is fed into the model, that the model will produce valuable output on harms.

Cyber-harm to Cyber-VaR

Associating harms to Cyber-VaR values is perhaps the most challenging task of the model. Harm is a heterogeneous notion and may comprise of qualitative data. Cyber-VaR though requires only quantitative values. Reasoning about qualitative data in quantitative terms, requires approximations of the reality. Thus, there is a need to assign a value of loss to every type of harm that will reflect to reality as accurate as possible. This value in money loss from a harm would be the value to lose in the Cyber-VaR model. In order to calculate this value, we need historic data of claims and data sets where organisations have revealed the overall cost they encountered due to cyber-incidents. We would then need to establish which types of harm occurred in these scenarios. It is critical to establish the rules for analysing datasets. Identifying which features in these datasets are relevant is the cornerstone to an effective approximation of reality. Data elaborating on the effectiveness of controls to mitigate specific types of harm should be considered as well. The presence of controls and their influence on the final loss may be determined from claim forms and if specific controls are deemed essential in reducing the impact of a harm, then this should be taken into account in estimating a loss probability distribution.

Another important factor for linking harm to cyber-VaR is the likelihood of the attack occurring. Each harm type is associated with a specific asset as described in the previous section. Therefore, it is possible to reason about the attack surface that may cause such harms to assets. To estimate a likelihood of attack probability, we need to gain insight from current practices which organisations follow. As described in earlier sections, organisations have a good understanding of the threat landscape which they face and how it changes. They base their reasoning on vulnerability scans, historical data on attacks, CERT online forums, black-hat conferences and reports from security

[24] Advisen. Advisen Cyber Dataset. http://www.advisenltd.com/analytics/advisens-cyber-dataset/

vendors. All these datasets may be used to infer a probability distribution for the likelihood of specific types of attacks occurring. Of course, the effectiveness of controls in place should be considered. Data relevant to determining the effectiveness of controls to specific attack types should inform the probability function and reduce the likelihood of an attack occurring when appropriate. The Advisen dataset may also be used to capture trends in types of attacks.

The final step is to reason about the motivation of the attackers. Participants during the interviews noted that organisations have a good understanding on which type of attackers are motivated to act against them. Data from organisations such as Cyence[25] may be of use to determine a distribution for estimating the motivation of an attacker. In addition, the location and the context within which organisations operate will influence this distribution. Observing how the number of attacks fluctuates in a particular sector through various reports from security vendors could be another source of data to be taken into account.

Effectiveness of Risk Controls

Risk controls are another critical component in our model. They protect assets by reducing the likelihood of successful attacks and/or reducing the harm suffered by assets if attacks are successful. Through these two points, controls can be projected across each of the three model levels. In our work thus far (see spreadsheet), we have mapped controls (using the CSC20) to the assets they protect, considered their mitigation nature (i.e., targeting harm and/or attack likelihood), and also recorded general reports of their effectiveness. The topic of control effectiveness is a particularly important one for our research as it would allow us to consider the value of implementing one control versus another, and broadly benefits to compliance. From our analysis, we found the topic of control effectiveness to be rather underexplored and underreported. There is little data to support how effective controls might be at protecting an asset, and even less data concerning to what extent a control would reduce cyber-harm or attack likelihood (leading to cyber-VaR).

As an initial step to the broad requirement of having data to relate control effectiveness to harm and attack likelihood, we have examined the CSC20 again. CSC20 is useful here because along with the controls that are proposed, there is also a definition of specific types of data that may be used by an organisation to assess the effectiveness of the control (our spreadsheet now presents such details). This therefore attempts to give each organisation the ability to have some internal understanding of effectiveness – of course, this is only applicable after the control has been implemented for some time to allow the necessary data assessment and reflection.

If we consider *CSC-7: Email and Web Browser Protections* as an example, the CSC documentation states that data which would be useful to gather includes: *number of unsupported email clients that have been detected on the organisation's systems (by business unit); number of events of interest that have been detected recently when examining logged URL requests made from the organisation's systems (by business unit); the percentage of devices that are not required to utilize network based URL filters to limit access to potentially malicious websites (by business unit); and the percentage of the organisation's users that will inappropriately respond to an organisation sponsored email phishing test (by business unit)*.

If control-effectiveness information provided via metrics such as those mentioned above could be closely mapped to assets or correlated with the protection they provide to assets, this might provide an opportunity for the actual effectiveness of controls to be applied to the asset level. This would therefore allow us to propose statements such as: "Based on the assessed metrics, Control X is 60% effective at protecting Asset Y, and 80% at protecting Asset Z". This could be evaluated against

[25] Cyence A Unique Cyber Risk Modeling Platform For The Insurance Industry https://www.cyence.net/

another control's effectiveness or be used as basis to determine the residual risk level maintained by the organisation.

In terms of the relating control effectiveness to the cyber-harm and cyber-VaR levels, similar data requirements to those mentioned above also exist. Here, the aim would be to use the nature of the control (e.g., whether it addresses harm or likelihood) and based on that nature, consider the extent to which harm and/or likelihood would be reduced through its implementation. One set of data that would be required is that of the *loss probability distributions of organisations* (as was mentioned above), and how they have traditionally been impacted by controls; for instance, as certain controls have been implemented by an organisation, how much has the probability, or amount, of losses shifted.

Additionally, data pertaining to *trends in previous attacks and the proportion of times in which controls in place prevented the attacks' success or reduced its harm* could also be useful at informing the model. This may draw on some of the metrics data available with control sets such as CSC20. This would allow us to extend our work beyond the impact of controls on the cyber-harm and cyber-VaR levels, to potentially consider how much an improvement in a control's effectiveness would have on those higher layers. Using such information, we would then be able to compare whether it may be better to invest in new controls to reduce harm and/or cyber-VaR, or to increase the effectiveness of existing controls.

6. The Relative Effectiveness of Risk Controls and the Value of Compliance

The aim of this project is to explore the relative effectiveness of risk controls to the security posture of an organisation, and so the value of compliance to security standards and frameworks prescribing said controls. We have fulfilled this aim by critically reflecting on the key concerns of organisations, namely, organisational assets (i.e., things of value to the organisation), cyber-harms (i.e., the range of negative impacts that can result from cyber-attacks) and cyber-VaR (i.e., the likely loss if a cyber-attack occurs); and also on how the security standards and controls act to address these concerns. This is both in terms of general protection of assets by controls and more specifically on how exactly controls protect these assets, for instance, in the reduction of attack likelihood or exposure to harm.

To assist in our analysis of the effectiveness of controls we have designed a model that is capable of relating risk controls to assets, cyber-harms and cyber-VaR. The value of the model is in its detailed coverage of each of these levels, and the facility it provides to reason within and across these levels. For instance, the model we propose is novel in its provision of a simple approach to create a system of assets, link these to related harms that could occur, and map these to value-at-risk for an organisation. As controls are selected, these could be mapped to related assets, and their broad impact accounted for in the context of harm and value-at-risk.

To validate our model and shed light into how the presence of controls may affect assets, harms and the Cyber-VaR level, we engaged in qualitative research and interviewed security professionals with extensive expertise in both business and technical aspects of security. Our findings suggest that organisations have a good understanding of how to assess the criticality of assets when IT infrastructure is concerned. Evaluating business processes and more abstract notions such as culture or reputation still proves a challenging task. The way assets are linked remains an area which organisations have limited understanding, especially when cloud services and remote devices are considered. This low level of understanding may influence the organisations' decisions on the criticality of assets. Our model provides a language to describe interconnections and allows

organisations to better understand the value of their assets (either physical assets or more abstract assets).

Our findings further identified that financial and reputational harms are the most prominent types of harm. Organisations are ill-equipped to measure the impact of harms in general and a more qualitative measuring system is adopted in the form of harm tables. Once these tables are designed, organisations decide on which controls to implement based on their risk appetite. The notion of cascading harm was acknowledged by participants as important, however, there is no evidence that organisations consider its effects when addressing threats. They rather respond to incidents in a more ad-hoc manner.

Regarding cyber-VaR, our findings suggest that being able to assess how harmful a situation for a set of critical assets is, as well as what is the likelihood of a successful attack that may result on such a harm or the motivation of an attacker to execute such an attack, is the first step towards obtaining a reliable outcome. Participants reasoned that determining the effectiveness of controls that mitigate impact or protect assets holds a pivotal role in estimating cyber-VaR.

While the model created provides the foundation to examine the impact of controls across the levels of harm and value, we have encountered a dearth of industry and academic data regarding the true effectiveness of controls. This factor affects our ability to operationalise our model by using effectiveness data and practical insight[26]. This lack of data is apparent both in our literature-based review and the subsequent period of consultation with stakeholders about which controls are found to be most effective and why. Additional challenges rise by the fact that the effectiveness of a specific control may depend on the presence of another control. Our findings suggest that the interdependency of controls is an area which some security experts consider, however it is still a concept in its infancy and further research should be conducted to provide data on how controls depend upon each other.

Another interesting and relevant finding that arises from our analysis is that control selection is often not driven by effectiveness, but rather by regulation, legislation or trends in threats. This is important to note because it highlights a potential disconnect in controls selected by companies, which could be the reason for current inadequacies in organisation's security postures.

Given the level of importance that understanding the effectiveness of controls has for our model, in combination with the lack of data about control effectiveness that exists, this report has also outlined the key data requirements to operationalise the model. These requirements address the areas of mappings between assets and harms and harms and cyber-VaR, in addition to specifically presenting aspects for control effectiveness. Claim forms and datasets such as Advisen may provide useful insights in assessing the effectiveness of controls. A key requirement for our system is to identify the appropriate characteristics for every dataset. Data obtained through metrics for technical controls (such as CSC) must be correlated to data from claim forms to obtain a better picture of how effective controls were in preventing or mitigating an attack.

As an example, there might be a metric showing that 30% of people working for an organisation are prone to phishing attacks. However, not in all cases a successful phishing attack will lead to data exfiltration. Linking data indicative of the effectiveness of technical controls to data from claims may

[26] The ideal would have been to use existing data about the effectiveness of a control and then map that across our model to ascertain the level of protection (in terms of assets, cyber-harm and cyber-VaR) it offers, and then compare that with the implementation of another control. Key questions would be – does it perform better? Does it overlap significantly? Do both controls actually address the same issues, so potentially are redundant? Through these types of questions, we would be able to determine the value of complying to controls (individually or together), and the overall impact on an organisation's security posture.

provide a better understanding on the overall effectiveness of a control in terms of harm and cyber-VaR.

Our model is the first and decisive step in determining the effectiveness of controls. Further research however, should be conducted:

- Implement a prototype software tool of the model proposed, which would be capable of determining the potential range of impacts of a risk control upon exposure to harm. This might be developed with a selection of estimate probability distributions based upon knowledge in the community, and with which it would be possible to test the sensitivity of results in a range of scenarios.
- Design a methodology for learning the impacts of risk controls within an organisation using software sensors with an organisation's infrastructure. This would enable the collection of data to establish the probability distributions required by the Model (in 1 above) through aggregation of results across multiple organisations and identification of general patterns. This approach would have the added benefit of allowing organisations to consider results tailored to their specific operations.
- Additionally, further consideration is required of how different datasets may be linked to provide quantitative evidence on how effective controls are. This new approach should take into consideration the interdependency of controls how effectiveness of the ecosystem of controls may change if certain controls are not present. Further exploration of historic data is required to identify features that will be abstract enough to provide useful information regarding the likelihood of an attack, even when the data is considered obsolete (i.e. on systems which are not used any more).
- The Model should be extended to address other classes of harm, from natural disaster or accidental insider actions (for example, where our research in other projects leads us to believe that current risk controls are often inadequate).
- We should also consider expanding the findings from the qualitative research. A possible next step could be to conduct large scale questionnaires, focusing on which controls are widely accepted in the industry and what metrics are used to determine their effectiveness. Additionally, interviews and focus groups could take place to emphasise on how the interconnection of assets may change the way organisations perceive assets, as well as identifying how cascading harm may occur and which types of harm are triggered. Interviewing lawyers will shed light into how recent developments in legislation may influence the way organisations reason about controls and whether cyber-insurance will become a norm, enabling insurers to suggest a set of desirable controls to hedge risk.
- Specific research should be conducted into the relationship between harm, and an assets level of digitisation. We need to know if it is the case that the level of digitisation has a consequence for the likelihood of susceptibility to successful cyber-attack, and the potential for harm and cascading harms within an organisation. New controls might be suggested.

- Specific research should be conducted into the value of unpredictability in control usage as a mechanism for improving cyber-defenses to reduce cyber-harm. By understanding cyber-Var a better risk-transfer model for the value of cyber-risk mitigation investment or cyber-risk insurance can be developed.

- In further analysing cascading harm the types of cascading harm organisations share across industries or wider must be considered to understand the systemic risk organisations face and the potential this creates for catastrophic cyber-harm scenarios.
- Another interesting piece of research may revolve around 'Kill chains'. Kill chains are attack orientated, however, our model focuses on harm-propagation. By overlaying the different steps in the harm propagation to a similar to kill chain format, we may be able to form a 'crisis response' kill chain for mitigating impact and treating risk.
- Another aspect on which we will focus is determining the balance between threat detection, data loss prevention and business functionality. Often controls may impose obstacles in the manner that a business functions. A balance should be reached to ensure the financial viability of the organisation while maintain an appropriate security posture.

Acknowledgments

We gratefully appreciate the funding and assistant of Novae Group plc for this research.

Appendices

Appendix 1: CIS Critical Security Controls (CSC) 20 Background

The CIS Critical Security Controls (CSC) are a set of 20 prioritised and well-vetted actions, activities and tools that have been put forward to assist organisations in improving their state of security[27]; these controls are often referred to as the SANS20. They have been derived from an understanding of the threat environment (including the main types and vectors of attack) and current technologies used within organisations. At this time the most recent version of controls is CSC version 6.1, released in August 2016. This also highlights another advantage of this control set, i.e., it is updated regularly and in response to the current threat environment.

The 20 controls in CSC version 6 are as follows:

- CSC 1: Inventory of Authorized and Unauthorized Devices - Actively manage (inventory, track, and correct) all hardware devices on the network so that only authorized devices are given access, and unauthorized and unmanaged devices are found and prevented from gaining access.
- CSC 2: Inventory of Authorized and Unauthorized Software - Actively manage (inventory, track, and correct) all software on the network so that only authorized software is installed and can execute, and that unauthorized and unmanaged software is found and prevented from installation or execution.
- CSC 3: Secure Configurations for Hardware and Software on Mobile Device Laptops, Workstations, and Servers - Establish, implement, and actively manage (track, report on, correct) the security configuration of laptops, servers, and workstations using a rigorous configuration management and change control process in order to prevent attackers from exploiting vulnerable services and settings.
- CSC 4: Continuous Vulnerability Assessment and Remediation - Continuously acquire, assess, and take action on new information in order to identify vulnerabilities, remediate, and minimize the window of opportunity for attackers.
- CSC 5: Controlled Use of Administrative Privileges - The processes and tools used to track/control/prevent/correct the use, assignment, and configuration of administrative privileges on computers, networks, and applications.
- CSC 6: Maintenance, Monitoring, and Analysis of Audit Logs - Collect, manage, and analyse audit logs of events that could help detect, understand, or recover from an attack.
- CSC 7: Email and Web Browser Protections - Minimize the attack surface and the opportunities for attackers to manipulate human behaviour though their interaction with web browsers and email systems.
- CSC 8: Malware Defences - Control the installation, spread, and execution of malicious code at multiple points in the enterprise, while optimizing the use of automation to enable rapid updating of defence, data gathering, and corrective action.
- CSC 9: Limitation and Control of Network Ports, Protocols, and Services - Manage (track/control/correct) the ongoing operational use of ports, protocols, and services on networked devices in order to minimize windows of vulnerability available to attackers.
- CSC 10: Data Recovery Capability - The processes and tools used to properly back up critical information with a proven methodology for timely recovery of it.
- CSC 11: Secure Configurations for Network Devices such as Firewall Routers, and Switches - Establish, implement, and actively manage (track, report on, correct) the security configuration of network infrastructure devices using a rigorous configuration management and change control process in order to prevent attackers from exploiting vulnerable services and settings.

[27] CIS. The CIS Critical Security Controls (CIS Controls) https://www.cisecurity.org/critical-controls.cfm

- CSC 12: Boundary Defence - Detect/prevent/correct the flow of information transferring networks of different trust levels with a focus on security-damaging data.
- CSC 13: Data Protection - The processes and tools used to prevent data exfiltration, mitigate the effects of exfiltrated data, and ensure the privacy and integrity of sensitive information.
- CSC 14: Controlled Access Based on the Need to Know - The processes and tools used to track/control/prevent/correct secure access to critical assets (e.g., information, resources, systems) according to the formal determination of which persons, computers, and applications have a need and right to access these critical assets based on an approved classification.
- CSC 15: Wireless Access Control - The processes and tools used to track/control/prevent/correct the security use of wireless local area networks (LANS), access points, and wireless client systems.
- CSC 16: Account Monitoring and Control - Actively manage the life cycle of system and application accounts – their creation, use, dormancy, deletion – in order to minimize opportunities for attackers to leverage them.
- CSC 17: Security Skills Assessment and Appropriate Training to Fill Gaps - For all functional roles in the organisation (prioritising those mission-critical to the business and its security), identify the specific knowledge, skills, and abilities needed to support defence of the enterprise; develop and execute an integrated plan to assess, identify gaps, and remediate through policy, organisational planning, training, and awareness programs.
- CSC 18: Application Software Security - Manage the security life cycle of all in - house developed and acquired software in order to prevent, detect, and correct security weaknesses.
- CSC 19: Incident Response and Management - Protect the organisation's information, as well as its reputation, by developing and implementing an incident response infrastructure (e.g., plans, defined roles, training, communications, management oversight) for quickly discovering an attack and then effectively containing the damage, eradicating the attacker's presence, and restoring the integrity of the network and systems.
- CSC 20: Penetration Tests and Red Team Exercises - Test the overall strength of an organisation's defences (the technology, the processes, and the people) by simulating the objectives and actions of an attacker.

Each of these controls contains between 4 and 14 sub-controls detailing specific steps that should be taken to achieve the requisite level of security.

Appendix 2: Case Scenarios

This appendix presents a set of real-world case scenarios of cyber-attacks and the resulting harms.

Discussion of Scenarios

The Sony cases

In April 2011, amid unstable economic conditions, Sony announced that personal information for 77 million PlayStation Network (PSN) subscribers as well as 24.6 million Sony Online Entertainment accounts had been exposed due to an external breach[28]. The data breach involved information about account logins, passwords, credit card details, purchase histories and billing addresses. Sony's facilities in Japan were also heavily impacted from the earthquake of March 2011 resulting in the suspension of several critical operations, which rendered the cyber-attack well-timed to inflict maximum damage. Sony had to get its PSN services offline the day following the attack[29] to assess the extent of the incident, resulting in loss of revenue, incurred response costs regarding identifying and addressing the vulnerabilities, notifying the customers and calculated a rough estimate of $171 million costs. This figure, however, did not include the punitive damages from lawsuits, costs from identity theft, any other misuse of stolen credit cards and the loss of business and market capitalisation[29].

In late April 2011, Sony provided a comprehensive recovery plan and an accurate calculation of the costs inflicted from the earthquake, they were still yet unable to calculate the full organisational harm from the cyber-attack[28]. The aggregated impact of the earthquake and the data breach resulted in a significant decrease in Sony's market evaluation as depicted in stock-exchange markets. Sony's share price dropped 19% after the earthquake, a drop equivalent to the general economy but soon recovered 50% of this loss. After the cyber-attack, however, the Sony's price, unlike the rest of the Japanese economy, sustained a 12% loss and the security weaknesses revealed once Sony had restored service, prolonged the recovery phase[28].

Three years after these incidents, in November 2014, confidential data from Sony Pictures was once again leaked. The data included more than 30,000 internal documents, 170,000 emails, social security numbers of Sony's employees, personnel reviews and medical histories, and movies which had not yet been released. The same cyber-attack paralysed all of Sony's systems, rendering the online database of stock footage unsearchable, the telephone system offline, computers and servers unusable; this was described by the FBI as an "unprecedented digital assault that would have felled 90 percent of companies it hit"[30].

Sony was forced to replace a large number of its systems, set up a hotline for identity fraud, provide psychological counsellors for employees and organise seminars on data security. Following the attack, Sony's employees received emails threatening their families if they did not denounce Sony, their credit cards were available for sale on the dark market and some witnessed their bank accounts exceeding credit limits. A survey conducted by the Identity Theft Resource Center regarding victims of identity theft, reported that victims experience "denial, frustration, rage, fear, betrayal, and powerlessness in the days, weeks, and years after the violation"[30]. Class-action lawsuits from employees were filed; either because Sony did not notify those whose data was leaked or over fears of how personal leaked information could be potentially used. Furthermore, the press

[28] Dark Reading. Sony data breach clean-up to cost 171 million dollars. 2011. http://www.darkreading.com/attacks-and-breaches/sony-data-breach-cleanup-to-cost-\$171-million/d/d-id/1097898 (10 August 2016, last accessed).
[29] PWC. Limiting the impact of data breaches: The case of the Sony Playstation network. 2011. http://www.strategyand.pwc.com/reports/limiting-impact-data-breaches-case (10 August 2016, last accessed).
[30] Slate. Inside the Sony hack. 2015. http://www.slate.com/articles/technology/users/2015/11/sony employees on the hack one year later.html (10 August 2016, last accessed).

unfolded Sony's diversity issues which were discussed extensively in the content of the leaked emails [30,31].

The JP Morgan case

JP Morgan Chase, one of the largest banks in the US, reported that hackers obtained administrator access to several of their servers. Information regarding names, phone numbers, email and physical addresses of account holders was exfiltrated, affecting 76 million households and seven million small businesses. JP Morgan announced an increase in their cybersecurity budget of $250 million per year[32]. The company was forced to replace the majority of its IT infrastructure, a process which was time-consuming and hindered the daily lives of employees. The remaining budget was spent hiring more than 1,000 employees to monitor the company's systems[33]. Of significant interest are the two long-term effects which resulted from this hack. The majority of the customers whose information was leaked were obliged to monitor their finances in fear of fraud, while they received fake emails directing them to impostor websites for financial exchanges. As a result many became victims of financial fraud. The second effect was the replacement of the security chief because of their inadequate collaboration with federal authorities, in an attempt to try to control the investigation and obscure the leakage of information[33].

The Ashley Madison case

In July 2015, 33 million accounts and personal information about people registered on Ashley Madison, a website facilitating extramarital affairs, were leaked[34]. The core principle of Ashley Madison's business model was privacy and security to build a trust relationship with their customers. The cyber-attack therefore had dramatic consequences for the reputation of the company, not only because it exposed the vulnerabilities of the system but because it proved that Ashley Madison's promise to delete data upon customers request was not kept[35]. Because of this practice, Ashley Madison became liable to lawsuits[35], with many organisations soliciting litigants on Twitter[36]. What is of great interest in this case, however, are the repercussions of what was coined as "collateral damage" which are peculiar to the nature of the services the website offered.

Once the data was publicly available and easily searchable, customers became susceptible to blackmail, with professional and personal ramifications[36]. Many of the leaked email addresses contained the ".mil" domain, indicating people who serve in the US military. Adultery, however, is a crime in the US military and members of Ashley Madison were subject to a year of confinement or dishonourable discharge[35]. In a similar vein, owners of 1,200 ".sa" email addresses were exposed to a potential death sentence, which is the punishment in Saudi Arabia for adultery. New practices of cybercrime emerged, with criminals threatening to expose people whose email addresses were found in the Ashley Madison dataset to their "significant other"[37], unless the amount of $225 was paid in bitcoin. Public figures were coerced into "painful personal admissions"; others were divorced, while the Toronto police reported two suicides potentially linked to the cyber-attack[37].

[31] Variet. 2015. Sony hack attack opens minefield of legal questions that has Hollywood worried. http://variety.com/2015/ biz/news/sony-hack-attack-opens-minefield-of-legal-questions-that-has-hollywood-worried-1201471664 (10 August 2016, last accessed).
[32] Guardian. JP Morgan Chase reveals massive data breach affecting 76m households hack, 2015. http://www.theguardian.com/business/2014/oct/02/jp-morgan-76m-households-affected-data-breach (10 August 2016, last accessed).
[33] Tech Week Europe. JP Morgan security exec reassigned after breach, 2015. http://www.techweekeurope.co.uk/e-management/jobs/ jp-morgan-exec-reassigned-171644 (10 August 2016, last accessed).
[34] InfoSec Institute. Ashley Madison revisited: Legal, business and security repercussions, 2015. http://resources.infosecinstitute.com/ashley-madison-revisited-legal-business-and-security-repercussions (10 August 2016, last accessed).
[35] Verge. The mind-bending messiness of the Ashley Madison data dump, 2015. http://www.theverge.com/2015/8/19/ 9178855/ashley-madison-data-breach-implications (10 August 2016, last accessed).
[36] Guardian. Top data security expert fears traumatic aftermath in Ashley Madison hack, 2015. https://www.theguardian.com/technology/2015/aug/19/ashley-madison-hack-outcome (10 August 2016, last accessed).
[37] National Post Ashley Madison aftermath: Confessions, suicide reports and hot on the hackers trail, 2015. http://news.nationalpost.com/news/canada/ashley-madison-aftermath-confessions-suicide-reports-and-hot-on-the-hackers-trail (10 August 2016, last accessed).

Outline of Assets and Harms from Cyber-Attack Scenarios

Target Breach between Thanksgiving and Christmas 2013

Assets

- Computers at third-party vendor
- Web application portal for the use of Target's vendors
- Internal network
- Internal servers
- Point-of-Sale systems
- Customer data (including names, mailing addresses, phone numbers, email addresses) of up to 60 million customers
- Customer data (credit card details) of up to 40 million customers

Harm

- Compromised customer data (unauthorised access to customer data)

(Intermediate)

- Angry customers (lashed out at the company's customer service hotline's perpetual busy signal)
- Drop in sales
- Profits dropped almost 50 percent from the same time the previous year
- Announced 10 percent discount the weekend before Christmas
- Forensic investigator costs
- PR / Media response costs
- CEO fired / resigned as a result of the breach
- CIO fired / resigned as a result of the breach
- Target's share price dipped from $62 before the crisis to $56 one month later. At this writing, Target's share price is now at around $82 [JUNE 2015].
- Serious instances of fraud (for customers)

(Long term)

- Offered free credit monitoring for one year for affected customers
- An estimated $252 million paid by Target to manage the breach (an estimated $90 million offset by insurance)
- Hit with class-action lawsuit
- Legal fees
- A $10 million pot in escrow set aside for customers who can prove their accounts were seriously compromised (this relates to the class-action lawsuit)
- Additional training for employees on how to better keep customers' information safe
- Overhauled its security systems to identify internal and external risks to shoppers' personal info
- $100 million for more advanced registers and other technology to process new, safer cards
    - Rolled out (new) EMV-compliant POS terminals in all of its stores nationwide
    - Reissuing its store-branded REDcards as chip-and-PIN cards
- Reimbursement to thousands of financial institutions as much as $67 million for costs incurred (agreement struck with Visa Inc. on behalf of banks and other firms that issue credit and debit cards)
- Payment of $39.4 million to resolve claims by banks and credit unions that said they lost money because of breach (agreement via MasterCard Inc.)

References

http://www.zdnet.com/article/the-target-breach-two-years-later/
http://www.huffingtonpost.com/eric-dezenhall/a-look-back-at-the-target_b_7000816.html
http://www.wsj.com/articles/target-reaches-settlement-with-visa-over-2013-data-breach-1439912013
http://www.reuters.com/article/us-target-settlement-idUSKBN0MF04K20150319

TalkTalk Breach October 2015

Assets

- Website applications
- Computer software
- Computers
- Database
- Customer personal data (nearly 157,000 customer's details – names, addresses, dates of birth, phone numbers)
- Customer personal data (nearly 15,000 customer's payment details – bank account numbers & sort codes)

Harm

- Compromised customer data (unauthorised access to customer data)
- Offline websites (DDoS)

(Intermediate)

- Costs for cybersecurity firm hired to investigate the hack
- Angered customers
- Offered free service upgrades to customers
- Negative media reports (E.g., customers having their bank accounts cleared out, even though none of the data stolen could be used to access bank accounts)
- Criticism of company by information commissioner's office
- Lost 101,000 customers
- Loss of customer trust
- PR / Media response costs
- Closed down online sales operations
- Gained fewer customers
- Customers victim of scams
- Criticism of company by security experts
- Lower pre-tax profits (pre-tax profit for 2016 fell to £14m, compared with £32m last year)
- CEO & company having to defend itself & give evidence in governmental committees
- £400,000 fines by regulatory body (ICO)
- Potential ransom demands

(Long term)

- Offering free credit monitoring to prevent fraudsters from setting up credit cards in customer's name
- Suffered costs of £60million (disruption of services and exceptional costs) / some estimates say £85 million

References

https://www.theguardian.com/business/2016/feb/02/talktalk-cyberattack-costs-customers-leave
https://www.theguardian.com/technology/2015/oct/23/talktalk-criticised-for-poor-security-and-handling-of-hack-attack
http://www.independent.co.uk/news/business/news/talktalk-fine-data-breach-theft-customers-information-stolen-record-penalty-a7346316.html
https://www.ft.com/content/2144b2f2-1813-11e6-b197-a4af20d5575e
https://www.theguardian.com/business/2015/dec/15/talktalk-hack-could-not-have-been-prevented-by-cyber-essentials

Ukraine Power Station Attack December 2015

Assets

- Computers
- Account credentials
- Internal networks (SCADA)
- Power management applications
- Firmware on critical devices
- Power distribution centres
- Power substations (30 substations)
- Circuit breakers at power substations
- Backup power supplies of power distribution centres
- Call centre

Harms

- Compromised computers (unauthorised access to systems)

(Intermediate)

- Legitimate operator access to power station controls blocked
- Disabled backup power supplies for power distribution centres (where operators are based)
- Confused power station operators
- Power outage for approximately 230,000 customers (in an especially cold time of the year) for one to six hours
- Frustrated customers
- Customers without heating
- Call centres unavailable (denial-of-service conducted by attackers to stop calls to the energy company by customers)
- Customers unable to find out updates on the problems
- Costs for cybersecurity firm hired to investigate the hack

(Long term)

- Systems not fully operational for months after attack, e.g., some systems manually operated
- Public fear from knowledge that cyber-attacks can cause disruptions in critical national infrastructure

References

https://www.wired.com/2016/03/inside-cunning-unprecedented-hack-ukraines-power-grid/
https://ics-cert.us-cert.gov/alerts/IR-ALERT-H-16-056-01
https://ics.sans.org/media/E-ISAC_SANS_Ukraine_DUC_5.pdf
http://arstechnica.co.uk/security/2016/01/first-known-hacker-caused-power-outage-signals-troubling-escalation/

Appendix 3: Controls and Vulnerability to Threats

In this appendix we present a table of our detailed findings with respect to the CSC 20 controls and the extent to which they may be vulnerable to threats and attacks. The table has three columns: Controls and sub-controls; Descriptions of these controls, which are taken directly from the CSC 20 documentation; and Inherent vulnerability and weakness to attack, where we highlight what aspect the control is targeting (i.e., attack prevention, or attack detection and limitation), the inherent vulnerabilities of the control and the way in which an attacker may exploit those weaknesses. In this main document, we present CSC 1 to 10 as an example of the analysis. Further detail on the inherent vulnerabilities of the other controls can be found in the attached spreadsheet.

Controls and sub-controls	Descriptions	Inherent vulnerability and weakness to attack
CSC 1: Inventory of Authorized and Unauthorized Devices	**Actively manage (inventory, track, and correct) all hardware devices on the network so that only authorized devices are given access, and unauthorized and unmanaged devices are found and prevented from gaining access.**	**The inherent vulnerability and weakness to attack of this control is detailed below according to its specific sub-controls.**
CSC 1.1	Deploy an automated asset inventory discovery tool and use it to build a preliminary inventory of systems connected to an organization's public and private network(s). Both active tools that scan through IPv4 or IPv6 network address ranges and passive tools that identify hosts based on analysing their traffic should be employed.	This control generally seeks to prevent threats from launching successful attacks. The inherent vulnerability here is that virtual machines and wireless devices may periodically join the network, rendering the inventory of devices dynamic and complex. An attacker could use virtual machines or access to them to join the corporate network to launch an attack.
CSC 1.2	If the organization is dynamically assigning addresses using DHCP, then deploy dynamic host configuration protocol (DHCP) server logging, and use this information to improve the asset inventory and help detect unknown systems.	This control generally seeks to detect threats before they launch successful attacks. An attacker that has a legitimate MAC address may impersonate it to get onto the network undetected.
CSC 1.3	Ensure that all equipment acquisitions automatically update the inventory system as new, approved devices are connected to the network.	This control generally seeks to prevent threats from launching successful attacks. An inherent vulnerability here is that not many systems can achieve integration with an authoritative asset inventory and the asset acquisition process. An attacker may target the inventory systems to disrupt the effectiveness of the control.

CSC 1.4	Maintain an asset inventory of all systems connected to the network and the network devices themselves, recording at least the network addresses, machine name(s), purpose of each system, an asset owner responsible for each device, and the department associated with each device. The inventory should include every system that has an Internet protocol (IP) address on the network, including but not limited to desktops, laptops, servers, network equipment (routers, switches, firewalls, etc.), printers, storage area networks, Voice Over--IP telephones, multi--homed addresses, virtual addresses, etc. The asset inventory created must also include data on whether the device is a portable and/or personal device. Devices such as mobile phones, tablets, laptops, and other portable electronic devices that store or process data must be identified, regardless of whether they are attached to the organization's network.	This control generally seeks to prevent threats from launching successful attacks. The inherent vulnerability here is that virtual machines may have false MAC addresses, which may be exploited by an attacker for an attack. It may be also difficult to keep such an inventory up-to-date therefore affecting its utility and the amount employees rely on it. An attack may attempt to exploit the fact that maintaining such complete listings is very difficult, especially given the large amount of devices (personal and otherwise) that employees have.
CSC 1.5	Deploy network level authentication via 802.1x to limit and control which devices can be connected to the network. The 802.1x must be tied into the inventory data to determine authorized versus unauthorized systems.	This control generally seeks to prevent threats from launching successful attacks. An inherent vulnerability is that validating MAC addresses implies the creation of a whitelist of authorized systems to connect to the network. This whitelist must be regularly updated or could, for instance, be allowing devices no longer authorised. An attacker could use a one-authorised device to gain access to the network (this is likely if the whitelists are not updated immediately).
CSC 1.6	Use client certificates to validate and authenticate systems prior to connecting to the private network.	This control generally seeks to prevent threats from launching successful attacks. The challenge and potential area of vulnerability here is that it requires all systems that want to connect to the network to have a client certificate installed. This may not be practical for certain devices or certain types of networks (e.g., guest networks). To add to the vulnerability above, an attacker could simply steal or gain access to a legitimate system, and then use it to connect to the network.
CSC 2: Inventory of Authorized and Unauthorized Software	**Actively manage (inventory, track, and correct) all software on the network so that only authorized software is installed and can execute, and that unauthorized and unmanaged software is found and prevented from installation or execution.**	**The inherent vulnerability and weakness to attack of this control is detailed below according to its specific sub-controls.**
CSC 2.1	Devise a list of authorized software and version that is required in the enterprise for each type of system, including servers, workstations, and laptops of various kinds and uses. This list should be monitored by file integrity checking tools to validate that the authorized software has not been modified.	This control generally seeks to prevent threats from launching successful attacks. An inherent vulnerability of this control is that if there are lapses in the monitoring of the file's integrity or in its updating, then it may miss unauthorised software/versions. An attacker might damage the effectiveness

		of this control by changing the lists, if they are not appropriately or regularly monitored. This would mean that they may be able to have malware executed on corporate machines.
CSC 2.2	Deploy application whitelisting that allows systems to run software only if it is included on the whitelist and prevents execution of all other software on the system. The whitelist may be very extensive (as is available from commercial whitelist vendors), so that users are not inconvenienced when using common software. Or, for some special-purpose systems (which require only a small number of programs to achieve their needed business functionality), the whitelist may be quite narrow.	This control generally seeks to prevent threats from launching successful attacks. It also, in some ways, protects attacks from spreading (e.g., if malware has been downloaded to a machine, it cannot be executed because the software is not whitelisted). An inherent vulnerability is the reliance on extensive, updated and 'complete' whitelists. An attacker could try to circumvent this control by attacking the whitelist maintained by organisations. Though it does not seem widely possible, it would be concerning if attackers found a way to impersonate legitimate software (e.g., via copying signatures or attaching to them, such as macros in Microsoft Office).
CSC 2.3	Deploy software inventory tools throughout the organization covering each of the operating system types in use, including servers, workstations, and laptops. The software inventory system should track the version of the underlying operating system as well as the applications installed on it. The software inventory systems must be tied into the hardware asset inventory so all devices and associated software are tracked from a single location.	This control generally seeks to allow the detection of threats before they launch successful attacks. The potential vulnerability here is the dependence on inventory tools (which may themselves be not perfect) to identify and track all systems and applications.
CSC 2.4	Virtual machines and/or air-gapped systems should be used to isolate and run applications that are required for business operations but based on higher risk should not be installed within a networked environment.	This control generally seeks to prevent threats from launching successful attacks. A potential vulnerability that could be exploited here is human error, where the same user devices (e.g., USB keys) are used to access air-gapped systems and normal networked systems. An attacker may attempt to reach an air-gapped (or secured) network by relying on human error, e.g., free USB keys, or spear-phishing emails (with malicious payloads).
CSC 3: Secure Configurations for Hardware and Software on Mobile Device Laptops, Workstations, and Servers	**Establish, implement, and actively manage (track, report on, correct) the security configuration of laptops, servers, and workstations using a rigorous configuration management and change control process in order to prevent attackers from exploiting vulnerable services and settings.**	**The inherent vulnerability and weakness to attack of this control is detailed below according to its specific sub-controls.**
CSC 3.1	Establish standard secure configurations of operating systems and software applications. Standardized images should represent hardened versions of the underlying operating system and the applications installed on the system. These images should be validated and refreshed on a regular basis to update their security configuration	This control generally seeks to prevent threats from launching successful attacks. The potential vulnerability here is that organisations need to keep an updated secure configuration for each system. The timeframe between an updated image being publicly

	in light of recent vulnerabilities and attack vectors.	available and the organisation implementing the necessary changes is key in this regard. This is not always straightforward for organisations to do and therefore needs time and checking before changes are made. An attacker may try to attack a system in the time period where a vulnerability is published, but organisations have not yet updated their systems to address it. This would therefore particularly affect organisations that are slow to update.
CSC 3.2	Follow strict configuration management, building a secure image that is used to build all new systems that are deployed in the enterprise. Any existing system that becomes compromised should be re--imaged with the secure build. Regular updates or exceptions to this image should be integrated into the organization's change management processes. Images should be created for workstations, servers, and other system types used by the organization.	This control generally seeks to limit the spread of successful attacks. A potential inherent vulnerability of this control is that the use of personal devices may render the configuration management of new devices difficult. While very difficult to do, if an attacker were able to compromise a single core image, then their changes would be replicated to several computers. This would have a significant impact on an organisation.
CSC 3.3	Store the master images on securely configured servers, validated with integrity checking tools capable of continuous inspection, and change management to ensure that only authorized changes to the images are possible. Alternatively, these master images can be stored in offline machines, air- gapped from the production network, with images copied via secure media to move them between the image storage servers and the production network.	This control generally seeks to limit the spread of successful attacks and allow recovery from them. The inherent vulnerability here is the threat posed by insider threats that may want to attack these master images. Currently there seems to be no mechanisms directly mentioning this here. An attacker that is capable of accessing these master images or blocking access to them (whether physically or otherwise), could be a significant threat to the organisation.
CSC 3.4	Perform all remote administration of servers, workstation, network devices, and similar equipment over secure channels. Protocols such as telnet, VNC, RDP, or others that do not actively support strong encryption should only be used if they are performed over a secondary encryption channel, such as SSL, TLS or IPSEC.	This control generally seeks to prevent threats from launching successful attacks. The inherent vulnerability in this control is that there still may be instances where remote administration occurs over unsecure channels, be it linked to legacy systems or human mistakes. An attacker may attempt to exploit human mistakes or misunderstandings such that the control is not followed and connections made are not via a secure channel.

CSC 3.5	Use file integrity checking tools to ensure that critical system files (including sensitive system and application executables, libraries, and configurations) have not been altered. The reporting system should: have the ability to account for routine and expected changes; highlight and alert on unusual or unexpected alterations; show the history of configuration changes over time and identify who made the change (including the original logged-in account in the event of a user ID switch, such as with the su or sudo command). These integrity checks should identify suspicious system alterations such as: owner and permissions changes to files or directories; the use of alternate data streams which could be used to hide malicious activities; and the introduction of extra files into key system areas (which could indicate malicious payloads left by attackers or additional files inappropriately added during batch distribution processes).	This control generally seeks to detect threats that may lead to successful attacks. Inherently, the control depends on identifying changes in critical files and routine or expected changes. The success of the control will depend on the quality of the system tasked to monitor, analyse and alert of changes; People change permission settings for convenience or add/ delete files (even for critical systems) during their working duties, so this will also need to be considered. An attacker may attempt compromising the integrity checking tools or could aim to hijack the account of a legitimate user to make illegitimate system changes.
CSC 3.6	Implement and test an automated configuration monitoring system that verifies all remotely testable secure configuration elements, and alerts when unauthorized changes occur. This includes detecting new listening ports, new administrative users, changes to group and local policy objects (where applicable), and new services running on a system. Whenever possible use tools compliant with the Security Content Automation Protocol (SCAP) in order to streamline reporting and integration.	This control generally seeks to detect threats that may lead to successful attacks. The control depends on automating configuration monitoring. The success of the control will depend on the quality of the system tasked to monitor and analyse changes which may occur in configuration settings but also on how well policies are mapped into monitoring practices. An organisation that does not know its baseline for used ports or services may be an easy victim for a capable attacker.
CSC 3.7	Deploy system configuration management tools, such as Active Directory Group Policy Objects for Microsoft Windows systems or Puppet for UNIX systems that will automatically enforce and redeploy configuration settings to systems at regularly scheduled intervals. They should be capable of triggering redeployment of configuration settings on a scheduled, manual, or event--driven basis.	This control generally seeks to prevent and to some extent limit the spread of successful attacks. The control depends on automating configuration monitoring. The success of the control will depend on the quality of the system tasked to monitor and analyse changes which may occur in configuration settings but also on how well policies are mapped into monitoring practices. An organisation that does not know its baseline for used ports or services may be an easy victim for a capable attacker.

CSC 4: Continuous Vulnerability Assessment and Remediation	**Continuously acquire, assess, and take action on new information in order to identify vulnerabilities, remediate, and minimize the window of opportunity for attackers.**	**The inherent vulnerability and weakness to attack of this control is detailed below according to its specific sub-controls.**
CSC 4.1	Run automated vulnerability scanning tools against all systems on the network on a weekly or more frequent basis and deliver prioritized lists of the most critical vulnerabilities to each responsible system administrator along with risk scores that compare the effectiveness of system administrators and departments in reducing risk. Use a SCAP-validated vulnerability scanner that looks for both code-based vulnerabilities (such as those described by Common Vulnerabilities and Exposures entries) and configuration--based vulnerabilities (as enumerated by the Common Configuration Enumeration Project).	This control generally seeks to prevent threats from launching successful attacks. The time that elapses between the public announcement of a vulnerability, the release of a patch for the system, the occurrence of the vulnerability scan and the time required to implement the patch is crucial; definitions of discovered vulnerabilities across multiple platform classification schemes are not necessarily standardised An attacker that is capable of obtaining access to zero-day attacks could be a significant threat to the organisation.
CSC 4.2	Correlate event logs with information from vulnerability scans to fulfil two goals. First, personnel should verify that the activity of the regular vulnerability scanning tools is itself logged. Second, personnel should be able to correlate attack detection events with prior vulnerability scanning results to determine whether the given exploit was used against a target known to be vulnerable.	This control generally seeks to prevent threats from launching successful attacks. It may also provide information to controls detecting attacks. This control depends on the experience of personnel to identify exploits of well-known vulnerabilities and correlate these with attack detection events. An attacker may attempt to exploit human mistakes in the logging process. An alternative could be to exploit the correlation algorithms and execute attacks which will create logs that will be considered part of the network's normal activity.
CSC 4.3	Perform vulnerability scanning in authenticated mode either with agents running locally on each end system to analyse the security configuration or with remote scanners that are given administrative rights on the system being tested. Use a dedicated account for authenticated vulnerability scans, which should not be used for any other administrative activities and should be tied to specific machines at specific IP addresses. Ensure that only authorized employees have access to the vulnerability management user interface and that roles are applied to each user.	This control generally seeks to prevent threats from launching successful attacks. It also aims at limiting lateral movement from the attackers. The inherent vulnerability in this control lies in the presence of insider threat. Organisations should have in place policies to deter insider attacks. An attacker may attempt to exploit human mistakes and obtain administration credentials with phishing attacks etc. An insider may obtain access as well and execute attacks.

CSC 4.4	Subscribe to vulnerability intelligence services in order to stay aware of emerging exposures, and use the information gained from this subscription to update the organization's vulnerability scanning activities on at least a monthly basis. Alternatively, ensure that the vulnerability scanning tools you use are regularly updated with all relevant important security vulnerabilities.	This control generally seeks to prevent threats from launching successful attacks. Organisations depend on third parties to discover vulnerabilities; The challenge is in ensuring that the vulnerability descriptions organisations receive are both human and machine readable, and that the machine readable format is integrated to systems organisation have in place to act upon vulnerabilities (Source: Tripwire). If an attacker gains knowledge of which vulnerability intelligence services the organisation subscribes to, there is the potential to execute attacks that are not covered in these services. Alternatively, an attacker may obtain knowledge of unpatched systems and execute attacks on these parts of the network.
CSC 4.5	Deploy automated patch management tools and software update tools for operating system and software/applications on all systems for which such tools are available and safe. Patches should be applied to all systems, even systems that are properly air gapped.	This control generally seeks to prevent threats from launching successful attacks. Patches for in-house and/or custom applications/integrations may be challenging to automate (Source: Tripwire); Organisations rely on third parties providing patches before the vulnerabilities are well-known and being exploited; Some systems may not detect or install patches correctly due to an error by the third party providing the patch or the administrator (Source: SANS 20) In cases where are delays in the patch management or failures which are not reported, an attacker may obtain knowledge of unpatched systems and execute attacks on these parts of the network.
CSC 4.6	Monitor logs associated with any scanning activity and associated administrator accounts to ensure that this activity is limited to the timeframes of legitimate scans.	This control generally seeks detect malicious behaviour in the network in terms of reconnaissance. The inherent vulnerability in this control lies in the presence of insider threat. Organisations should have in place policies to deter insider attacks. An attacker may mask reconnaissance activities as scanning activities if the analysis of the monitoring is not effective.
CSC 4.7	Compare the results from back-to-back vulnerability scans to verify that vulnerabilities were addressed, either by patching, implementing a compensating control, or documenting and accepting a reasonable business risk. Such acceptance of business risks for existing vulnerabilities should be periodically reviewed to determine if newer compensating controls or subsequent patches can address vulnerabilities that were previously accepted, or if conditions have changed, increasing the risk.	This control generally seeks to prevent threats from launching successful attacks. It also aims at limiting lateral movement from the attackers as well as providing additional information on detecting attacks. This control implies a risk assessment in place mapping critical assets from a business perspective to network and technical assets. The success of the control depends on how well organisations understand how the infrastructure supports services and core business capabilities the organisation; Time to

		install a patch depends on third parties making available such a patch or whether a system is bespoke/ in-house; Experience of personnel responsible to judge if the time is within the accepted risk or not is crucial. In cases where are delays in the patch management or failures which are not reported, an attacker may obtain knowledge of unpatched systems and execute attacks on these parts of the network.
CSC 4.8	Establish a process to risk-rate vulnerabilities based on the exploitability and potential impact of the vulnerability, and segmented by appropriate groups of assets (example, DMZ servers, internal network servers, desktops, laptops). Apply patches for the riskiest vulnerabilities first. A phased rollout can be used to minimize the impact to the organization. Establish expected patching timelines based on the risk rating level.	This control generally seeks to prevent threats from launching successful attacks. It also aims at limiting lateral movement from the attackers as well as providing additional information on detecting attacks. This control implies a risk assessment in place mapping critical assets from a business perspective to network and technical assets. The success of the control depends on how well organisations understand how the infrastructure supports services and core business capabilities the organisation; Experience of personnel responsible to judge if the time is within the accepted risk or not is crucial. In cases where are delays in the patch management or failures which are not reported, an attacker may obtain knowledge of unpatched systems and execute attacks on these parts of the network.
CSC 5: Controlled Use of Administrative Privileges	**The processes and tools used to track/control/prevent/correct the use, assignment, and configuration of administrative privileges on computers, networks, and applications.**	**The inherent vulnerability and weakness to attack of this control is detailed below according to its specific sub-controls.**
CSC 5.1	Minimize administrative privileges and only use administrative accounts when they are required. Implement focused auditing on the use of administrative privileged functions and monitor for anomalous behaviour.	This control generally seeks to prevent successful attacks, and to assist in their detection. The success of the control depends on the experience of the personnel, the likelihood of human (administrator) errors being made (e.g., logging into systems as administrator when not necessary), and the efficiency of the monitoring/detection system. An attacker that manages to compromise or gain access to an administrative account may be able to launch an attack on the organisation before being detected.
CSC 5.2	Use automated tools to inventory all administrative accounts and validate that each person with administrative privileges on desktops, laptops, and servers is authorized by a senior executive.	This control generally seeks to prevent successful attacks. The success of the control depends on the automated tools (likely third-party tools) and how effective they are at inventorying all of the administrator accounts.

		An attacker could circumvent this control by gaining access to the administrative accounts, such as accounts of legitimate users or accounts setup for now defunct purposes (e.g., development or pen-testing).
CSC 5.3	Before deploying any new devices in a networked environment, change all default passwords for applications, operating systems, routers, firewalls, wireless access points, and other systems to have values consistent with administration--level accounts.	This control generally seeks to prevent successful attacks. The potential vulnerability here is that this requires the organisation to have a perfect understanding of all the devices on their network. Furthermore, users tend to change passwords and use memorable words therefore even though default passwords may not exist, there still may be weak passwords used. An attacker may attempt to exploit the use of weak passwords, even though they may not be the default ones.
CSC 5.4	Configure systems to issue a log entry and alert when an account is added to or removed from a domain administrators' group, or when a new local administrator account is added on a system.	This control generally may be used to help detect attacks. The inherent vulnerability here is that there needs to be an immediate follow-up of these log entries and alerts, or attacks could quickly follow and compromise corporate systems. An attacker may seek to disrupt this control by first changing up system configurations from posting alerts. Another option for the attacker is to ensure that a significant amount of alerts are launched thereby indirectly suggesting to the security operator that the alert system is not functioning properly.
CSC 5.5	Configure systems to issue a log entry and alert on any unsuccessful login to an administrative account.	This control generally may be used to help detect potential attacks. The inherent vulnerability is that there may be instances where administrators forget or fail to enter correct passwords, and therefore generate alerts. These alerts may change the baseline for such alerts, which may actually result in it being 'normal' for 1-2 alerts to be made. In instances where these alerts are legitimate (e.g., it is truly an attacker), they may therefore be missed as they are within the baseline. An attacker may be particularly careful in how many times they attempt to login to administrative accounts, such that they remain below the 'normal' baseline.
CSC 5.6	Use multi-factor authentication for all administrative access, including domain administrative access. Multi-factor authentication can include a variety of techniques, to include the use of smart cards, certificates, One Time Password (OTP) tokens, biometrics, or other similar authentication methods.	This control generally seeks to prevent successful attacks. Some multi-factor tools are less secure than others, and this will need to be factored in by organisations selecting which controls to implement. Moreover, some users tend to skip two-factor authentication for convenience if an alternative is provided.

		An attacker capable of gaining access to the multiple factors would be able to access corporate systems. Moreover, if the attacker could block access to one of the tokens, this would block the users from gaining access themselves (thus, a denial-of-service type attack).
CSC 5.7	Where multi-factor authentication is not supported, user accounts shall be required to use long passwords on the system (longer than 14 characters).	This control generally seeks to prevent successful attacks. The vulnerability here is that long passwords are not necessarily secure ones. This is especially true given that users need to remember the password so may therefore opt for a word (or string of words) that is easy to guess. An attacker could exploit the fact that users will pick memorable passwords, and attempt to circumvent the control by guessing the password based on the user (e.g., name of family members, favourite teams etc.).
CSC 5.8	Administrators should be required to access a system using a fully logged and non-administrative account. Then, once logged on to the machine without administrative privileges, the administrator should transition to administrative privileges using tools such as Sudo on Linux/UNIX, RunAs on Windows, and other similar facilities for other types of systems.	This control generally seeks to prevent successful attacks. The vulnerability in this context is that it is difficult to mandate that administrators to follow these practices and some may even tend to skip multiple access to systems for convenience. One potential area where an attacker may be able to gain administrative access – especially on Windows – is if they have installed a key-logger. Therefore, even though mechanisms have been put in place to reduce administrative privileges, this attack will be possible.
CSC 5.9	Administrators shall use a dedicated machine for all administrative tasks or tasks requiring elevated access. This machine shall be isolated from the organization's primary network and not be allowed Internet access. This machine shall not be used for reading email, composing documents, or surfing the Internet.	This control generally seeks to prevent successful attacks. A potential vulnerability here is that users tend to access systems from multiple devices for convenience. In practice therefore, maintaining a fully isolated machine may be challenging. There is also the fact that this device may become infected if USB keys are placed into it (e.g., to copy needed software). An attacker may attempt to reach an isolated machine by relying on human error, e.g., free USB keys with malicious payloads.
CSC 6: Maintenance, Monitoring, and Analysis of Audit Logs	**Collect, manage, and analyse audit logs of events that could help detect, understand, or recover from an attack.**	**The inherent vulnerability and weakness to attack of this control is detailed below according to its specific sub-controls.**
CSC 6.1	Include at least two synchronized time sources from which all servers and network equipment retrieve time information on a regular basis so that	This control generally seeks to prevent threats from launching successful attacks. It also aims at limiting lateral movement from the

	timestamps in logs are consistent.	attackers as well as providing additional information on detecting attacks. There is an inherent vulnerability due to the fact that computer clocks are prone to drifting. Things may become complicated when dealing with time stamps from different hour-zone locations.
CSC 6.2	Validate audit log settings for each hardware device and the software installed on it, ensuring that logs include a date, timestamp, source addresses, destination addresses, and various other useful elements of each packet and/or transaction. Systems should record logs in a standardized format such as syslog entries or those outlined by the Common Event Expression initiative. If systems cannot generate logs in a standardized format, log normalization tools can be deployed to convert logs into such a format.	This control generally seeks to prevent threats from launching successful attacks. It also aims at limiting lateral movement from the attackers as well as providing additional information on detecting attacks. It may be difficult to attain a standardized format for logs; Log normalization tools may have their own inconsistencies and may introduce another hurdle when monitoring or attempting to detect an incident. Attackers may attempt to obtain access and either change the logging information or force systems to stop logging events.
CSC 6.3	Ensure that all systems that store logs have adequate storage space for the logs generated on a regular basis, so that log files will not fill up between log rotation intervals. The logs must be archived and digitally signed on a periodic basis.	This control generally seeks to prevent threats from launching successful attacks. It also aims at limiting lateral movement from the attackers as well as providing additional information on detecting attacks. An inherent vulnerability lies in the way logs are stored. There is a possibility that during the storage process logs may become corrupted or damaged depending on how they are archived. Attackers may attempt to obtain access and either change the logging information or force systems to stop logging events.
CSC 6.4	Have security personnel and/or system administrators run biweekly reports that identify anomalies in logs. They should then actively review the anomalies, documenting their findings.	This control generally seeks to detect threats and lateral movement from the attackers. The inherent vulnerability in many detection systems lies in the number of false positive alerts they generate. Managing the number of false positive and negative alerts may become an overwhelming task. Attackers may choose to create noise to change the norm in what constitutes an anomaly in the logs. They may also create too many alerts for security personnel to review in a timely manner, masking the real attack.
CSC 6.5	Configure network boundary devices, including firewalls, network-based IPS, and inbound and outbound proxies, to verbosely log all traffic (both allowed and blocked) arriving at the device.	This control generally seeks to prevent threats from launching successful attacks. It also aims at limiting lateral movement from the attackers as well as providing additional information on detecting attacks.

CSC 6.6	Deploy a SIEM (Security Information and Event Management) or log analytic tools for log aggregation and consolidation from multiple machines and for log correlation and analysis. Using the SIEM tool, system administrators and security personnel should devise profiles of common events from given systems so that they can tune detection to focus on unusual activity, avoid false positives, more rapidly identify anomalies, and prevent overwhelming analysts with insignificant alerts.	This control generally seeks to detect threats and lateral movement from the attackers. There is not a standardised format for logs. Different systems will provide different logging events. There is an inherent vulnerability when software and applications provide complex logs which are challenging to process. Also SIEM tools may provide a big number of false positive events if not configured properly. Attackers may choose to create noise to change the norm in what constitutes an anomaly in the logs. They may also create too many alerts for security personnel to review in a timely manner, masking the real attack.
CSC 7: Email and Web Browser Protections	**Minimize the attack surface and the opportunities for attackers to manipulate human behaviour though their interaction with web browsers and email systems.**	**The inherent vulnerability and weakness to attack of this control is detailed below according to its specific sub-controls.**
CSC 7.1	Ensure that only fully supported web browsers and email clients are allowed to execute in the organization, ideally only using the latest version of the browsers provided by the vendor in order to take advantage of the latest security functions and fixes.	This control generally seeks to prevent successful attacks. The inherent vulnerabilities in this control are based on the fact that email clients and browsers may also come with (unknown) vulnerabilities; There may be a delay in the organisation's software update cycle which results in updates not being immediately applied (e.g., updates may need to be checked for compatibility with local systems before being implemented). An attacker may seek to exploit the control by either (a) finding a new vulnerability in the control and attacking the organisation that way, or (b) conducting an attack immediately as the vulnerability has been released – this takes advantage of the fact that it usually takes organisations some time before they implement latest updates/patches.
CSC 7.2	Uninstall or disable any unnecessary or unauthorized browser or email client plugins or add-on applications. Each plugin shall utilize application / URL whitelisting and only allow the use of the application for pre-approved domains.	This control generally seeks to prevent successful attacks. The inherent vulnerability is that it relies heavily on organisations having a complete understanding of their systems and software - which may not be the case in many organisations especially considering bring-your-own-device paradigms. An attacker may seek to exploit the fact that employees use personal devices (which are not subject to stringent security requirements as mentioned) to access corporate systems and services. This may therefore provide them with a new vector to attack the organisation (i.e., via a user and personal

		devices).
CSC 7.3	Limit the use of unnecessary scripting languages in all web browsers and email clients. This includes the use of languages such as ActiveX and JavaScript on systems where it is unnecessary to support such capabilities.	This control generally seeks to prevent successful attacks. The vulnerability that may still exist here is due to the fact that if organisations fail to block users in changing web browser and email client settings, they make modifications which put the organisation at risk. An attacker may seek to trick users into changing settings (e.g., via an ActiveX or Java game) and then launch an attack – this would directly impact the effectiveness of this control.
CSC 7.4	Log all URL requests from each of the organization's systems, whether onsite or a mobile device, in order to identify potentially malicious activity and assist incident handlers with identifying potentially compromised systems.	This control generally seeks to detect successful attacks. The potential concern for organisations here is that logging all URL requests is ideal but will generate a significant amount of log data. Unless the organisation has people, tools and systems able to properly analyse the extent of these logs in order to detect potentially malicious activity, the utility of logs will be limited. If an attacker was able to compromise an organisation's systems, they could choose to generate a large variety of legitimate traffic and hide a request or (especially redirect request) within it. This would therefore rely on the organisation picking up and acting on the request in later logs. Depending on how quickly they respond to the request, the attacker could launch an attack.
CSC 7.5	Deploy two separate browser configurations to each system. One configuration should disable the use of all plugins, unnecessary scripting languages, and generally be configured with limited functionality and be used for general web browsing. The other configuration shall allow for more browser functionality but should only be used to access specific websites that require the use of such functionality.	This control generally seeks to prevent successful attacks. From a vulnerability perspective, it is not clear how relaxed the second browser configuration should be - this is a challenge organisations and also is dependent on each specific organisation. Regardless, they will need to have a comprehensive understanding of risks before they allow/deny configurations. Depending on how straightforward it is to switch between configurations, if an attacker is able to move from the more secure to less secure configuration, they may be able to conduct some form of an attack (e.g., downloading software from Google Drive or Dropbox – both recognised sites).

CSC 7.6	The organization shall maintain and enforce network based URL filters that limit a system's ability to connect to websites not approved by the organization. The organization shall subscribe to URL categorization services to ensure that they are up-to-date with the most recent website category definitions available. Uncategorized sites shall be blocked by default. This filtering shall be enforced for each of the organization's systems, whether they are physically at an organization's facilities or not.	This control generally seeks to prevent successful attacks. A key vulnerability here is that approved websites may be compromised therefore URL filtering will not be helpful in those cases. Moreover, this control relies on either a complete list of approved websites or a complete list of blacklisted websites - each is difficult to find and maintain An attacker may use an approved site as a platform to conduct an attack or launch/download a malicious payload.
CSC 7.7	To lower the chance of spoofed email messages, implement the Sender Policy Framework (SPF) by deploying SPF records in DNS and enabling receiver-side verification in mail servers.	This control generally seeks to prevent successful attacks. This control will be susceptible to any vulnerabilities that exist within the Sender Policy Framework (SPF).
CSC 7.8	Scan and block all email attachments entering the organization's email gateway if they contain malicious code or file types that are unnecessary for the organization's business. This scanning should be done before the email is placed in the user's inbox. This includes email content filtering and web content filtering.	This control generally seeks to prevent successful attacks. The inherent vulnerability in this context is that malicious files are often disguised or attached to files appearing to be legitimate (e.g., PDFs, DOCs), therefore these may not be filtered. An attacker may look to exploit the fact that most businesses use PDFs, DOCs etc. and conduct their attacks in that way, e.g., embedding a malicious payload set to execute only when the file is run.
CSC 8: Malware Defences	**Control the installation, spread, and execution of malicious code at multiple points in the enterprise, while optimizing the use of automation to enable rapid updating of defence, data gathering, and corrective action.**	**The inherent vulnerability and weakness to attack of this control is detailed below according to its specific sub-controls.**
CSC 8.1	Employ automated tools to continuously monitor workstations, servers, and mobile devices with anti-virus, anti-spyware, personal firewalls, and host-based IPS functionality. All malware detection events should be sent to enterprise anti-malware administration tools and event log servers.	This control generally seeks to prevent threats from launching successful attacks. It also aims at limiting lateral movement from the attackers as well as providing additional information on detecting attacks. These controls rely heavily on existing rule sets and known vulnerabilities and attacks. This reality does sometimes limit their ability to address new attacks; There is also the assumption that users will not attempt to circumvent these tools for malicious or benign purposes; Challenges with false positives.
CSC 8.2	Employ anti-malware software that offers a centralized infrastructure that compiles information on file reputations or have administrators manually push updates to all machines. After applying an update, automated systems should verify that each system has received its signature update.	This control generally seeks to prevent threats from launching successful attacks. It also aims at limiting lateral movement from the attackers as well as providing additional information on detecting attacks.

CSC 8.3	Limit use of external devices to those with an approved, documented business need. Monitor for use and attempted use of external devices. Configure laptops, workstations, and servers so that they will not auto-run content from removable media, like USB tokens (i.e., "thumb drives"), USB hard drives, CDs/DVDs, FireWire devices, external serial advanced technology attachment devices, and mounted network shares. Configure systems so that they automatically conduct an anti-malware scan of removable media when inserted.	This control generally seeks to prevent threats from launching successful attacks. There are inherent vulnerabilities due to the fact that users may still insert external devices and run malware (knowingly or unknowingly); Additionally, anti-malware tools are often dependent on lists of known vulnerabilities and attacks. Attackers may try to exploit users with phishing attacks and trick them to change the configuration controls.
CSC 8.4	Enable anti-exploitation features such as Data Execution Prevention (DEP), Address Space Layout Randomization (ASLR), virtualization/containerization, etc. For increased protection, deploy capabilities such as Enhanced Mitigation Experience Toolkit (EMET) that can be configured to apply these protections to a broader set of applications and executables.	This control generally seeks to prevent threats from launching successful attacks. It also aims at limiting lateral movement from the attackers. DEP without ASLR is not robust enough to prevent arbitrary code execution in most cases, and for ASLR, the absence of DEP can allow an attacker to use heap spraying to place code at a predictable location in the address space. They are best used together but even then their combined effectiveness is heavily dominated by the effectiveness of ASLR. (https://blogs.technet.microsoft.com/srd/2010/12/08/on-the-effectiveness-of-dep-and-aslr/)
CSC 8.5	Use network-based anti-malware tools to identify executables in all network traffic and use techniques other than signature-based detection to identify and filter out malicious content before it arrives at the endpoint.	This control generally seeks to detect threats already present in the network. The sophistication of the specific control is crucial in being able to accurately detect executables and filter them out as they pass over the corporate network. It is also worth mentioning that this additional scanning could have an impact on network performance and business services. Attackers may obtain knowledge of the version of the anti-malware running on the network and execute attacks which are not flagged as malicious in those security vendors.
CSC 8.6	Enable domain name system (DNS) query logging to detect hostname lookup for known malicious C2 domains.	This control generally seeks to prevent threats from launching successful attacks. It also aims at detecting the presence of threats in the network. This control requires a comprehensive and up-to-date list of all malicious C2 domains in order to function properly. Inherent vulnerabilities lie on how up-to-date the list is. Attackers may obtain knowledge of the C2 domains which are included in the black list and execute attacks which are not included in this list.

CSC 9: Limitation and Control of Network Ports, Protocols, and Services	Manage (track/control/correct) the ongoing operational use of ports, protocols, and services on networked devices in order to minimize windows of vulnerability available to attackers.	The inherent vulnerability and weakness to attack of this control is detailed below according to its specific sub-controls.
CSC 9.1	Ensure that only ports, protocols, and services with validated business needs are running on each system.	This control generally seeks to prevent successful attacks, and potentially to limit them (e.g., limit the use of unauthorised ports, protocols, services). The vulnerability in this case that is unless this control is centrally administered, it could be a complicated task and one which might lend to machines being overlooked or missed. There is also the challenge that depending on the organisation and the tools selected to implement this control, accurate configuration of the network ports, protocols and services might be a complex task. An attacker may try to circumvent this control by either (a) ensuring that unauthorised actions (e.g., remote access or exfiltrating data) is conducted via known ports, protocols or services, or (b) searching for instances where machines have been overlooked and it is possible to run or connect to unauthorised ports.
CSC 9.2	Apply host-based firewalls or port filtering tools on end systems, with a default-deny rule that drops all traffic except those services and ports that are explicitly allowed.	This control generally seeks to prevent successful attacks (though firewalls may also be used for detection). An attacker that targets services and ports that are explicitly allowed may not necessarily be prevented. A potential weakness is also that there is a reliance on the effectiveness of selected firewalls and port filtering tools.
CSC 9.3	Perform automated port scans on a regular basis against all key servers and compare to a known effective baseline. If a change that is not listed on the organization's approved baseline is discovered, an alert should be generated and reviewed.	This control generally seeks to detect potential attacks or avenues of attacks. The inherent vulnerability with this control is that there needs to be a clear known effective baseline (discovering this baseline could be a difficult task in itself), and also updates to this baseline must be communicated and implemented continuously for this control to be effective. Moreover, regular port scans against production systems could result in disruption of services. The opportunity for an attacker is the case where the effective baseline is not clearly known and therefore the attacker may be using open ports on corporate systems (e.g., to connect remotely or access data).
CSC 9.4	Verify any server that is visible from the Internet or an untrusted network, and if it is not required for business purposes, move it to an internal VLAN and give it a private address.	This control generally seeks to prevent against potential attacks. There were no inherent vulnerabilities found.
CSC 9.5	Operate critical services on separate physical or logical host machines, such as DNS, file, mail, web, and database servers.	This control generally seeks to prevent against potential attacks.

		There were no inherent vulnerabilities found.
CSC 9.6	Place application firewalls in front of any critical servers to verify and validate the traffic going to the server. Any unauthorized services or traffic should be blocked and an alert generated.	This control generally seeks to prevent against potential attacks. A potential vulnerability here is that an attack that targets services that are explicitly allowed may not necessarily be prevented. Also, this control does not directly consider the use of firewalls to examine data being transferred from servers (arguably this may be covered by another CSC control however). To render this control ineffective, an attacker could target authorised services with malicious payloads. This may eventually reach users or may actually open applications themselves to compromise.
CSC 10: Data Recovery Capability	**The processes and tools used to properly back up critical information with a proven methodology for timely recovery of it.**	**The inherent vulnerability and weakness to attack of this control is detailed below according to its specific sub-controls.**
CSC 10.1	Ensure that each system is automatically backed up on at least a weekly basis, and more often for systems storing sensitive information. To help ensure the ability to rapidly restore a system from backup, the operating system, application software, and data on a machine should each be included in the overall backup procedure. These three components of a system do not have to be included in the same backup file or use the same backup software. There should be multiple backups over time, so that in the event of malware infection, restoration can be from a version that is believed to predate the original infection. All backup policies should be compliant with any regulatory or official requirements.	This control aims at mitigating harm from attacks and is not focusing on preventing or detecting threats. Depending on the time period between the last backup and the incident, some valuable data/software may still be lost; There may be situations where several back-ups contain undetected malware. Attackers may attempt to exfiltrate data from the back-up systems, instead of attacking the systems used on a daily basis. Attackers may also attempt to encrypt the data with attacks such as ransomware.
CSC 10.2	Test data on backup media on a regular basis by performing a data restoration process to ensure that the backup is properly working.	This control aims at mitigating harm from attacks and is not focusing on preventing or detecting threats. Backups may be tested in part, and not consider unique cases where several parts of the larger system are infected and need to be restored. Attackers may attempt to exfiltrate data from the back-up systems, instead of attacking the systems used on a daily basis. Attackers may also attempt to encrypt the data with attacks such as ransomware.

CSC 10.3	Ensure that backups are properly protected via physical security or encryption when they are stored, as well as when they are moved across the network. This includes remote backups and cloud services.	This control generally seeks to prevent threats from launching successful attacks on data stored on back-up systems. Inherent vulnerabilities may lie in weak protection mechanisms for backups (i.e. poor encryption protocols). These weak mechanism may introduce other risks and points of failure (i.e., theft / corruption of data) Attackers may attempt to exfiltrate data from the back-up systems, instead of attacking the systems used on a daily basis. Attackers may also attempt to encrypt the data with attacks such as ransomware.
CSC 10.4	Ensure that key systems have at least one backup destination that is not continuously addressable through operating system calls. This will mitigate the risk of attacks like CryptoLocker which seek to encrypt or damage data on all addressable data shares, including backup destinations.	This control generally seeks to prevent threats from launching successful attacks on data stored on back-up systems.

Appendix 4: Focus Group and Interview Questionnaire

The focus groups and interviewees will be asked to answer the questions **in their opinion**.

1. Assets
   a. What assets within organisations might be harmed by cyber events?
   b. What makes an asset critical in your view?
   c. What process do you go through to identify critical assets?
   d. How well are organisations aware of the interactions (or relationships, or dependencies) between assets?

2. Harm / loss
   a. What are the types of losses and harms that could result from cyber-attacks?
   b. [Show our categorisation and list of losses / harms] Are there any losses or harms that are not covered in this listing? Why would you add them?
   c. How would you determine the degree of loss or harm which may occur as a result of a cyber-attack on their assets?
   d. Any tools currently used to measure such losses?
   e. To what extent do organisations consider and model aggregate (i.e., secondary, tertiary, etc.) losses or harms that may occur as a result of a cyber-attack?

3. Controls
   a. How would you make decisions on which risk controls to apply to assets? What are some of the key factors considered, or approaches taken?
   b. How would you decide what security standards to adopt (e.g., ISO vs. SANS20 vs. Cyber Essentials)?
   c. To what extent are connected assets and aggregate (i.e., secondary, tertiary, etc.) harms considered in control selection?
   d. As a cyber insurer, which controls / control sets do you believe are most important or which ones do you recommend? Why? [**Question only for cyber insurers i.e., not asked in mixed cohort focus group**.]
   e. What would trigger a decision to review your position on the answer to 3f? [**Question only for cyber insurers i.e., not asked in mixed cohort focus group**.]

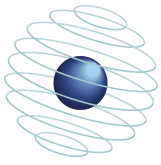
4. Control effectiveness
   a. How would you reason about (determine) the effectiveness of a risk control?
   b. What are the types of ways in which organisations can measure a control's effectiveness?
   c. What are some of the types of data that would need to be collected to measure control effectiveness?
   d. Residual risk is the risk remaining after a risk mitigation measure has been applied – do you agree with this description?
   e. Do you seek to measure or determine the levels of residual risk? If so, how?

5. Control dependencies
   a. Do you regard and treat risk controls as dependent upon each other?
   b. To what extent do organisations have a good understanding about the interdependencies between risk controls?
   c. How would you go about identifying and understanding the interdependencies between risk controls?
   d. Are there are any interdependencies that you consider risky and monitor accordingly?
   e. How does the effectiveness of controls impact the interdependencies between controls?

6. Cyber-VaR
    a. Are you familiar with the term Cyber Value-at-Risk? What does it mean to you?
    b. How would organisations attempt to reason about (determine) Cyber Value-at-Risk for an asset or set of assets?
    c. How would you attempt to reason about (determine) Cyber Value-at-Risk for a set of assets?
    d. Traditionally, to determine the level of risk to an asset, an organisation needs to consider the likelihood of the asset being attacked. How would an organisation determine this likelihood?
    e. [Present our approach to reason about Cyber-VaR; from assets to harm to VaR] What are your thoughts on this approach? How might it be further enhanced?

7. Cyber-insurance
    a. What are your thoughts on cyber-insurance, what part does it play in businesses today?
    b. What might be the driving factors which cause companies to purchase cyber-insurance?
    c. What might be the driving factors which cause companies not to purchase cyber-insurance?
    d. How does cyber-insurance compare to business interruption coverage or data loss insurance?
    e. How could cyber-insurance better position itself to be easier or more effective to purchase?

DEPARTMENT OF
COMPUTER
SCIENCE



CYBER
SECURITY
OXFORD