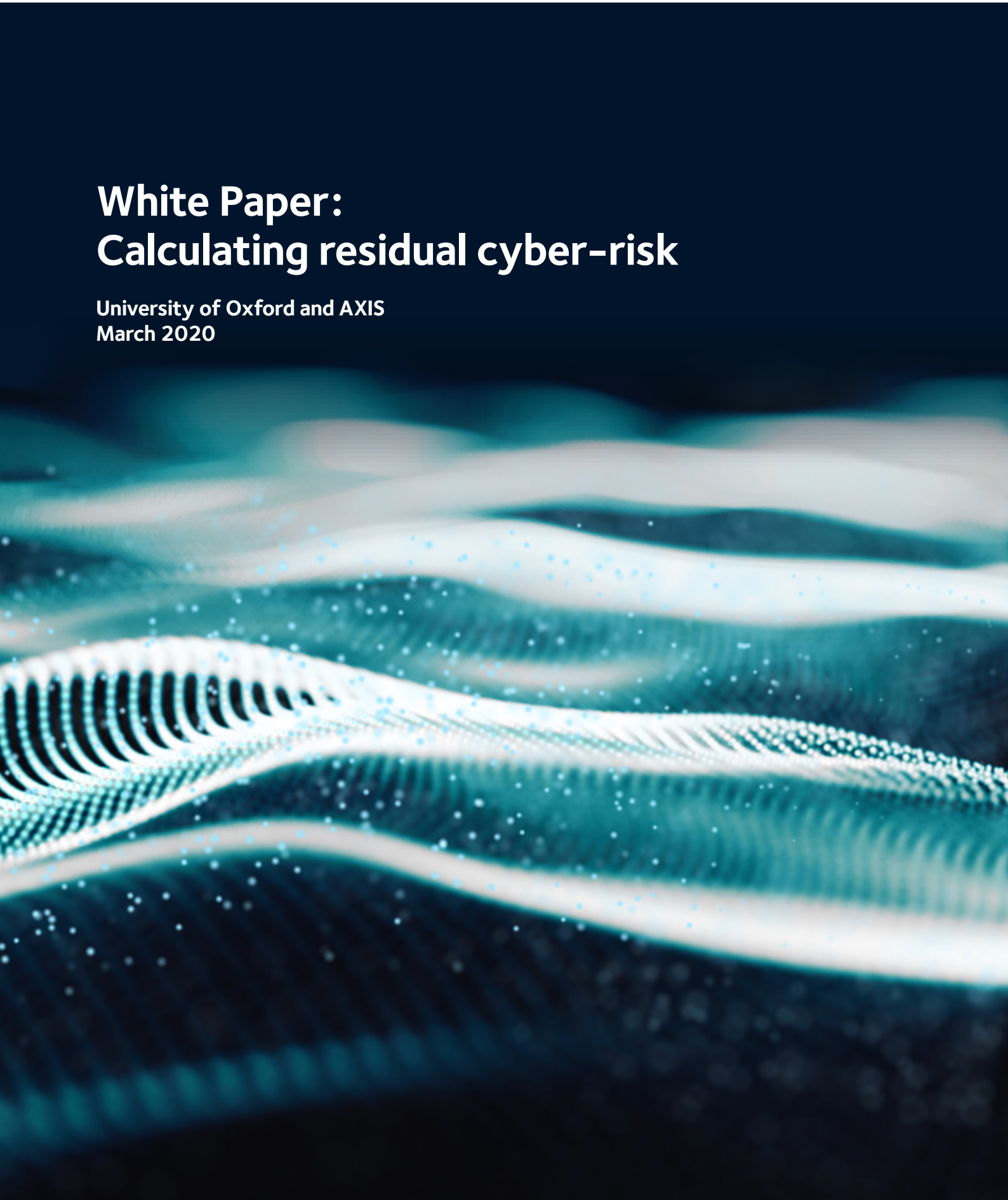
UNIVERSITY OF OXFORD

DEPARTMENT OF COMPUTER SCIENCE

AXIS

# White Paper: Calculating residual cyber-risk

University of Oxford and AXIS
March 2020



# Foreword

Dan Trueman,
AXIS

Senior executives and boards are taking more interest in their exposure to these threats. Consequently, the language and frames of reference used to articulate cyber-exposures have moved from technical issues to financial challenges. Business impact and potential financial loss resonate more keenly with senior stakeholders than risks measured by red, amber, green - they also align more naturally with the language of insurance and risk transfer.

Against this backdrop and a belief that there is a more commercial way to view cyber-risk, AXIS Capital has sponsored a research project, undertaken by Oxford University's Department of Computer Science. Under the control of Professor Sadie Creese and Professor Michael Goldsmith, the research examines alternative ways of quantifying cyber-risk and applying this understanding to risk transfer strategies.

The project began life looking at the quantification of cyber-risk using the Value-at-Risk (VaR) methodology. With its roots in the financial services industry, the model uses just three variables to quantify risk: potential losses, probability of losses and timeframes. Oxford proposes a new take on the model, introducing three more variables: control effectiveness, control dependencies and harm propagation.

This proposed model is important: it helps articulate cyber-risk in financial terms, supports commercial decision making, and helps create balance between protecting an organization and operating a business. It generates data that allows businesses to make decisions about their risk appetite, cybersecurity investments, and other risk mitigation and transfer strategies more confidently. It helps predict the potential losses arising from cyber-attacks and illustrates the pros and cons of their control mitigation strategies and configurations. It helps identify and quantify residual risk, helping calculate insurance limits and coverages.

The future will see the team investigating the predictive powers of the model in taking account of the most pressing threats currently faced by organisations, and ultimately exploring how we can reason about the potential for aggregated risk in cyberspace.

# Residual risk matters

It would be hard for any organisation to convince itself that it does not face some kind of risk from a cyber-attack. Exact numbers are debated, but it is generally agreed that we face a growing level of threat from increasingly able threat actors. The threat ecosystem is so broad that just by doing business online you will be exposed to prevalent cyber-criminality, even if you do not suffer a highly targeted attack.

Cyber-attacks are not new. The truth is that this has been happening for decades. But as we have become more dependent on digital technology and infrastructure, as we have created more value in digital assets and our use of the Internet, then the potential for harm resulting from cyber-attack has grown significantly. This harm takes many forms and can be monetised. Examples include: the onward sale of stolen assets; the charging of ransoms to maintain access to critical services and data; and the market advantage that can result from reputational damage or loss of secrets. Thus fuelling the ecosystem of actors and organisations ready to take their part in conducting such attacks, and the development of the tools and weaponry which enable them.

We can observe some of the most profitable and effective attacks trending in breach reports and the insurance losses being claimed. Of course, this is not the entire story. The development of new attack tools is a continuous activity. When new software vulnerabilities or zero-days that can be exploited to gain access to systems are discovered, they often initially defeat our perimeters. The issue for any particular organisation is whether the potential impact of the attack is large enough to justify being concerned, and if there is a high enough likelihood of it being realised to motivate taking action.

But deciding which response to take isn't simple. The knowledge and expertise for selecting risk controls exists in a limited supply of workforce. Cyber-harm is not always tangible to non-experts, and the need for taking action is often argued from worst case scenarios that many can feel are unlikely to arise (unless they have had personal experience of how bad things can be when an attack is successful). It can be hard to know whether action needs to be taken to address a threat that has not yet struck – since action costs, and therefore this may compromise investment that could have been made elsewhere.

Compounding these difficulties is the fact that there is little data on performance of different risk control methods – how do we know which to use and when? How do we know whether the return-on-investment is going to deliver the risk reduction that we need? We learnt in the '90s that even if we build our systems from components with verifiable high-levels of integrity (meaning that we have proven them to be free of certain types of attack surface), we can still be vulnerable to attack. This is because the way in which we operate our systems matters. Delivering cybersecurity is as much about how we orchestrate the controls, bringing them together to effect cybersecurity, as the components themselves.

We have to accept a level of risk – a residual risk that exists even after we have deployed our defences. Being cyber-secure means accepting insecurity, but attempting to manage it so we can be resilient; that, should the worst happen, it cannot be devastating. This makes cyber-risk a risk that must be managed, and to do that we must understand the nature of the risks that we face.

## In order for us to carry risk, a number of conditions need to hold.

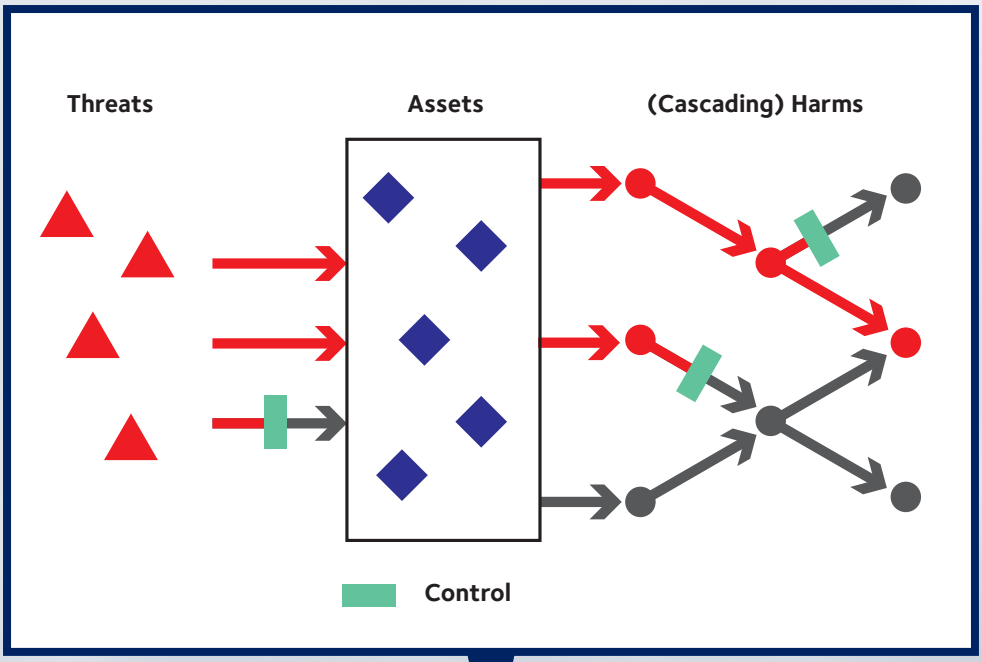
1. There must be a motivated and able threat.
2. There must be an attack surface which can be exploited.
3. A successful attack must result in some kind of loss or harm.

# Answering the question: "*Have we done enough?*" is challenging.

Since no organisation has unlimited budget, the reality of operating in the face of cyber-risk is that we have to try and focus our resources towards those risks with the capacity for greatest harm. We also need to know whether we are using our risk controls effectively, since if we are not then our residual risk may be far larger than we realise.

We believe that understanding the Cyber Value-at-Risk (CVaR) can provide a means to decide whether cybersecurity investments are adequate. Specifically, our CVaR model (Modelling the cyber gap (https://www.cs.ox.ac.uk/projects/ACVAR/Modelling_the_cyber_gap.pdf)) is the first to allow analysis that takes account of the controls we use and the way in which they can prevent harm from propagating when attacks successfully occur. Crucial to this, is understanding how effective our risk controls are and who in the organisation is best placed to oversee their use.

Our conceptual model for CVaR involves four different concepts: assets, cyber-harms (types of losses from cyber-attacks), controls and threats. We have now implemented an algorithm for calculating such a CVaR for an organisation, and in this white paper we present early insights that the research has produced.




Attack surface includes human beings as well as technology; therefore, you must inevitably consider that such a surface exists as there will be at least one human being working in your organisation.

# Harm propagates and the ripples can result in loss

The truth is that harm propagates; it is never solely the initial harm resulting from an attack that forms the entirety of the losses suffered. It initiates a ripple effect and each of the harms that result will have some kind of loss associated with them. These ripples need to be considered if we are to calculate CVaR, since if they are not then we may take overly optimistic views on the residual risk exposure. Understanding how risk controls can limit such ripples can demonstrate return-on-investment and underpin business cases.

The question of whether one has invested enough in managing the risk will matter significantly should an attack succeed. Where there is impact on the general public – whether it represents the loss of personal data, or reduced access to critical services – those with responsibility for overseeing organisations will be held to account by the public, by the press, civil society, and increasingly by the regulator and governing officials. Even where the loci of harm do not include the general public, similar accounts will need to be made to current customers, business partners, supply chains and competitors.

It will not be sufficient to rely on a lack of inherent knowledge about cyber-risk across organisations; the cyber-risk is no longer new and therefore some ability to identify and manage risk effectively is expected.

## Harm Propagation Factors and Healthiness Conditions for Cyber Value-at-Risk Calculations

To ensure that the results of CVaR calculation are plausible and reliable, we have established conditions that should hold under all circumstances. They are intuitive conditions which relate the factors that impact harm propagation to the overall CVaR:

- Increase in volume of threat and potential attack vectors increases CVaR
- Increase in threat likelihood increases CVaR
- Increase in risk control effectiveness decreases CVaR
- Increase in value of assets increases CVaR
- Increase in interdependency of assets increases CVaR

Any mathematical model for calculating CVaR should also meet the following, which taken in conjunction with the above forms a set of *healthiness conditions*:

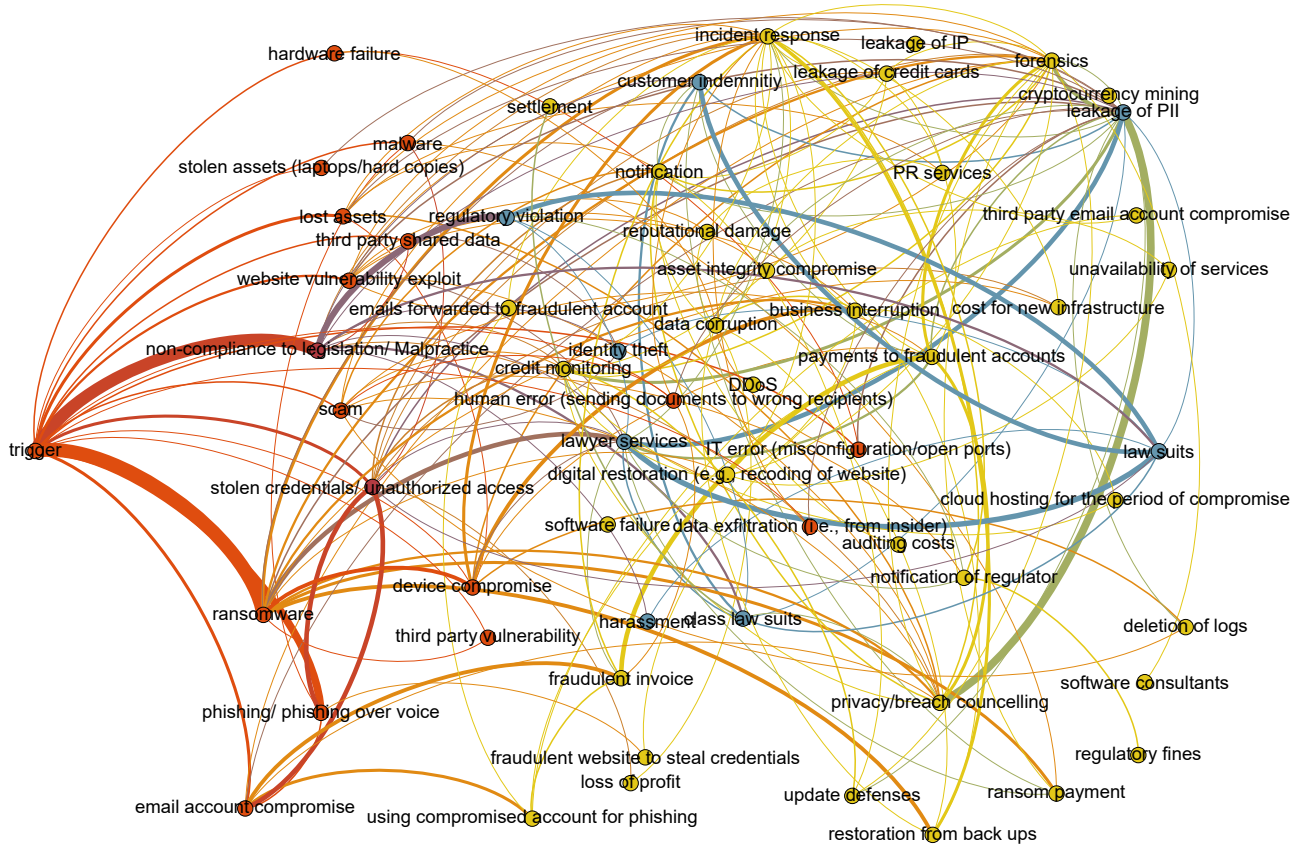
- The highest loss value in a CVaR distribution cannot exceed the sum of all losses relating to components of the system.

Practically we can use these conditions as tests providing evidence of the successful implementation of the mathematical model for CVaR in software.

# The direct impact of an attack is only the beginning,

## and what follows is a cascade of harms. These cascades are the reason why losses resulting from cyber-attacks can become so large.

To understand the potential for harm propagation, consider the graph below. This shows the harms occurring in more than 70 real cyber-insurance claims. The initial harm resulting from the trigger is simply the beginning, and what follows is a cascade effect. Such cascades are the reason why losses resulting from cyber-attacks can become so large. We call these 'harm trees' and use them in our system models so we can take account of how harm propagates. This means that we can model the range of potential aggregate losses that can result from a single attack event, mediated by the performance of the cyber-defense risk controls in place.

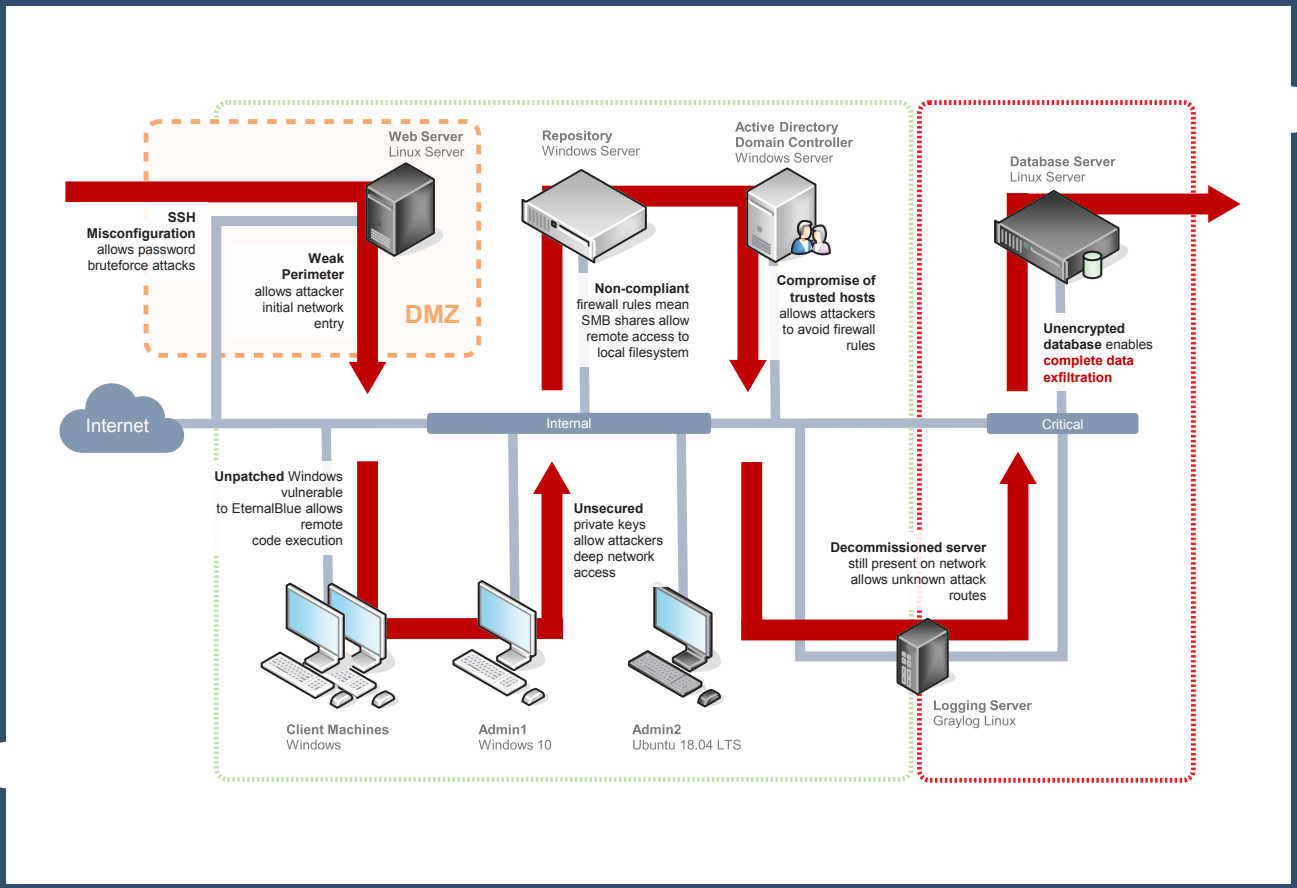



Graph representation of the most common harm ripple effects observed in the collected data. The root of each harm tree is the attack that triggers the harms, and all other nodes are harms that propagate from it. All nodes connected to the trigger node (left in the figure) are initial attacks, and the remainder of the nodes are the resulting harms. Edges are directed, the direction of the curved edges marks the direction clockwise. The thickness of the edges represents the frequency of that link, i.e., the thicker the more frequent. Note that frequency is not determined by a single edge, but by the addition of all edges running into a node. Colours are used to better differentiate the edges, without any specific meaning.

# Risk controls act together to create a compound effect

The very fact that harm propagates means that there is a connection between the risk controls we use; in the face of a single threat or attack, the level of losses we will realise is directly related to how we orchestrate multiple controls to limit the harm propagation. This not only involves technology, but also the practices and people we put around them.

Our work with the CVaR model has facilitated our demonstration of the real impact cybersecurity practice and the use of controls can have on our ability to contain the losses resulting from an attack. For the first time, we can take account of the possible effectiveness of controls and how that can directly impact how harm propagates and therefore the range of losses we are exposed to. Our model shows that the inter-dependencies between controls are real, and, therefore, we should be able to invest in cybersecurity in such a way as to maximise the compound effect and the strength of our overall cybersecurity posture.

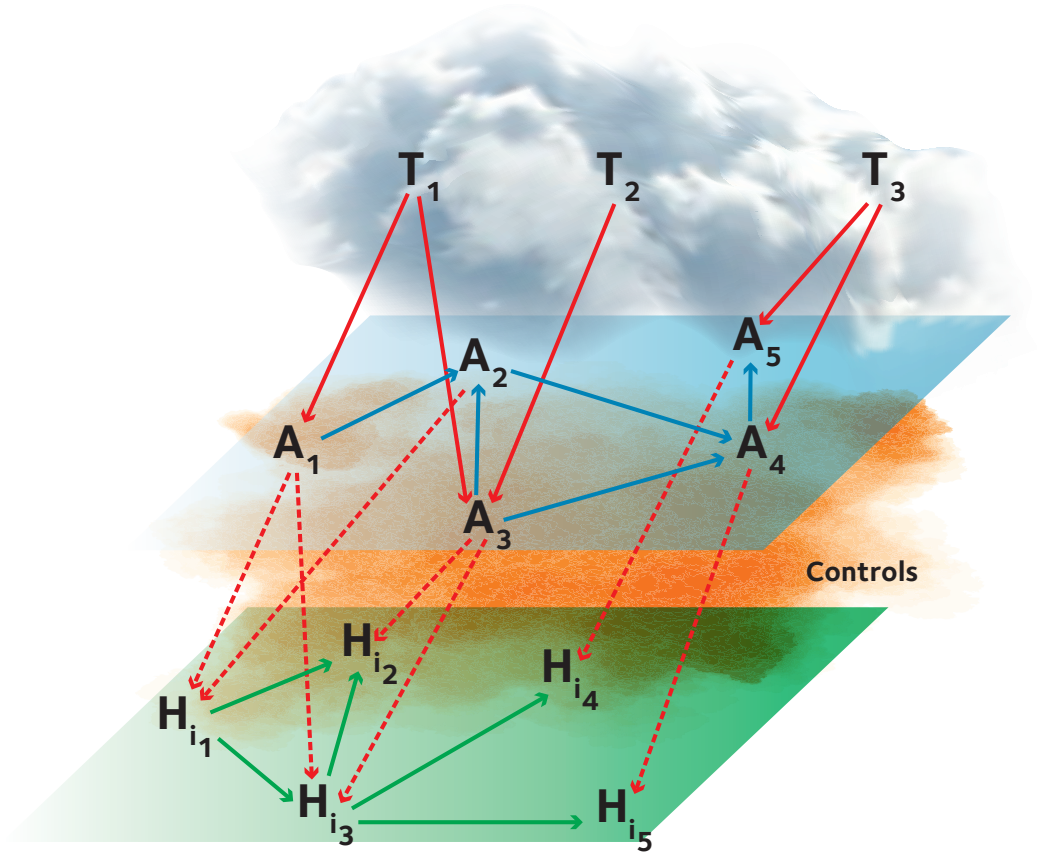


# Cyber Value-at-Risk [CVaR]

CVaR is in essence a probabilistic density function for losses from cyber-incidents. Calculating a CVaR for an organisation produces a range for total losses and the likelihood that they will occur.

Our treatment of losses is deliberately broad, and considers all harms that might arise from a cyber-attack or incident, and the summation of all losses that relate to those harms.

We say CVaR for an organisation is effectively constructed by summing the CVaR for all assets, given all threats of interest. Our knowledge of the threat that we may face could be highly sophisticated and therefore we can scope towards understanding our CVaR in the face of it. Or, we may wish to assume that a hypothetical threat exists with unbounded capacity to attack – allowing us to consider a situation where a threat emerges that has not been foreseen. The method remains the same.




Our approach takes account of the variability of risk-control effectiveness by abstracting away any environmental details (such as operational processes, configurations, use practices) to a simple binary model whereby a control is either effective or ineffective. An effective control both prevents harm to the asset AND prevents onward harm propagation. An ineffective control does not protect the asset AND may allow onward harm propagation.

This is an abstraction which recognises the lack of empirical data available on performance of risk controls; generating a more complex or subtle model would be too highly contrived without the evidence upon which to base it. But it is still a powerful abstraction – the model explicitly recognises that harm can propagate and allows for it to occur in the face of controls being in place in order to make a calculation of the CVaR range.

These two graphs show the CVaR calculation for identical systems, save for one difference: in the second scenario an additional risk control has been added. Specifically, in this case a data-encryption mechanism limits the harm-propagation potential should the data be exfiltrated. For both scenarios, we have calculated the CVaR for three different control-effectiveness assumptions. Note that the most extreme (if low-probability) losses are suppressed in the second scenario even with low control effectiveness, and that the effective cap on losses in the medium-effectiveness case comes down from around $60M to around $20M. The probability of the loss being contained to single figures with high control effectiveness is greatly enhanced.

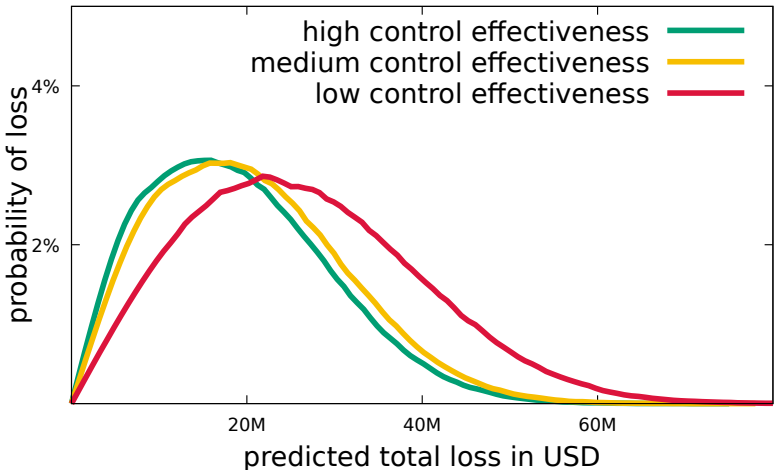



Figure 4. Monte Carlo simulation of CVaR with various levels of control effectiveness.

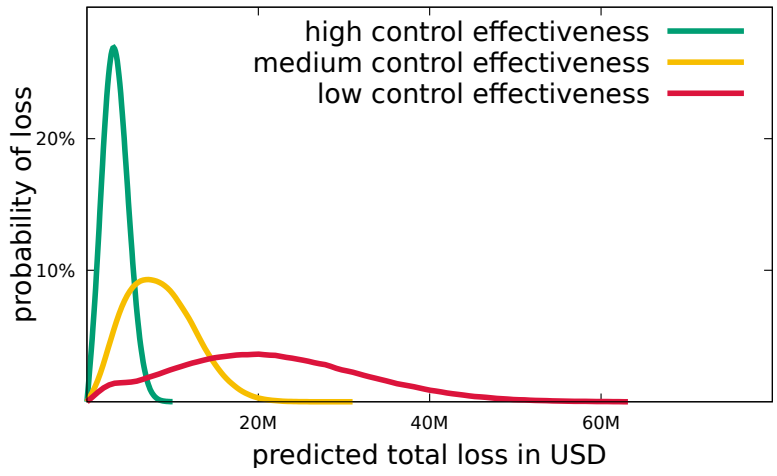



Figure 5. Same Monte Carlo simulation of CVaR with various levels of control effectiveness including an additional security control, which stopped the harm propagation so reducing the likelihood of losses.

# Knowledge of threat can help you take more risk

If you don't face a real threat, then you don't need to moderate your actions to avoid the related risks. If you know the capability and likely strategies of the threat you face, then you can adapt your cybersecurity practice to not only prevent such threats, but also detect and manage any which penetrate your perimeter.

Up until now this intuition has been embedded within the practice of cybersecurity by mandating the consideration of threat in risk assessments – only allowing risks to be enumerated where threat is determined to exist. Over time we expect to face new threats, and that our analysts will need to adapt our defences in order to deliver organisational resilience.

Within CVaR, we can explore the consequence of facing additional threats and compare how different security control configurations can limit residual risk as threat increases or changes. This means that it should be possible to identify where return on security investment is more worthwhile and where a decision to not invest can be made, thus delivering a platform upon which we can base strategic decisions on cybersecurity investments.

We can observe that, to effect the controls required for cybersecurity, we need a team that spans beyond the remit of the CTO or CIO. The majority of the technical controls are likely to be the responsibility of the chief information security officer (CISO). But there are procedural controls and security programmes for which HR, Operations, CISOs and data protection officers (DPOs) will need to coordinate for a successful implementation. Finally, more strategic controls need the attention of the executive leadership, board members, legal counsel and those involved in compliance and risk management. Risk appetite will be set at the highest level of oversight, and agreement will need to be reached on the cybersecurity recommendations made by the executive.

In conclusion, without the CVaR harm-propagation model, we would be left with worst- and best-case estimates and no method for taking account of how the use of heterogeneous risk controls, with variable effectiveness, can help to reduce entire organisational residual risk (as opposed to simply protecting specific assets).

**We introduce threat into a CVaR model by virtue of harm trees.** Selecting which harm trees to include in any CVaR calculation is important as each will potentially increase the CVaR. Ignoring harm trees that might relate to threats that will be faced is optimistic and could result in some losses not being accounted. Likewise, including harm trees that may relate to attacks that are never actually witnessed could be overly cautious or pessimistic about the level of risk exposure, and result in risk-control spend that could have been reduced. Of course, CVaR models can help us understand the consequence of ignoring real threats, demonstrating how the under-investments in controls manifest as enablers for harm propagation.

In other words, it is in taking account of harm propagation that we can begin to reason about organisational residual-risk levels – and in the recognition that controls can be faulty or ineffective in preventing harm propagation that we begin to get realistic about the true nature of the residual risk we face.

Our next steps will be to put this model into use in order that we can begin to refine its predictive capabilities and so that we can begin to benefit from this additional insight when planning how to invest in cybersecurity.

## This research has created the following six key takeaways.

1. **The cyber-threat faced by organisations is broad**, which means even if you face no threat from a targeted attack, you are still exposed to cyber-criminality. The attack source includes humans as well as technology.

2. **The direct impact of an attack is only the beginning**, and what follows is a cascade of harms. These cascades are the reason why losses resulting from cyber-attacks can become so large.

3. **Accounting for interdependency between controls is crucial**, and unfounded investment in cybersecurity can not only be wasteful but can also reduce the effectiveness of existing controls.

4. **Control implementation matters**, and mistakes can render defenses ineffective and the attack surfaces larger. Artefacts and tools left during the control configuration process can assist the attacker in escalating privileges and moving laterally.

5. **Knowledge helps us take more risk**, the more you know about the capability of the cyber-threats you are exposed to, the better you can adapt to prevent, detect, and manage them.

6. **Resilience requires broad knowledge**, since no organization has unlimited budget, the reality of operating in the face of cyber-risk is that resources must be focused towards risks with the capacity for greatest harm. CVaR helps estimate cyber-risk and allocate resources towards optimal control configuration.

UNIVERSITY OF OXFORD

DEPARTMENT OF COMPUTER SCIENCE

AXIS