On market concentration and cybersecurity risk

Dan Geera, Eric Jardine b and Eireann Leverett c

aIn-Q-Tel, Cambridge, USA; bVirginia Tech, Blacksburg, USA; cConcinnity Risks, UK

ABSTRACT
Market concentration affects each component of the cybersecurity risk equation (i.e. threat, vulnerability and impact). As the Internet ecosystem becomes more concentrated across a number of vectors from users and incoming links to economic market share, the locus of cyber risk moves towards these major hubs and the volume of systemic cyber risk increases. Mitigating cyber risk requires better measurement, diversity of systems, software and firms, attention to market concentration in cyber insurance pricing, and the deliberate choice to avoid ubiquitous interconnection in critical systems.

1. Introduction

Trends towards market concentrations are the new normal in IT systems, on the Internet, and across the World Wide Web. Concentration is often the cumulative result of myriad small, independent, and freely taken choices, though the deliberate acts of organisations to absorb competition ought not be minimised. Online, mechanisms of preferential attachment often dominate (Barabási and Albert 1999; Barabási 2014; Jardine 2017a). People tend to frequent online shops, services, and software that is already widely used. Everybody uses Facebook, Microsoft, Apple, Tik Tok, Snapchat, Whatsapp or Google, to name a few, because everyone else uses these same services. In many instances, the reward for interconnection, as famously dictated by Metcalfe’s Law, is proportional to the number of possible two-way connections, that is, proportional to the number of nodes squared (Metcalfe 1995; Metcalfe 2013; Zhang, Liu, and Xu 2015). The implication is that the greatest user and corporate value is often found at the most frequented places.

The cumulative result of numerous independent choices is a space with extraordinarily big platforms, systems and providers. The examples of such points of concentration range across operating systems (e.g. Microsoft Windows; Apple OS), protocols (e.g. WPA2; WiFi), e-commerce sites (e.g. Amazon; Alibaba), social networking sites (e.g. Facebook; Twitter), content delivery networks (e.g. Akamai; Cloudflare), cloud computing services (e.g. AWS; Azure), antivirus vendors (e.g. Symantec; ESET), aggregation platforms (e.g. Reddit), and so forth.

This tendency towards market concentration has a number of effects on systemic cyber risk, many of which mirror similar concentration risk effects in other sectors of the economy and society (Gürtler, Hibbeln, and Vöhringer 2010; Mandelbrot 2013; Zhang...
et al. 2013; Dhaliwal et al. 2016). These effects play out at every level of the cyber risk equation, that is, across threat, vulnerability and impact. While in no way exhaustive of the complexity in this domain, this paper details three theses on the relationship between concentration and cyber risk.

Thesis One: Market concentration has risk redistribution effects, often changing who gets targeted. Large operating systems, platforms, protocols and organizations often act as magnets for malicious activity.

Thesis Two: Market concentration can increase systemic vulnerability. Smaller organizations can sometimes improve their security posture by transferring their security to larger organizations. The degree of security improvement (if any) depends, however, on the volume of attacks, the distribution of attacks across organizations, and base rates of security performance across firms. In many cases, transfers of security to larger providers, such as Cloud services or CDNs, can actually increase levels of vulnerability under quasi knowable circumstances. Optimization is key.

Thesis Three: When points of major market concentration fall to malicious attack, the impact is far more significant than in more distributed systems due to significant (and often under recognized) issues of scale, complexity and ecosystem interdependency.

Overall, these three theses suggest that market concentration tends to have both redistributive and accentuating effects on cyber risk. Market concentration, broadly defined, tends to shift the risk of being attacked from individuals and minor nodes towards major players. Market concentration can also exacerbate systemic vulnerability and significantly increase the odds of high impact, system-wide failure. Dealing with cyber risk requires dealing with market concentration.

To make our case, the article unfolds this way: The following sections first detail the cyber risk (i.e. threat, vulnerability and impact) equation and define our conceptualisation of market concentration. Extending that line of argument, each section plots the interaction of patterns of market concentration and a discrete part of cyber risk, showing in detail and with illustrative examples how the former influences the latter. We conclude with policy suggestions for mitigating cyber risk in highly concentrated systems.

2. Defining cyber risk and market concentration

Cyber risk, like risk in any adversarial space more generally, is a function of threat, vulnerability and impact (Cox 2008; Hubbard and Seiersen 2016; Coburn, Leverett, and Woo 2019). The threat component of the equation is the probability that an organisation will be targeted by malicious activity. It can range from 0 per cent, not attacked, to 100 per cent, definitely attacked in a given period. The vulnerability component of the cyber risk equation captures what happens if an organisation is targeted by a malicious actor. Some intrusion attempts will fail. Others will succeed. The vulnerability component captures this variance by assigning a probability that a targeted organisation can mount a successful defence, given both the competence of the attacker and the skills of the defender. This component, too, ranges from 0 per cent to 100 per cent, i.e. always fails at security to always succeeds. The final component of the formula is impact, which denotes the dollars, hours of downtime, or some other measure of fallout from a malicious attack. Multiplied
together, the three components provide a sense of an organisation’s level of cyber risk over a given period.

Formally, the scope of market concentration in an ecosystem can be measured by the Herfindahl-Hirschman Index (HHI). The HHI is measured as the sum of the square of each organisation’s market share. The result of this mathematical process is a number ranging from 1 to 10,000, with perfect monopoly at 10,000 and a perfectly competitive system at an HHI value of 1 (Hirschman 1964; Hirschman 1980; Rhoades 1993). In practice, markets with an HHI above 2,500 are considered very concentrated.

More amorphously than a precise HHI number for each sector of the economy and Internet, concentration for our purposes can take several forms. Market concentration is about size, scale, centralisation and interconnectedness. It is, in the broadest terms, the amount of records, dollars, users, decision-making prerogative or degree centrality (incoming/outgoing links) a platform, provider, protocol, code base or organisation has when compared to the sum of all other like entities. Concentration can unfold along geographical space, time and name space. Systems with high levels of market concentration have, relative to the rest of the participants in that area of activity, big companies, big platforms, big providers, big intermediaries, big everything. Framed differently, concentrated organisations, platforms, or protocols have lots of links, customers, users or clients. Concentrated systems are ones where the choices of a big few affect the outcomes of the many. Concentrated systems can also be viewed as the opposite of diverse systems, meaning their component parts are more similar than not. Concentrated code bases, for instance, can be marked by something as simple as the proportion of used software written in the same programming language, e.g. Cobalt, anyone?

Concentration has several benefits economically. For example, firms plausibly emerge at all in order to contend with market-based transaction costs (Coase 2012). Bigger firms can also leverage their size to capture increasing return to scale and scope, improve their efficiency through greater task specialisation, and make longer term investments that maximise their revenue and profit growth. Too much scale, centralisation and concentration, of course, can each produce their own set of economic disadvantages, but there are clear economic reasons why organisations tend to get bigger, centralise their decision-making, and expand their proverbial footprints.

Yet concentration, as we detail in depth in the following sections, can also result in security/cyber risk consequences. Table 1 presents a non-exhaustive set of examples of the types of concentration we have in mind when we use the phrase, ‘market concentration’. Table 1 also details: (1) the reason it might be economically rational to concentrate

<table>
<thead>
<tr>
<th>Type of concentration</th>
<th>Potential economic benefits</th>
<th>Potential security consequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Users</td>
<td>Network externalities</td>
<td>The value of breaching the system grows supra-linearly as the user base increases</td>
</tr>
<tr>
<td>Homogeneous code base</td>
<td>Wide labour pool of potential coders to develop new platforms</td>
<td>Lower cost to develop and employ malware at the same or greater reward to the attacker</td>
</tr>
<tr>
<td>Homogenous platform use</td>
<td>Lower training and maintenance cost for a system</td>
<td>Increase the value of a vulnerability in the platform for an attacker</td>
</tr>
<tr>
<td>Centralised decision-making (ecosystem or</td>
<td>Improved efficiency of decisions; clear strategic direction; improved quality of service;</td>
<td>Homogeneous outcomes, if exploitable, lead to widespread compromise</td>
</tr>
<tr>
<td>organisational) Interconnection</td>
<td>reduced cost for components</td>
<td>Third party risks and ripple effects</td>
</tr>
</tbody>
</table>

Table 1. Forms of concentration and their economic, security, and cyber risk consequences.
and (2) some of the security consequences that can emerge from each type of concentration when viewed from an attacker’s perspective.

Table 1 is meant to highlight some of the diversity of possible cases that could be included under a comprehensive analysis of the interrelationships between market concentration and cyber risk. Our analysis below draws upon different cases that exemplify certain patterns and trends. We selected the cases largely for their data richness. They are not systematic, but are indicative of what the general theses in each section suggest we ought to observe (King, Keohane, and Verba 1994). In other words, the cases do not suggest that concentration always produces the detailed effects on security, only that the logic is there and some supportive evidence can be readily found.

Next, we detail how concentration interacts discretely with each component of the cyber risk equation and generally across all three component parts. These interactions affect both the distribution of risk and its amount of risk present in the system.

3. Thesis 1: market concentration and threat

Given a set of attack interests, capabilities and capacities, the threat component of the risk equation is about who gets attacked in the first place. More formally, and from the defender’s perspective, it is the probability that a malicious actor will expend the resources and time to try and compromise a given target’s systems. Threat matters because an organisation can be said to have zero current risk from malicious actors if it is never attacked. Of course, zero past risk only correlates with, and does not determine, future risk.

Not all organisations or individuals are equally likely to be targeted by malicious attacks, and some attacks are opportunistic rather than targeted (or rather they target a technology such as default SSH credentials, instead of an organisation or individual). Certainly, some famous cyber intrusions are simple attacks of convenience. Equifax’s failure to patch Apache Struts in 2017 created an easy way in for whoever compromised the system, suggesting more of an attack of opportunity than intent per se, though revelations of potential Chinese state involvement complicate the previously known story (Newman 2017). Other attacks, such as those that targeted the Democratic National Committee before the 2016 election, are highly deliberate, targeted and persistent.

Concentration directly affects the threat component of the risk equation by changing the malicious actor’s opportunity space. Concentration, in this sense, has predominately distributional effects on the threat component of cyber risk, namely, concentration leads to risk transference. Who gets targeted is a function of potential reward for malicious actors, which is, in turn, a function of a potential target’s size or network centrality. Increased size and position correlate with malicious actor payout, whether that reward is financial, disruptive or geopolitical.

As a market concentrates, the risk of being targeted by a cyberattack transfers from small, less central organisations to the major hubs or their counterparties (e.g. the Target Breach). For smaller organisations that might not be able to assume the financial cost of greater levels of cybersecurity, the transference of risk from small to big organisations would be a blessing. From the perspective of the highly concentrated organisations, the increased volume of attacks that need to be dealt with impose additional operating costs. In the Internet economy, being a major hub is often necessary for profitability (Jardine 2017a), but it also means that your platform will act like a magnet for malicious activity.
The history of malware targeting personal computer operating systems exemplifies this process of risk transference. For years, Windows OS was heavily targeted by malware designers. Mac OS, in contrast, maintained a mythology of a malware free system. This initial state of affairs, however, was not necessarily just a function of Macs having better OS security; divergent malware development rates for the two platforms is well explained by simple economic incentives.

Malicious actors should want to attack points with the highest reward ratio. And indeed, a formal model shows that market share drives malware development markets. Assuming equal levels of OS vulnerability, malware designers will target only the OS with the largest market share when the ratio of the dominant OS to the next largest OS is greater than the protection potential of the dominant OS overall (O'Donnell 2008). This formal model suggests that rational malicious actors would not even waste time designing malware to target Mac OS until it accounted for at least one-sixth of the OS market. Other formal models show similar trends in the concentration of malicious actor activity on dominant platforms (Arce 2018). In risk-based terms, the concentration of users on Windows OS for much of the history of personal computing had a redistributive effect on the threat component of risk. It transferred the risk of being targeted by malicious actors in the first place from Mac and Linux OS users to those who used Windows.

These earlier trends have had a legacy effect on risk among software platforms, making Windows still a more targeted system than alternative operating systems. But the final point that this example illustrates is more general: points of concentration promise the most reward and will be targeted the most frequently. Users of secondary systems, platforms or services might give up some benefits in terms of coordinated efficiency or positive network externalities, but might also reduce their cyber risk.

4. Thesis 2: concentration and vulnerability

The vulnerability component of cyber risk captures the expected/realized outcome of an attack on an organisation. It is a simple truism that attempted attacks sometimes succeed and sometimes fail. Attacks vary considerably in their sophistication and persistence. Organisational defenses, both technical and human, similarly vary, ranging from thorough to haphazard. Together, the interaction of the sophistication of the attacker, the defensive profile of the organisation, and a dose of luck produces some volume of successful and failed intrusions. In cyber risk terms, each organisation assuming they are attacked in the first place has a probability of being compromised over a given period. At the extremes, well-defended networks faced with unsophisticated attackers are likely to remain secure, while poorly defended networks faced off against competent and persistent opponents are likely to be breached.

Concentration affects the vulnerability component of the cyber risk equation. Three distinct effects emerge, each of which is unpacked more below. First, concentration in large providers can lead to an average reduction in the individual vulnerability of users, website operators and organisations due to the often greater security services provided by these major hubs. Second, since concentrated nodes tend to get targeted more often, joint probabilities suggest that they might be vulnerable across a large enough sequence of attacks. Third, platform and service concentration can also lead to increases in complexity, which can have negative effects at the level of software, increasing the proportion of
vulnerable code and worsening time-to-patch rates, leaving more systems vulnerable for longer windows of time. The first effect suggests that individuals can leverage scale to reduce their risk. The second and third suggest that at a systemic level, concentration can increase vulnerability and worsen cyber risk.

4.1. A first positive: professionalism, scale and security

Concentration is not all bad. Some individuals and organisations can improve their cyber risk profile, to a degree, by using concentrated services. Concentrated nodes can be, on average, better at providing security than less concentrated systems. The financial sector provides an example. Smaller firms (measured by revenue) actually suffer larger direct losses from cyber security events than bigger firms. This outcome is ‘possibly due to lower absolute investment in IT security’ (Bouveret 2018, 8). Security for smaller, less concentrated firms can be improved by transferring some or all of their security provision to bigger players. These security gains can be exemplified through the example of anti-virus software (AV), though other services such as spam filters or content delivery networks behave similarly. In each case, a combination of professionalism and scale advantages translates over into increased security effectiveness to some degree.

AV technology can allow users to leverage concentrated resources nested within large commercial firms to better protect the security of their devices. Scale effects are not the same as the effects of professionalisation. AV allows users to outsource personal device protection to a third party. By employing AV, users basically say that dealing with malware is better done by professionals. Malware development occurs at a remarkably fast pace. According to the AV Test Institute, upwards of 350,000 new malware variants are flagged every day (AV Test Institute 2019). In May of 2019 alone, 9.31 million new malware variants were recorded. Contending with this volume of malicious activity is challenging for individual users and organisations. Using AV is often a rational protective measure.

Professional services can provide better security than average users could produce on their own. While such security outsourcing can lead to maladaptive behavioural changes that may ultimately worsen security outcomes especially over the short run (Jardine n.d.), the record of the empirical effectiveness of commercial AV systems suggests there is malware protection to be had. One slightly older empirical investigation of AV effectiveness, for instance, found that 62 per cent of malware is effectively stopped at the water’s edge at the time of first exposure. Up to another 16.5 per cent is identified and blocked within one month’s time (Sukwong, Kim, and Hoe 2011).

Not all of this effectiveness, however, follows from the work of professionals. Market concentration also matters. Identification of new malware requires a bird’s eye view of network activity and trends, best had from a large user base and sensor network. Market concentration, in this case, is an effective driver of improved malware identification. The more users or sensors that an AV company has looking for new malware variants, the greater the ability of the company to identify emergent problems, produce signatures to protect users from new malware, and, by extension, provide better security for users. Clearly, some portion of the protections that are to be had from using AV come from outsourcing security from the individual to a professionalised third party. Yet some distinct gains also plausibly emerge as a function of the size of the third party itself: more sensors and users make AV malware detection better.
Spam filters in email would be another complementary example. Gmail can block spam very effectively by leveraging its significant user base to both crowdsource potentially malicious messages through manual tags and automatically detect spam and phishing attempts through automated matching of phrases given their huge corpus of available text. Content delivery networks provide similar benefits. CDNs can leverage in a coordinated way their capacity to deal with network traffic. The larger the CDN, the more capacity they can marshal and the better the protection they can, on average, provide. Other examples beyond these more technical services likewise apply. For instance, the Payment Card Industry Security Standards Council is effective at setting security standard adoption (such as PCI DSS) precisely because they are concentrated (Woods and Moore 2019).

In this sense, individual cyber risk can sometimes be reduced by concentrating security resources. Individuals alone are vulnerable. Individuals relying on concentrated firms can improve their baseline vulnerability rate considerably. Using commercial AV, content delivery networks, large email clients or cloud computing resources can, in other words, result in improved cyber risk outcomes for individual users.

4.2. A first negative: concentration, security and repeated attacks

Yet, even the extra security that the concentrated hubs can provide can also be overwhelmed. Here, the effect of market concentration on the threat component of cyber risk (i.e. who gets targeted in the first place) has a knock-on effect on the vulnerability of major nodes. The key mechanism at work is what is known as joint probability. The simplest example is what happens every time you flip a coin. Each round, regardless of what happened before, there is a 50 per cent chance that the outcome will be either heads or tails. The last fifty flips could have turned up heads, yet the next round will still be heads with a 50 per cent probability. This notion jars intuition and is in fact correct only when each toss is considered as a discrete event. Looking at the outcomes of the various coin flips as a sequence provides a fundamentally different answer. The probability that 51 sequential coin flips will come up heads is terrifically small (like 4.44e-16 small).

Joint probability is likewise a problem for cyber risk, and concentration can exacerbate this issue. If the concentrated points of the ecosystem – those with a lot of users, clients, incoming links, etc. – get targeted by malicious actors more because of their size, then they need to fend off huge volumes of attacks. Over most meaningful amounts of security incidents, the probability that the organisation that is defending a network will be successful across all attacks falls to nearly zero remarkably quickly.

Imagine, for example, three separate organisations. Given an ecosystem of attackers of various competency, assume that one has a 90 per cent probability of successful defence (Organization A). Another (Organization B) has a 99 per cent probability of successful defence. The last is very secure, with a 99.9 per cent (Organization C) probability of successful defence. Each of these organisations has a reasonably good chance of defending against any individual intrusion attempt. In other words, every individual port scan or phishing email would be deflected 90 per cent, 99 per cent, or 99.9 per cent of the time.

But, what happens across these organisations over a sequence of otherwise independent attacks is another story (Figure 1). As with the flip of 51 heads in a row, the
probability that the hypothetical organisations can successfully defend against every attempted attack drops to nearly zero in no time. For Organization A, successful defence across the full sequence of attacks drops below 1 per cent by the 44th attack (90%^44). Organization B remains resilient for longer, but drops below a 1 per cent probability of successful defence across all attacks by the 459th intrusion attempt (99%^459). Organization C, finally, is by far the most robust of the three (pointing to the exponential effects of joint probabilities). At the 459th attack, Organization C still has a 63.18 per cent chance of successfully defending against all attempted intrusions. In fact, it takes 4,608 attempted attacks for the probability of successful defence for Organization C to fall below the 1 per cent mark (99.9%^4,608). This is clearly a much-improved outcome, but still potentially insufficient at Internet scale. For context, the US Department of Defense alone receives 36 million potentially malicious (non-legitimate) emails per day (Konkel 2018).

Figure 1. Repeated attacks and the vulnerability of highly targeted hubs.

Joint probability also lends some clues as to the best security-maximizing step for smaller firms. A common argument in favour of using Cloud providers or CDNs is that their scale contributes to improved security performance, making them a better choice for smaller organisations than in-house security. Imagine the same top-performing organisation as above, which is effective at security in a single run 99.9 per cent of the time. Picture this organisation as a large concentrated firm, akin to a CDN or major Cloud provider. Imagine further a smaller organisation that needs to provide security without any internalised benefits of scale and is effective 75 per cent of the time. The smaller organisation would be better off moving its security provision to the larger organisation in the event that they were both attacked once. In a hypothetical one-shot world, the smaller firm would be 24.9 percentage points more secure by transferring its security provision to the larger organisation.
In a world of multiple and unevenly distributed attacks (i.e. the real world), the best security-maximizing choice for the smaller firm becomes less clear and subject to a process of optimisation. The distribution of malicious activity, the total volume of attacks in the ecosystem, and the baseline security performance of the organisations in question all interact to determine the optimal choice. Figure 2 showcases three hypothetical attack-clustering scenarios.

The first scenario is one of extreme clustering. The comparative net target value of the big firm is so great that all attacks go towards the concentrated organisation in the ecosystem. Since the threat component of the cyber risk equation then equals zero for the small organisation (i.e. zero attacks * 75% successful defense * impact = zero), the small organisation is better off not transferring its security provision to the larger, better at security organisation. Linking the fate of the small firm to the big firm if the big firm suffers from all the attacks in the ecosystem can result in rapidly overwhelmed defences and a worse security outcome for the smaller organisation. Indeed, in a world where the small firm simply passes under the radar of a malicious actor, any volume of attacks makes coupling the small organisation’s fate to the bigger firm a bad security deal.

In a world without a perfect clustering of attacks, the smaller organisation needs to optimise. In the second scenario in Figure 2, the smaller organisation does get attacked one time during the period in question. It has a 75 per cent chance of successfully defending against the attempted intrusion. From the small organisation’s perspective, it can get more security by transferring its security provision to the larger organisation, as long as that organisation is attacked 287 or fewer times. Any more attacks, and it again becomes rational for the smaller organisation to maintain its own security, everything else being equal. The last scenario is like the second, but the smaller organisation gets attacked twice over the hypothetical period. The rate of clustering of attacks against the bigger organisation, in other words, is slightly less than before. In this case, the joint probability that the small organisation will successfully defend against both attempted intrusions is 56.25 per cent (i.e. 75%\(^2\)). There is, as a result, a wider range of security improvements that could be had if the small organisation transfers its security provision to the larger organisation. Indeed, if the bigger organisation is attacked 575 times, then the two

![Figure 2. Outsourcing security in a world of clustered attacks.](image-url)
organisations are equally vulnerable. Any fewer attacks against the bigger firm and the smaller organisation would be better off letting the big organisation do the heavy lifting of security.

All this suggests that a concentrated ecosystem, by attracting malicious activity to the main hubs, also potentially produces negative knock-on effects for the vulnerability component of the cyber risk equation. Over enough attempted intrusions, the joint probability of successful defence across a large number of attacks asymptotically approaches zero. This analysis also highlights the important security optimisation choices that smaller, less targeted organisations need to make. Sometimes transferring security to a larger and more skilful organisation can be of benefit, but it depends upon how clustered the attempted intrusions in the ecosystem are, how many attacks there are in total, and what base rates of security performance are at play.

Additionally, the analysis showcases the importance of multilayered organisational resilience, which could be conceptualised as placing sequential barriers in front of malicious actors as opposed to a single external firewall model. Interestingly, multiple barriers through which a malicious actor needs to successfully pass in order to gain meaningful access to a network or data flips the joint probability analysis above on its head: with each new well-defended gate through which a malicious actor needs to pass, it becomes exponentially more likely that the network will remain secure. For some firms such as CDNs or Cloud providers, maintaining redundant systems that are sufficiently distinct (perhaps even geographically distributed) would help to make joint probabilities work in the favour of the defender.

Yet, even factoring in resiliency, the defender needs to be effective at their primary duty every time, while the attacker need only succeed once to compromise the CIA triangle (confidentiality, integrity and access) at the core of information protection (Geer 2018b). Market concentration can make that harder to do.

4.3. The second negative: concentration, scale and complexity

Concentration is multidimensional. Concentration involves scale, which is part of the reason why individuals can potentially improve their security profile by relying on large services. Concentration as it implies scale also involves increased complexity. Nothing big is simple. Issues of complexity can play out at two levels: the baseline volume of faulty code and the time it takes larger organisations to patch flaws.

The largest edifices of code are, everything else being equal, likely to be the most complex. The link between concentrations of software use and software complexity is fairly clear. Highly used software will have some combination of a decisive incumbent user advantage over existing alternatives, a strong competitiveness factor that helps unseat incumbents in a world of preferential attachment (Barabási and Albert 1999; Barabási 2014) and diffuse functional design, allowing people to accomplish multiple task and potentially innovate, depending upon its openness, on top of the code (Zittrain 2008). In such cases, but especially with regard to wide-ranging function, more complexity of code tends to be correlated with more users.

Yet, by dint of that very complexity, concentrations of code are also the most likely to contain defective routines that can be found and exploited by malicious actors. One part of the equation is simply a function of the number of lines of code: more code, more
potential defects, everything else being equal. But the accumulation of vulnerabilities might not be constant at scale, since scale entails complexity and, as Steven Johnson puts it, ‘complexity characterizes the behaviour of a system or model whose components interact in multiple ways and follow local rules, meaning there is no reasonable higher instructions to define the various possible interactions’ (Johnson 2001, 19). Put another way, the possibility of a flaw is not linearly more likely as the edifice of code grows. It is something else and it is hard to even say what rate of increase might be at play by dint of its very complexity.

The problem is particularly acute in today’s global market for platforms and services. The market, especially with newer technologies such as the Internet of Things, is based around rapid innovation, first-to-market behaviour, and the use of stock code with overlaid tweaks for function. Products are often developed, sold, shipped, and then security is layered on afterwards if the market wants the good and can sustain the product. This pattern gives rise to persistent vulnerabilities, that are easy to introduce, slow to be discovered, and often difficult to fix because they are so embedded in how a product is architected. These vulnerabilities can be weaponized. When they are, unpatched systems become vulnerable to malicious intrusion. Framed differently, the result of the need for speed is reused code, which is often selected based upon its expressiveness (Turing-completeness). The trouble is, ‘the very code that has the greatest probability of being reused is the code that has the greatest probability of being rich enough in complexity to obscure exploitability’ (Geer 2018a). A few simple lines of code are likely to be error free; a complex edifice of code is disproportionately likely to have numerable hidden vulnerabilities that increase vulnerability and attendant cyber risk. Code that is self-modifying (think machine learning) may not even be analysable (Geer 2019).

Isolated flaws are fixable. When found, patches can be developed and software can be updated. With static code, constant observation and sufficient time, any given code base, no matter how complex, will eventually rid itself of most errors (Ozment and Schechter 2006). Unfortunately, at least two of these conditions often do not exist, as code is rarely static and time is always of the essence. Regardless, assuming no adverse interactions, the faster software can be patched, the less vulnerable users of that software become. Software patches are, however, embedded in complex socio-technical systems, involving software designers, bug hunters, malicious actors, operators of the systems running the code and users who often demand continuous service. All that is to say: just because a vulnerability exists does not mean it will be found; just because it is found does not mean it will be exploited; just because it is exploited does not mean it will be patched; and just because it is patched does not mean, necessarily, that the patch will be deployed in any sort of polynomial time (Ablon and Bogart 2017; Herr, Schneier, and Morris 2017; Geer 2019).

Out of this socio-technical assemblage of software development and maintenance comes a simple security metric, known as time-to-patch rates or remediation velocity (Kenna Security and Cyentia Institute 2019). Time to patch is the time it takes for an organisation to correct a certain proportion of vulnerabilities on its systems after the announced discovery of the vulnerability and the issuance of a correction by a vendor. Generally, faster patch rates are better, but organisations need to balance a multitude of incentives and costs, meaning that immediately patching all vulnerabilities is neither cost effective or even always a risk reducing step. After all, installing patches
can also break things and prevent you from doing business, which may cost more than suffering a breach.

Many factors affect organisational time-to-patch rates (Kenna Security and Cyentia Institute 2019). Organisational size, itself directly influenced by patterns of market concentration, also matters. Concentration can have three potential effects on time-to-patch rates through the concept of returns to scale. First, increased scale might improve patch rates by reducing the average time-to-patch as scale increases (a pattern of increasing returns to scale). Larger firms, for example, might have disproportionately more resources to devote to patching vulnerabilities. Second, scale might matter little at all for remediation velocity (constant returns to scale). Lastly, there could be maladaptive effects at work, where increased scale leads to a worsening time-to-patch rate (decreasing returns to scale). For example, a larger organisation might be more centralised and complex, generating bureaucratic choke points that limit the timely remediation of security vulnerabilities.

Each scenario is possible, but empirically scale tends to have a negative effect on remediation velocity (Kenna Security and Cyentia Institute 2019). Data from Kenna Security and The Cyentia Institute plotted in Figure 3 illustrate just how much worse medium and large firms fare compared to smaller organisations on the time-to-patch measure, across both exploited and not yet exploited vulnerabilities. In a total sample of 300 separate enterprises, smaller firms (1–500 employees) routinely outperform both medium (500–5,000 employees) and large (more than 5,000 employees) organisations. Consistent with the idea that increasing scale and complexity are a harm to patch rates, the largest gap is between small and large firms, with medium size organisations falling in the middle in terms of performance. For instance, compared to small firms, it takes a median time of nine days longer for large organisations to patch 25 per cent of unexploited vulnerabilities, 47 days to patch 50 per cent, and 157 days to patch 75 per cent (Kenna Security and Cyentia Institute 2019). All firms take less time to patch exploited vulnerabilities, which makes sense since these flaws are known to be weaponized. However, larger firms once again take the longest time to patch 25 per cent, 50 per cent or 75 per cent of known,

![Figure 3. Time-to-patch and the cyber risk of large organisations.](image-url)
exploited vulnerabilities in their systems. The evidence suggests that scale harms organisational patch rates, making more systems vulnerable for longer. Market concentration can worsen systemic vulnerability, generating more cyber risk. While individuals can, at times, improve their personal security by using large services for AV protection, website content delivery, or as an email provider, scale sometimes plays out badly for cyber risk in aggregate. Because larger nodes are attacked more frequently, the odds that large providers of various stripes will not be beaten eventually, to the detriment of their users, are vanishingly small over a sufficiently large number of attacks. Likewise, at the level of computer software, market concentration tends to worsen the odds that exploitable vulnerabilities exist. Additionally, larger organisations tend to be worse at patching defects in their software systems, creating longer windows of vulnerability.

5. Concentration and impact

If targeted and then compromised, what happens next (i.e. the fallout or impact of a cyber-attack) is hugely variable, but likewise dependent on patterns of concentration, among other factors. One way to conceptualise concentration and scale is to think of big organisations as being highly connected, in the sense that what happens to the big organisations ripples out and affects many others. A diffuse system is one with isolated pockets, without, in network theory terms, connecting edges to other nodes. In highly concentrated systems, everything is connected, especially to the big nodes (e.g. organisations or platforms) in the system. Such interconnections can take a number of forms, but readily include links between organisations via data sharing agreements, common supply chains and third party vendors. Interconnections create transmission pathways by which the negative consequences of a security incident can percolate through a whole system (Geer 2018a).

Content delivery networks (CDNs) are a good example of how a cluster of users on a single service can amplify the potential impact of major security incidents. CDNs move content closer to the edge nodes of the network and allow users to both get better service quality and improved protections from distributed denial of service (DDoS) attacks. Because CDNs can leverage resources in a coordinated way, they tend to be more efficient than a loosely interlinked group of, say, individual website operators who are all providing their own bandwidth to handle incoming traffic. That efficiency means that there are fewer system-wide resources to handle big malicious events than would otherwise be the case, but the CDN can leverage the available resource in a more coordinated way, bringing more capacity to bear at any given point in time. Effectively, CDNs, like AV vendors before, can sometimes reduce individual user vulnerability by leveraging their increased scale in a coordinated way.

But when the CDN or other major Internet node fails, the effect of concentration on the impact element of the cyber risk equation can be large indeed. Take the 2016 attack on Dyn, a domain name system provider, as an example. The attack was well into the terabyte per second range (the largest attack to date is a 1.3 tbs attack on Github), but still represents only around one-hundredth of the latent DDoS potential worldwide (Leverett and Kaplan 2017). The attack temporarily disrupted the services of some 69 different Internet-based companies, ranging from AirBnB to Zillow, but including such big named brands as Netflix, Spotify, HBO, Fox News and Twitter. Cumulatively, the economic cost of the attack was huge. (The more interconnected, the more counterparty risk matters.)
The root of the problem in the Dyn case was that these companies all used Dyn’s services to resolve their DNS queries (this example showcases concentration in name space, since the DNS itself is a distributed protocol and Dyn, to a lesser degree, maintains a geographically distributed infrastructure). When this concentrated Goliath fell, the ripple effects were large indeed. Consider for a moment the counterfactual, where each company maintains its own DNS services. The system would be far less efficient and less able to handle major traffic requests. Yet when one point in the system failed, the scope of the potential damage or disruption (impact, to use the cyber risk term) would be minimal. AirBnb might fall for a period if hit by a 1.3 tbs attack, but this would not influence the provision of service by Spotify, HBO or any other online services. Decentralisation, as opposed to concentration, can limit the potential impact that any given cyber attack might have by precluding common-mode failure.

Lest we be fooled, the adverse effects of market concentration on the impact component of cyber risk are not confined to online services. Operating systems are another clear example. Microsoft is the dominant OS used in most PCs attached to personal and government networks. When a new zero day for Microsoft hits the wild, the effects can be large. The WannaCry attack in May 2017 is one such example. The malware made use of the Eternal Blue exploit, which was reportedly found by the National Security Agency (NSA), leaked online by the Shadow Brokers, and then used by some unknown party to launch a globe-spanning ransomware attack.

The exploit targeted systems running an unpatched version of Microsoft 7, 8 and XP. Over the span of a weekend, the attack spread to well over a hundred countries and affected some 200,000 devices, with the National Health Services in the United Kingdom getting hit particularly badly. The attack was aborted early by the actions of a single computer enthusiast/hacker named Marcus Hitchens, who accidently stopped the attack by registering the domain name kill switch, preventing its further spread (Solon 2017). WannaCry’s spread in response to a single exploit once again highlights the potential effects of market concentration on the impact component of cyber risk. When everyone is on the same system, not only does that system attract the bulk of malicious attacks, making it likely to fail over a large enough volume of attacks, but the potential impact of a successful attack moves from a private tragedy to a system-wide, cascading and potentially significant event.

Indeed, the problems of scale and interconnection are not as idiosyncratic as these examples might suggest. More systematic evidence suggests that (a) security events that come to involve multiple parties are disproportionately costly compared to single organisation events; and (b) larger organisational size tends to generate more pronounced downstream ripple effects, suggesting that the more big organisations there are in a sector, the more that knock-on effects ought to be expected (Riskrecon and Cyentia Institute 2019). The median cost of a breach of a single organisation is $77,200, with no downstream or upstream ripple effects. The median cost of a multiparty event is $999,500. On a frequency distribution, the extreme end cost (95%) for a single organisation is $16,000,000, while the same 95 per cent event involving multiple organisations through downstream and upstream interconnections is $417,362,204 (Riskrecon and Cyentia Institute 2019). The compromise of multiple organisations should obviously cost more than the breach of a single firm, but the cost of ripple effects due to interconnection is out of proportion with the number of total victims. The ratio of downstream to originating organisations affected by
a security incident in a multiparty breach is 8-to-1. The ratio of median costs for a single vs multiple party incident, however, is 13-to-1 (Riskrecon and Cyentia Institute 2019).

Big nodes measured by companies with more employees in this case also create bigger downstream ripple effects. An ecosystem with many big organisations is, therefore, likely to suffer from far more and far costlier security incidents as the effects ripple out across the various interconnected links, spreading from a central organisation to third party vendors, downstream clients and so forth. For example, while companies of every size as measured by the number of employees can cause a multiparty security incident, larger firms are more often the source of a multiparty security event, while small-to-medium sized enterprises are usually the recipients of such events, as detailed in Figure 4 (Riskrecon and Cyentia Institute 2019).

Concentration, in other words, negatively affects the impact component of the cyber risk equation. Bigger organisations or platforms have lots of users who can all be simultaneously affected by a security incident. Bigger organisations also tend to be more interconnected with others via supply chains, third party vendors and other service providers. In either case, concentration to the extent that it denotes greater scale and interconnection tends to create the scene for a disproportionately pronounced fallout from a security incident.

6. Contending with cyber risk in highly concentrated markets

Concentration influences all three components of the cyber risk equation: threat, vulnerability and impact. It plausibly causes a redistribution of risk that affects the threat component, shifting attacks from infrequently used nodes towards the larger hubs of activity. It also plausibly improves individual security outcomes in the median case by transferring the work of security from individuals to large hubs with advantages of professionalism and scale. These hubs can leverage their scale to produce more security but are at risk of being undone by being targeted so much that even their improved protections are simply overwhelmed. Lastly, larger hubs increase the odds of significant failures, worsening the potential negative outcome of a cyberattack.
The cumulative result of this analysis is fairly simple. Many forms of market concentration, from protocols and code to links, users, services, firms or platforms, both shifts and aggravates cyber risk. If that is the case, then the issue of cyber risk becomes an increasingly relevant concern as more of society’s functions transition to online-only. If the Internet does eventually become, as former Swedish Prime Minister Carl Bildt put it, ‘the infrastructure of all other infrastructures’, then trends in market concentration present some very real, society-wide cybersecurity challenges (Bildt 2015).

Norm building efforts aimed at preventing malicious use of the Internet and IT systems in the first place can help manage risk (Governance 2015; Bildt and Smith 2016; Cyber-space 2019). More precisely to the issue of concentration, remedial steps, at their core, advance along four lines:

First, cyber risk can be managed only to the extent that it can be measured. Models of managing cyber risk at the level of an individual firm suggest that expenditure should not exceed 37 per cent of the expected costs of a security incident (Gordon and Loeb 2002; Gordon and Loeb 2006; Baryshnikov 2012; Geer 2015; Gordon et al. 2015). Yet, accurately measuring cybersecurity events is hugely challenging, as measures often suffer from problems of the denominator (Jardine 2015, 2018, 2020), incompatible metrics (Brecht and Nowey 2013), insufficient attention to over time trends (Geer and Jardine 2017; Jardine 2020), measures distorted by political or economic incentives (Anderson et al. 2008; Anderson et al. 2013; Lawson and Middleton 2019), a lack of data transparency necessitating clever measurement techniques (Woods, Moore, and Simpson 2019), reporting biases (Florêncio and Herley 2011) and data aggregation problems (Jardine 2017b; Jardine 2020). Issues of technological flux likewise present a challenge where past data might supremely fail to predict future events. The task of producing better measures is not easy, but it is fundamental. Good risk mitigation policies require a concerted effort to better measure every aspect of the Internet ecosystem in a way that at least allows for the potential for data-driven policy.

Second, countering growing cyber risk that arises from trends toward greater market concentration requires the deliberate development of ecosystem diversity, even though such a process often runs afoul of the sound economic logic pushing towards greater scale and centralisation (Coase 2012; Geer 2018a). Redundant systems in large clusters are not sufficient. Redundancies understood as more of the same will not help protect against the cyber risks that follow from greater market concentration. Ten thousand redundant systems all running Microsoft Windows or Apache Struts can fall like dominos when these systems are compromised. Like in nature, diversity can increase resiliency and prevent cascading failures. Diversity of systems might introduce inefficiencies that hurt the bottom line, but they protect against costly large-scale failures. Promoting diversity, especially for critical services, is likely key if cyber risk is to avoid reaching or potentially pull back from a level where an Internet-equivalent of an extinction event is possible. Concrete methods for promoting diversity of systems could range from anti-trust legislation in an extreme to tooling government or corporate procurement of IT technologies and services to prevent homogeneity of providers/platforms.

Third, calculations of cyber risk insurance need to factor in ecosystem-wide trends in market concentration in order to accurately price risk, especially to the extent that cyber risk exhibits features that are distinct from other risk types (Biener, Eling, and Wirfs 2015). While organisational factors (such as corporate revenue and assets) are a common
feature of cyber risk insurance policies (Romanosky et al. 2019), each firm remains nested in an sector with distinct patterns of market concentration. The way in which proportional organisational size interacts with malicious actor incentives suggests that accurate insurance pricing requires knowing not just how big in absolute terms a prospective policyholder might be, but also how large of a share of the ecosystem (sector/industry) they represent. In highly concentrated spaces, premiums for the main nodes should be disproportionately high and payments for others comparatively low, everything else being equal. In more competitive markets, premiums weighted for asset value and organisational security procedures should trend towards a median value. Yet, issues such as the over-supply of insurers can inhibit effective security governance through insurance pricing mechanisms (Woods and Moore 2019; Woods, Moore, and Simpson 2019).

Fourth, for certain systems, measurement to manage risk and diversity might be insufficient and persistent disconnectedness might be needed (Geer 2018a). Interconnection in many aspects of life may increase risk, yet do so within manageable bounds. Certain aspects of a country’s critical national infrastructure, however, is likely too important to interconnect, even if such connection is done with the best security standards in mind. The implication, really, is that high impact events that affect many due interconnection and concentration might be limited via disconnected firebreaks. Senate Bill 79, which was recently included in the National Defense Authorization Act, proposes just such a move in the United States with regards to US energy infrastructure. To mitigate risk, certain parts of a nation’s infrastructure might be best left off the proverbial table.

7. Conclusion

In sum, there is an often malign association between trends in market concentration and the location and level of cyber risk within the system. Much of these patterns are not unique to cyberspace and could affect other facets of risk in finance and other sectors. Market concentration affects all three elements of the cyber risk equation, as it does to a degree in a host of other sectors, but particularly the financial sector (Gürtler, Hibbeln, and Vöhringer 2010; Taleb 2010; Mandelbrot 2013; Zhang et al. 2013; Kasman and Kasman 2015; Dhaliwal et al. 2016). Via the threat component, trends in market concentration redistribute risk, changing who gets targeted by a malicious actor in the first place. Everything else being equal, attackers tend to target the concentrated hubs because that is where the biggest reward resides. Via the vulnerability component, shifting market concentration levels can both reduce individual risk to a degree, but also increase systemic vulnerability because big services (players) are targeted so much that they eventually fail and larger organisations also exhibit the worst time-to-patch rates, leaving them and their users vulnerable for longer. Finally, via the impact component of the risk equation, concentrated nodes can mask unseen interdependencies that can give rise to massive, cascading failures.

Imagine a hypothetical example of perfect market concentration. All users cluster onto one platform. All attacks flow at that platform. Eventually that platform fails, and all users are compromised. This extreme example is a difference in degree, rather than a difference in kind, from the events that unfold due to current trends in market concentration. The implication is that countering trends in market concentration, a process that means at one level dealing with how people freely choose to use services and platforms (Barabási
and Albert 1999; Barabási 2014; Jardine 2017a), is essential for managing cyber risk on the Internet. Conversely, attempting to deal with cyber risk without also contending with trends in market concentration are likely to fail.

**Note**

1. We are grateful to Richard Willey at Akamai for helping me think this issue through.

**Disclosure statement**

No potential conflict of interest was reported by the author(s).

**Funding**

This paper is part of a wider project, ‘Incrementally Tailoring a Better Cyber Risk Score’, funded by a Comcast Innovation Grant. Grant number: 2019-145.

**Notes on contributors**

*Dan Geer* is a computer security analyst and risk management specialist. He is recognised for raising awareness of critical computer and network security issues before the risks were widely understood, and for ground-breaking work on the economics of security.

*Eric Jardine* is an assistant professor of political science at Virginia Tech and a research fellow at the Centre for International Governance Innovation (CIGI). His research focuses on the uses and abuses of the Dark Web, the measurement of trends in cybercrime data, and the inherent politics surrounding both anonymity-granting technologies and encryption. He has given a Tedx talk on emergent privacy challenges of the digital age, a congressional staffer briefing on the Dark Web and cryptocurrencies, and spoken numerous times at the United Nations Conference on Trade and Development about trends in online trust. His scholarly work has been published in a number of peer reviewed outlets, including New Media & Society, Journal of Cyber Policy, First Monday, Intelligence and National Security, Terrorism and Political Violence, and Studies in Conflict and Terrorism, among numerous others. He is the co-author, with Fen Hampson, of Look Who’s Watching: Surveillance, Treachery and Trust Online. More information can be found at his website, www.measuringcyber.com

*Éireann Leverett* once found 10,000 vulnerable industrial systems on the internet. He then worked with Computer Emergency Response Teams around the world for cyber risk reduction. He likes teaching the basics, and learning the obscure. He continually studies computer science, cryptography, networks, information theory, economics, and magic history. He is also fascinated by zero knowledge proofs, firmware and malware reverse engineering, and complicated network effects such as Braess’ and Jevon’s Paradoxes. He has worked in quality assurance on software that runs the electric grid, penetration testing, and academia. He likes long binwalks by the hexdumps with his friends. Éireann Leverett is a regular speaker at computer security conferences such as FIRST, BlackHat, Defcon, Brucon, Hack.lu, RSA, and CCC; and also a regular speaker at insurance and risk conferences such as Society of Information Risk Analysts, Onshore Energy Conference, International Association of Engineering Insurers, International Risk Governance Council, and the Reinsurance Association of America. He has been featured by the BBC, The Washington Post, The Chicago Tribune, The Register, The Christian Science Monitor, Popular Mechanics, and Wired magazine.

**ORCID**

*Eric Jardine* [http://orcid.org/0000-0002-2041-314X](http://orcid.org/0000-0002-2041-314X)

*Éireann Leverett* [http://orcid.org/0000-0001-6586-7359](http://orcid.org/0000-0001-6586-7359)
References


