Systemic Cyber Risk: A Primer

David Forscey, Jon Bateman, Nick Beecroft, and Beau Woods
Contents

Introduction 1

What Is Systemic Cyber Risk? 3

What Does Systemic Cyber Risk Look Like? 6

What Can Cause (or Mitigate) Systemic Cyber Risk? 9

How Can Sources of Systemic Risk Be Identified? 11

How Can Policymakers Respond? 13

Conclusion 16

About the Authors 17

Notes 19

Carnegie Endowment for International Peace 25

Aspen Institute 27
Introduction

There is growing concern about “systemic cyber risk”—the possibility that a single failure somewhere in cyberspace could cause widening ripples with catastrophic consequences. Whereas most cyber events have a narrowly defined set of victims, a systemic cyber incident could do damage on a national or even a global scale—threatening the digital infrastructure that entire societies, economies, and governments rely on to function. In the last few months alone, two very different events illustrated distinct versions of the problem.

On November 24, 2021, Chinese cybersecurity researchers disclosed a severe vulnerability in Log4j—a low-profile software utility embedded in millions, or perhaps billions, of consumer devices and enterprise systems around the world. The security flaw could permit hackers to take total control of vulnerable machines with relative ease. The job of fixing Log4j fell to a team of volunteer programmers at Apache, who took two weeks to release a security patch. By that point, the hacking had already begun. The first patch was then followed by a second patch and a third patch, as more security gaps were uncovered. Meanwhile, organizations struggled to apply these patches because Log4j is often hidden underneath layers upon layers of other software packages. Experts predict it will take years to fully resolve the issue. Until then, innumerable victims remain vulnerable to state-sponsored hackers, ransomware gangs, and other bad actors.

Compare the Log4j incident—a slow-rolling crisis actively abused by malicious actors—with another recent global event that was shorter, sharper, and completely accidental. On October 4, 2021, billions of users worldwide lost access to all Facebook services, including Instagram and WhatsApp. This happened because a small error during routine maintenance had unexpected and cascading consequences. An errant command was entered, and a bug in
Facebook’s auditing systems mistakenly allowed the command to run, disconnecting all data centers. Misjudging the situation, Facebook’s DNS servers reacted by automatically halting public advertisements, blinding the internet to Facebook’s online location. Meanwhile, widespread network failures blocked Facebook’s IT staff from accessing the affected systems, even physically, to restore them. Although the outage lasted only six hours, that was a lifetime for many small businesses, family networks, and others reliant on Facebook for their daily needs.

These different incidents point to a common set of underlying problems. While organizations and consumers have more tools than ever to protect their data from loss or compromise, improvements in individual defense have been offset by a heightened risk of systemwide events. Many sectors of the global economy now rely on the same set of critical technology products and services, concentrating risk into an unknown number of possible failure points. The potential for catastrophe increases as developing nations further digitize and as activities that were previously separated from the internet—like medical care or transportation—become networked. The worst cyber events can now cause bodily harm or deaths, political crises, and multibillion-dollar economic losses. As digital networks interlink with the physical world in complex, dynamic, and opaque ways, many observers fear new forms of fragility that no one understands.

The dangers come in various forms and are illustrated by an increasing number of large-scale cyber incidents. Before the Log4j crisis, the WannaCry and NotPetya hacks (2017), the Meltdown and Spectre vulnerabilities (disclosed in 2018), and the compromises of SolarWinds (discovered in 2020) and Microsoft Exchange (2021) all demonstrated how a single piece of faulty hardware or software could jeopardize critical systems worldwide. Before the Facebook outage, simple human errors had triggered previous outages of Amazon, Google, and Microsoft cloud services. There have also been physical disruptions, both malicious—like the 2020 Nashville suicide bombing that impaired regional telecommunications—and natural—like Hurricane Maria in 2017, which disrupted internet connectivity in Puerto Rico. So far, these and other high-impact cyber events have proven largely manageable. Nevertheless, they reveal latent risk factors and illuminate some potential triggers and pathways of a future systemic event. Modelers have warned that even more damaging cyber incidents are possible.

Despite rising worries about systemic cyber risk, the problem and potential solutions are poorly understood. “Systemic cyber risk” is a vague concept with no widely accepted definition. Moreover, tools and methodologies for finding and measuring sources of systemic cyber risk remain very limited. Cyberspace is incredibly complex, with billions of devices managed by millions of organizations. It is hard to assemble useful data on so many interdependencies, and models are still too crude to draw confident conclusions from what data does exist. Worst of all, efforts to bound, reduce, or manage systemic risk remain ad hoc and uncoordinated. A problem of this scale and complexity demands much broader and deeper collaboration among industries, across the public-private divide, and internationally.
This paper seeks to provide a common foundation for understanding and addressing systemic cyber risk. Building on prior research, it explores definitions of the problem, underlying contributing factors, and potential policy responses. Although much remains unknown about systemic cyber risk, including its true size and distribution, public and private sector leaders worldwide can and should act now to investigate, reduce, and manage the risk.

What Is Systemic Cyber Risk?

The concept of “systemic cyber risk” is grounded in the broader idea of “systemic risk”—the possibility that a single event or development might trigger widespread failures and negative effects spanning multiple organizations, sectors, or nations. Well-known systemic shocks with global consequences have included the 2008 financial crisis, the COVID-19 pandemic, and recent supply chain disruptions. These events illustrate how various forms of interdependency—whether financial, biological, logistical, digital, or otherwise—can increase volatility in large-scale systems by boosting the chance that any given failure will spread, often in unexpected ways (see figure 1).

Figure 1. Complex Interdependency Can Heighten Systemic Risk

Note: This 2003 chart, created by the Netherlands Organization for Applied Scientific Research (TNO), illustrates how interdependences linking critical infrastructure sectors create opportunities for a single failure to create second and third-order consequences. An updated version would doubtless be even more complex.
Systemic cyber risk is therefore a cousin of other systemic risks. Many cyber analysts recognize this family resemblance and have sought to draw insights from finance, public health, and other disciplines to fill gaps in their understanding of systemic cyber risk. Moreover, a systemic cyber event could actually be a cause or consequence of other types of systemic failure. For example, the COVID-19 crisis has led to remote work patterns in developed countries that have put new stresses on some internet services. Although the internet has proved remarkably resilient to these stresses, this example suggests how biological contagions might help precipitate future failures in cyberspace. Conversely, a coordinated destructive cyber attack on key pharmaceutical companies could potentially reduce the public’s access to and/or confidence in life-saving vaccines, enabling a cyber incident to exacerbate a systemic health crisis.

There is no single, widely accepted definition of systemic cyber risk, and most proposed definitions are vague. This is in part because systemic risk is a fuzzy concept. Because systemic risk can arise and manifest in innumerable ways, some commentators have resorted to saying that “we will know it when we see it.” But definitional challenges are also the result of important differences in perspectives among stakeholder groups and how they specify the relevant system. A system might be a company, a certain computer architecture, an economic sector, an entire nation, or a supranational region. What counts as “systemic” can look very different depending on the system at issue and one’s relationship to that system.

Parsing different definitions of systemic cyber risk is important because meaningful efforts to address the problem will require intensive collaboration across a diverse set of interest groups. No single player has the tools to fully understand, reduce, or manage systemic cyber risk. Rather, key information and policy levers are scattered in many places—including national security agencies, financial institutions, technology providers, cybersecurity firms, critical infrastructure operators, re/insurers, civil society organizations, and regulators—around the world. Building partnerships among these divergent players will require common terms of reference or, at least, mutual clarity on how and why definitions differ.

For example, the U.S. Cybersecurity and Infrastructure Security Agency (CISA) recently launched a Systemic Cyber Risk Reduction Venture. This effort focuses on cyber risks that are “concentrated” enough to pose “critical risks to the Nation’s security and economic security.” In other words, CISA defines the “system” at stake as the United States itself (or U.S. security)—a frame of reference that matches the agency’s core mission. CISA therefore prioritizes cyber risks that threaten its list of fifty-five U.S. “national critical functions,” such as maintaining the U.S. water supply and conducting U.S. elections. In one sense, the agency’s definition could be considered inclusive given the vast range of sectors covered. But in another sense, the definition could be construed as narrow because it focuses primarily on national impact, while ignoring harm suffered by other nations.
The financial system’s approach tends to be more sector-specific and international. The European Systemic Risk Board (ESRB), for example, defines a systemic cyber incident as one occurring “in the financial sector” that potentially threatens “serious negative consequences for the internal market and the real economy.” Unlike CISA, the ESRB is not concerned with disruptions to a U.S. city’s water supply but rather with cyber events that might destabilize international financial markets (by deleting critical transaction data, for example) or disrupt the global economy (perhaps by causing a prolonged outage of international shipping networks). The ESRB approach is grounded in its organizational mission of supervising the EU financial system’s overall stability.

A third approach to systemic cyber risk can be found in the insurance industry, which has experienced rapid growth in the issuance of insurance policies that pay out to victims of cybersecurity incidents. Many industry professionals consider a cyber risk to be systemic if it is uninsurable—based on the scale of potential losses; loss correlation across many clients, sectors, and regions; and the difficulty of modeling and hedging. The reinsurance firm Munich Re cites such factors to explain why a major internet failure would carry “systemic accumulation risk.” A 2017 report by the insurance company AIG defined systemic cyber risks simply as those cyber events “capable of impacting many companies at the same time.”

These are just a few ways to define systemic cyber risk, and they do overlap. An extremely catastrophic cyber event—such as the widespread loss of integrity and trust in critical financial transaction data—would likely qualify as systemic from almost anyone’s perspective. Less severe incidents are likely to elicit different views. For instance, CISA might consider the ransomware attack on the Colonial Pipeline in 2021 a systemic event. Yet the event’s impact on fuel distribution was short-lived, it did not cause financial market instability, and its insurance impacts were limited (because most losses were downstream of the victim and likely uninsured)—meaning that other stakeholders might take a different view.

Given the lack of shared terminology, how should cybersecurity experts classify incidents that have already occurred? Reports from the Swiss Re Institute in 2017 and the EastWest Institute in 2019 asserted that no previous cyber event had qualified as systemic. On the other hand, the risk analytics firm Cyberhedge has described the SolarWinds breach as systemic, and several of the “systemic” scenarios outlined by AIG in 2017 resemble subsequent real-world events, including (brief) outages of major cloud services. These diverging interpretations may complicate efforts to rally stakeholders toward collective action. Those who believe that systemic cyber events are already happening may feel greater urgency to act, while those who believe such events remain hypothetical might want to prioritize more clear and present cyber dangers.
What Does Systemic Cyber Risk Look Like?

Systemic cyber risk comes in multiple forms and can be categorized in different ways. One practical approach is to identify the different kinds of causal sequences that can amplify a trigger event—the spark that lights the tinder—and produce systemic consequences. The three sequences described here show how systemic failures can occur vertically, horizontally, or both.

**Chain Reaction (Vertical Failure)**

A problem affecting a single component of the overall system triggers a chain reaction affecting an ever-expanding range of dependent entities (see figure 2). For example, attackers take down an internet exchange point, disrupting operations for multiple entities that can no longer provide critical services for customers. The customers must then halt service for their own customers, and so on.

**Contagion (Horizontal Failure)**

Identical copies of a system component fail simultaneously or in rapid succession (see figure 3). One failure creates systemic effects without a chain reaction. For example, a malicious exploit affecting all copies of a software product allows attackers to disrupt core operations for all users of that product.
Another way to categorize systemic cyber risk is by understanding the different roles that cyberspace may play in a systemic event. Systemic cyber events can have cyber triggers, cyber pathways, cyber effects, or all of the above. A cyber trigger is an adverse event originating within cyberspace, such as a malicious intrusion into a computer network. A cyber pathway is the digital infrastructure (such as ubiquitous network protocols) through which an initial trigger propagates and is amplified. Cyber effects are harms experienced within cyberspace, such as the non-availability of a computer system.

These three variables can combine in many different ways and interact with noncyber factors. For example, the Colonial Pipeline incident was caused by ransomware (cyber trigger), which led the company to shut down pipelines for the United States’ East Coast fuel system (noncyber pathway), and ultimately disrupted surface transportation (noncyber effect). Conversely, a missile strike (noncyber trigger) that destroys a data center (cyber pathway) may cause data losses (cyber effect) as well as work stoppages at organizations that relied on the deleted data (noncyber effect).

Although observers disagree about which real-world cyber incidents (if any) qualify as systemic, a number of large-scale cyber events have illustrated the potential triggers and pathways. Most attention has focused on malicious hacks, malware incidents, and exploit-able vulnerabilities, including the following:

- **Morris Worm (1988):** An accidental software bug caused a computer worm to spread rapidly, crippling tens of thousands of computers and forcing important institutions to disconnect from the internet.\(^{21}\)

- **Internet Root Server Attack (2002):** A coordinated attack that disabled nine of thirteen root servers running the internet’s Domain Name System. One expert believes “it [was] clearly done with the intent to cripple or shutdown the Internet.”\(^{22}\)
• **SQL Slammer (2003):** A computer worm rapidly infected tens of thousands of machines and disrupted access to most of the world’s data query servers and networks.\(^{23}\)

• **Dyn (2016):** Criminals in control of the Mirai botnet launched a distributed denial-of-service attack against an important piece of the Domain Name System, blocking access to major websites including Twitter and PayPal.\(^{24}\)

• **WannaCry and NotPetya (2017):** The WannaCry attackers exploited a Windows vulnerability to launch global ransomware that compromised over 230,000 computer systems within a day.\(^{25}\) The NotPetya wiper, masquerading as ransomware, used a flaw in Ukrainian tax preparation software to spread the attack among international corporations, causing an estimated $10 billion in damages.\(^{26}\)

• **Meltdown and Spectre (2018):** Researchers disclosed critical vulnerabilities in common chips, including those used in the servers of leading cloud service providers. If exploited, the flaws would have allowed attackers to steal data from countless corporate customers.\(^{27}\)

• **VxWorks (2019):** In 2019, researchers revealed previously undisclosed security vulnerabilities inside an operating system called VxWorks, used in settings like aviation and industrial automation where physical safety is essential.\(^{28}\) This single dependency, obscured in the supply chain, imparts remotely triggerable failure modes to over 200 million devices across all critical infrastructure sectors.\(^{29}\)

• **Kaseya (2021):** Using a zero-day vulnerability (unknown to the vendor) in a popular IT management tool, a criminal ransomware group infected between 800 and 1,500 businesses in one attack.\(^{30}\)

While malicious cyber exploitation causes unease and grabs headlines, systemic cyber events can have many other triggers, both intentional and unintentional. The 2020 Nashville suicide bombing caused days-long regional outages of AT&T’s telecommunications networks. And in 2013, in California, suspected insiders cut fiber optic cables and disabled a power substation with gunfire, illustrating the dangers of physical attacks.\(^{31}\) Human errors have reportedly contributed to outages of Amazon Web Services (2017), Google Cloud (2018), and Microsoft Azure (2019) services.\(^{32}\) Natural disasters can also cause large-scale cyber effects. Lightning strikes in 2018 affected an Azure data center in Texas, triggering a cascading series of failures in the region and beyond. Hurricane Maria in 2017 led to internet and phone problems throughout Puerto Rico that persisted for more than a year.\(^{33}\)

The consequences of these and other noteworthy incidents were ultimately contained. No single cyber event has come close to rivalling the worst noncyber systemic events, such as the 2008 financial crisis or the COVID-19 pandemic. Nevertheless, many large-scale cyber incidents could have caused greater damage than they did—if a hostile perpetrator desired it, if fail-safes had malfunctioned, and so on.\(^{34}\) Malicious cyber operations are particularly concerning because while most adversaries have not yet demonstrated an intent to cause
systemic-level harms, their motivations could shift quickly for a variety of reasons. Moreover, if systemic vulnerabilities accumulate, it will become easier to cause a systemic cyber event, and the field of actors capable of doing so will grow. While nation-states have long been viewed as the primary actors capable of triggering systemic shocks via cyberspace, more criminals, extremist groups, and other mischief-makers may soon take a seat at the table.

In sum, recent events offer ample warning that systemic cyber risk is a real threat. Although the worst-case scenarios remain hypothetical, the list of large-scale cyber events continues to grow, with many of biggest incidents happening in just the last few years. These incidents reveal the diversity of forms that systemic cyber risk can take—a broad and expanding set of potential triggers, pathways, and effects. If present trends continue, then major cyber incidents will probably become ever-larger and more damaging in the coming years. And, eventually, a bona fide national (or global) catastrophe will occur in cyberspace. It is therefore urgent that policymakers understand what factors underlie these trends and what they can do to address them.

What Can Cause (or Mitigate) Systemic Cyber Risk?

Given its interconnected structure, many aspects of cyberspace naturally amplify systemic risk. Standardized digital technologies—including common protocols—create a highly interdependent web of technical and organizational linkages. At least five properties of cyberspace help to magnify systemic risk.

Risk Concentration

Many aspects of the digital economy have consolidated around common technologies, software products, and third-party service providers. This offers greater efficiency, capability, and agility for some organizations. But widespread dependence on fewer, more specialized services and products creates common vulnerabilities, data bottlenecks, and single points of system failure. This concentrates risk by magnifying the negative consequences of individual failures.

Complexity

Computer networks and the operational and human systems that interface with them are increasingly complex, which “ensures hidden levels of mutual dependence.” The internet’s shifting web of technical, contractual, and financial linkages prevents organizations from fully understanding which system components support which core functions for which organizations. Single points of failure can easily escape scrutiny.
Opacity

Many of the most important features of digital technology are intangible and proprietary in nature. This means that understanding systemic relationships requires specialized knowledge of and access to sensitive parts of privately operated technology systems. Yet companies often have commercial, legal, and other motivations to withhold such information from outsiders.

Scale

Cyberspace weaves together millions of vulnerable machines that often share identical, remotely exploitable vulnerabilities. This creates low marginal costs for high-impact attacks, incentivizing malicious operations designed to ensnare the maximum number of victims. It also increases the likelihood of a systemic incident even if an adversary does not intend to cause one.

Intelligent Opponents

The presence of intelligent adversaries greatly complicates the task of understanding and managing systemic cyber risk. Malicious humans act with purpose, patience, and technical savvy to tilt the odds, finding latent concentration risks or burrowing their way into the best defended organizations to trigger a chain reaction of failures. Patient opponents can sometimes turn an otherwise improbable sequence of events into a virtual certainty.

These cyber-specific properties are not the only factors amplifying systemic cyber risk. Fragility in real-world systems can set the conditions for an initial cyber trigger to cause widespread harm. For example, just-in-time shipping makes it more likely that a ransomware attack on an important port would cause logistical bottlenecks across multiple sectors. Excessive financial leverage increases the odds that an economically damaging cyber attack would prompt bankruptcies or market turbulence. Weak governance in a country could lead to a situation where localized infrastructure outages caused by a cyber intrusion ignite large-scale political unrest and social crises. These examples illustrate why studies of systemic cyber risk should look well beyond traditional cybersecurity concerns and seek to draw on a wide range of disciplines and data sources.

While multiple studies have examined the factors driving systemic cyber risk, there has been less analysis of possible countervailing trends and mitigating factors. For example, the continued growth of digital interconnection may be offset by a global rise of techno-nationalism and cyber sovereignty. The U.S. and Chinese technology sectors in particular are undergoing a partial “decoupling” that will affect many other countries and private actors around the world. Moreover, some major tech companies continue to invest in “walled garden” technology ecosystems with intentional limits on the available interconnections to
competitors’ platforms. Conceivably, a more fragmented digital world based on geographic and technological boundaries could limit the impact of some cyber incidents. Alternatively, balkanization might encourage state actors to carry out even more aggressive cyber attacks against their adversaries, because the attackers’ own digital sphere of influence would be relatively insulated from direct collateral damage.

Proactive steps to achieve resilience can reduce the likelihood and severity of systemic cyber events. However, much more study is needed to understand and measure various forms of resilience. Consider the 2017 infection of Maersk, the global shipping giant, by the destructive NotPetya malware. Resilience in that situation would have had many different elements. First, were Maersk’s backup systems and recovery plans in place (cyber resilience)? Second, did the company’s logistics networks have enough slack to absorb a cyber-triggered disruption (operational resilience)? Third, did Maersk’s competitors have the excess capacity and agility to fill in while the company recovered (sector-wide resilience)? Fourth, could Maersk’s customers tolerate delays in sending and receiving shipments (customer resilience)? As this example shows, systemic resilience can be just as complex as systemic risk. It merits more attention from researchers and policymakers.

### How Can Sources of Systemic Risk Be Identified?

Effective risk management requires some capacity to quantify the probability and severity of harmful events and locate specific trouble spots where interventions can best be applied. This is a major challenge with any complex system, and it is acute in the context of systemic cyber risk. Large-scale digital networks such as the internet change constantly and behave in unpredictable ways. Without better data and risk models, any attempts to map the many possible digital origins, pathways, and effects of a systemic shock would be incomplete at best and quickly become outdated.

Data collection faces major obstacles. Many organizations do not know their first-order cyber dependencies. Some struggle to implement even basic steps, such as maintaining software and hardware inventories. Thus, surveying individual organizations will not yield a complete picture of systemic-level points of failure. For example, some insurers have sought to identify portfolio-wide cyber risks by asking policyholders to report their most important software applications and services. But these reports, when compared to the results of external network scans, reveal many omissions. Additionally, organizations often do not want to disclose sensitive information about their dependencies to outside entities such as insurers or government bodies. They may fear losing competitive advantages, inviting regulatory and legal scrutiny, or providing a road map for hostile actors to exploit. Critical technology providers, in particular, often keep their technical architectures a closely guarded secret.
Moreover, collecting data on cyber dependencies is only useful when the right analytic methodologies and tools are used to process the data and predict how trigger events might propagate and cause systemic harms. Risk models have been developed for systemic cyber risk, but their usefulness is often hampered by a lack of transparency (the underlying methodologies remain secret) and consistency (different models cannot readily be compared on a like-for-like basis). Commercial or security interests can therefore limit the ability of a user to interpret why a model has generated a particular result.

More fundamentally, there are no historical case studies to help refine the modeling tools, because many of the most concerning systemic cyber scenarios have never occurred. Today’s models must therefore depend heavily on assumptions. For example, in 2021, RAND published “a quantitative model of cascading failures to estimate the potential economic damage associated with a given cyber incident.” Applying this model to Maersk’s NotPetya infection, RAND found the total economic cost may have been as little as $3 billion or as much as $57 billion. The dramatic range underscores the uncertainties involved in attempting to anticipate the effects and losses of highly consequential cyber events.

Thankfully, some concentration risks are fairly evident. It is widely recognized that the Domain Name System, the Border Gateway Protocol, GPS-based position/navigation/timing, and the Linux kernel are critical to the stability of the digital world. But there may be many other latent concentration risks, and some could be more dangerous. SolarWinds was a “little-known” and “obscure” company before it was hacked by Russian operatives. Many SolarWinds customers were users without even knowing it.

The difficulty of mapping and measuring systemic cyber risk presents at least two policy problems. First, without the ability to understand and communicate the probability and severity of systemic cyber events, decisionmakers cannot determine whether resources currently devoted to other problems (including other forms of cyber risk) should be redirected toward steps to mitigate systemic cyber risk. Second, without a clear picture of where the problem areas lie and which possible failure points deserve the most attention, policymakers cannot determine precisely which resources should be redirected and how.

Many observers believe that systemic cyber risk is a large and growing challenge and that such risks are distributed throughout many sectors of the economy and society, with innumerable triggers and pathways. Popular acceptance of these beliefs has dampened risk appetites among cyber re/insurers and regulators and has sparked some calls for large-scale governmental interventions to address systemic cyber risk. Meanwhile, although a minority, other observers argue that systemic cyber risk is smaller and less widely distributed than commonly believed. This camp believes that digital and operational systems have so far proven tolerably resilient to cyber events; insurance markets should assume more cyber exposure; and collective action should focus on a handful of sectors, technologies, and circumstances where systemic cyber risk concentrates. Definitively resolving this argument will require better data and methodologies than currently exist. But even as experts work to better understand and quantify systemic cyber risk, it is already clear that policymakers must do more to tackle the problem.
How Can PolicymakersRespond?

Generally speaking, addressing systemic cyber risk involves three types of interventions. First, policymakers can try to enhance the ability to identify and measure such risks. Second, they can seek to bound, reduce, or even eliminate some systemic cyber risks. And third, they can work to better manage whatever systemic cyber risk cannot be eliminated.

Given the breadth and complexity of the underlying problem, effective policy should involve a diversity of stakeholders: private actors (such as technology providers, cybersecurity firms, critical infrastructure operators, and re/insurers) as well as public actors (including regulators and national security agencies). International cooperation is also essential because systemic cyber risk is inherently global and systemic shocks are not bound by national borders.

Below are just a few of many initial ideas that merit further exploration.

Identify Unrecognized Software Dependencies

Some sources of risk concentration can be readily identified. For example, the U.S. National Institute of Standards and Technology (NIST) recently published a list of “critical software” categories, such as operating systems and web browsers. The categories were identified by their elevated privileges and trusted functionality. However, the VxWorks case indicates that organizations may depend on software that is not apparent to risk managers. Efforts to achieve a comprehensive understanding of software supply chains are therefore essential. To that end, transparency into the provenance of software offers a foundation for building a system-wide view of specific software dependencies.

One mechanism for promoting transparency could be a software bill of materials, intended to enable vendors to communicate the contents of their software products. In the United States, NIST has sketched the outlines of what a Software Bill of Materials must contain, pursuant to President Joe Biden’s Executive Order 14028 on “Improving the Nation’s Cybersecurity.” The Linux Foundation’s Core Infrastructure Initiative is another attempt to identify which elements of open source code are systemically critical. Google, too, has a new tool to reveal dependencies for open-source projects. There is significant additional space for industry to innovate in this area.

Promote Diversity of Products and Services in Narrow Segments

Dispersing digital dependence across multiple software products, computing hardware, or internet infrastructure services could “increase resiliency and prevent cascading failures” resulting from a single malicious compromise. Efforts by CISA to shift businesses away from total reliance on GPS signals for position, navigation, and timing data offer one
Governments can use tools such as grants, tax incentives, and antitrust policy to increase competition in key technology markets. Major investors could set up funds focused on nurturing new entrants in highly concentrated, critical IT segments. If successful, these newcomers would both generate profits for the investors and help to reduce systemic cyber risk across the rest of their portfolios.

Diversification should be pursued thoughtfully to manage potential tensions with other aspects of cybersecurity. For example, introducing alternative new IT products and services could be harmful in the short run if they happen to be less secure than the previously dominant technology. Diversification can also sometimes increase system-wide complexity and costs, making it harder to monitor and manage a system’s operations and risk profile. Efforts to diversify IT should go hand in hand with efforts to better manage network complexity and cost.

**Identify Systemically Important Digital Entities**

Following the 2008 global financial crisis, a number of countries and international bodies sought to better identify, monitor, and regulate “systemically important” financial institutions whose failures could have cascading consequences. The Financial Stability Board, which includes the G20 and several other countries, now works with other international and national regulators to designate global banks and insurers as systemically important. National legislation such as the U.S. Dodd-Frank Act complements this international scheme by creating additional designation authorities and imposing tailored regulation on designated entities.

A similar scheme could be developed for systemically important digital entities—those most at risk of triggering, propagating, or suffering the effects of systemic cyber events. Policymakers should consider several questions in weighing and designing such a scheme. First, what consequences would a designation have? Designated entities would presumably need to meet heightened resilience and transparency requirements. They might also need increased governmental support, such as intelligence collaboration. Second, what would be the designation criteria? While banks have well-understood and readily measured markers of systemic importance, such as total assets and leverage ratios, digital entities would need to somehow be evaluated based on their networked interconnections. Third, who would make the designations? Global bodies such as the International Telecommunication Union are one option, though their decisionmaking can be distorted by geopolitics. Finally, how can a designation scheme overcome political resistance and inertia? Tech companies may lobby against what they see as new regulatory burdens, and there has not yet been a cyber event catastrophic enough to galvanize significant global action.

While designing and implementing such a scheme would be challenging, some early efforts are already underway. In the U.S. House of Representatives, for example, a bipartisan bill introduced in 2021 would empower CISA to designate a subset of critical infrastructure (both digital and nondigital) as systemically important.
Pursue Scale in Hardening Elements of Non-substitutable Digital Infrastructure

Some sources of risk concentration are here to stay. It is probably impractical to reduce widespread dependence on some building blocks of cyberspace, such as popular cloud platforms, key open-source software, and internet infrastructure. These elements of digital infrastructure offer an attractive target for steering limited resources to harden systems against disruption and to shrink the pool of adversaries (and nonmalicious triggers) that might be capable of compromising them.

One approach would be to convene international, multistakeholder working groups, perhaps under the aegis of existing industry or international forums or trusted nongovernmental organizations. Each working group could focus on a different source of concentrated risk. For example, a cloud working group could vet and disseminate best practices, such as the use of mathematical techniques to find and improve the security of software components that are common to many cloud environments. An open-source working group could mobilize actors to more consistently enforce licensing agreements and to broaden successful efforts to reduce known vulnerabilities—such as the Internet Security Research Group’s effort to translate Linux (the internet’s most important operating system) into a more secure language. A public core working group could address recurrent problems in systems like the Border Gateway Protocol or the Domain Name System. By bringing together disparate stakeholders to address discrete areas of systemic cyber risk, these working groups could help to spur collaborative relationships and create processes that could be assessed, refined, and built upon over time.

Scale Response Capacity for the Have-Nots

An important element of resilience is a ready-made capability to respond and recover quickly from a given incident. But the incident response industry is already struggling to keep pace with routine intrusions. In the case of a horizontal or contagion systemic failure, this could leave most victims without timely access to response and recovery capabilities. Countries need the ability to quickly scale the provision of technical assistance to a wide range of actors.

There is no quick or easy solution to this problem, but several promising models merit expansion, at least to test their ability to operate at scale. In the United States, for example, there are National Guard cyber units as well as state-level efforts like the Michigan Cyber Civilian Corps or Wisconsin’s Cyber Disruption Response teams. Connecting local governments and business groups directly with communities of white-hat (or ethical) hackers could offer another way to deliver immediate emergency assistance when other avenues are unavailable.
**Enhance the Role of Insurance**

Any events that generate massive correlated losses at a global scale—especially business interruptions, a major worry in systemic cyber events—will exceed the response capacity of the private insurance market. However, it should be possible to design public-private partnerships in which private insurers cover some elements of systemic cyber risk while governments step in after a certain threshold. Such an arrangement would have many benefits, including drawing more private capital into the marketplace and creating an orderly, pre-planned mechanism to compensate victims.\(^6\)

Public-private partnerships have demonstrated their potential in other challenging insurance areas, including extreme weather and natural catastrophe, nuclear energy, trade credit, and terrorism. These partnerships can offer various forms of support, including risk pooling and backstops. In the United States, the Cyberspace Solarium Commission recommended looking at the need for a federal backstop for losses from catastrophic cyber events. The European Insurance and Occupational Pensions Authority has likewise noted that the “threat of systemic risk events coming from cyber incidents might require responses from both the government and the industry to provide adequate insurance capacity in support of the real economy.”\(^6\)

**Conclusion**

The growing awareness of systemic cyber risk is revealing two central challenges: complex systems are difficult to understand, and coordinated action is hard to incentivize at the system level. While these problems are not confined to cyber risk, the ever-expanding capabilities of digital technology, coupled with deepening dependence throughout society, mean that systemic cyber risk has the potential to spread harm with a unique combination of speed, scale, and uncertainty.

While systemic cyber risk has become a hot topic, it deserves even more (and closer) attention than it has so far received. Many key questions remain unanswered. Exactly how serious and widespread is systemic cyber risk? What are the most important aggravating and mitigating trends? Where can we find better definitions, data, and models to identify systemic cyber risk? And crucially, how can disparate private and public stakeholders around the world come to together to tackle this global problem? It may take years to arrive at satisfactory answers. But now is the time to pose these questions and begin taking tangible action—before a truly catastrophic cyber event occurs.
About the Authors

David Forscey is a former managing director of the Aspen Cybersecurity Group. He previously worked on state cybersecurity strategy and policy at the National Governors Association Center for Best Practices. He was also a national security fellow at Third Way and a cybersecurity fellow at New America. He is a graduate of the University of Virginia and Georgetown University Law Center. He currently serves in the Joint Cyber Defense Collaborative at the Cybersecurity and Infrastructure Security Agency.

Jon Bateman is a fellow in the Technology and International Affairs Program at the Carnegie Endowment for International Peace. He previously worked as a senior intelligence analyst, policy adviser, and speechwriter in the U.S. Defense Department, most recently serving as special assistant to the Chairman of the Joint Chiefs of Staff.

Nick Beecroft is a nonresident scholar in the Technology and International Affairs Program at the Carnegie Endowment for International Peace. His background is in the insurance industry, where he focused on the development of cyber insurance, and UK military intelligence.

Beau Woods has more than twenty years of experience advising on intersectional risk management, cybersecurity, and strategy through his various roles with the Atlantic Council, the I Am The Cavalry initiative, and Stratigos Security, and as a nonprofit board member.

The views expressed in this paper do not represent the position of any U.S. government agency or department.
Acknowledgments

The authors wish to thank Marjory Blumenthal, Tracie Grella, Jay Healey, Trey Herr, Ariel (Eli) Levite, George Perkovich, Milena Rodban, and Jonathan Welburn for their valuable feedback. The authors are solely responsible for the final paper, and the views expressed here do not represent the position of any U.S. government agency or department.
Notes


39 Tyson Macaulay has said that “governments have an urgent need to achieve a clearer understanding of the often-opaque interdependencies between critical infrastructure sectors, and to take steps to mitigate the chances of cascading chain reactions”; see “The Danger of Critical Infrastructure Interdependency,”
Centre for International Governance Innovation, 2019, https://www.cigionline.org/articles/danger-critical-infrastructure-interdependency/. Dave Clemente has noted that “understanding and managing the risk that arises from these dependencies is rarely straightforward or transparent”; see “Cyber Security and Global Interdependence: What Is Critical?,” Chatham House, February 2013, https://www.chathamhouse.org/sites/default/files/public/Research/International%20Security/0213pr_cyber.pdf. In 2018, AIR, now part of the analytics company Verisk, has noted that the “sources of systemic cyber risk are many—similar to land mines in a minefield—and the insurance industry's inability to navigate this minefield and quantify this risk is limiting the growth of this line of business”; see “Rethinking Systemic Cyber Risk—An Approach for Growth,” 2018, https://www.air-worldwide.com/sites/default/files/Publications/Brochures/documents/issue-brief-rethinking-systemic-cyber-risk-an-approach-for-growth. The EastWest Institute has observed that to “deal with this inherent challenge, many insurers and reinsurers are looking for innovative ways to improve their underwriting and risk modeling methodologies”; see “Cyber Insurance and Systemic Market Risk,” 30. Months after the VxWorks disclosures described, the same vulnerabilities were discovered in five additional products used in yet more critical infrastructure systems—much to the surprise of the manufacturers and customers; see Vijayan, “Millions More Embedded Devices Contain Vulnerable IPnet Software.”


44 Ibid.


Another example of a concentrated risk emanating from open-source software is the Network Time Protocol, used by billions of digital devices to synchronize time—and which was overseen by one person as of 2016; see Babcock, “Harlan Stenn Tackles NTP Security Issues as Big Move Looms.”


Even short of a global internet calamity, targeted BGP attacks present a source of concentrated risk because they enable attacks on a variety of vital network services, including anonymity systems like Tor, public key infrastructure (for example, encrypted communications), and cryptocurrency infrastructure. See Yixin Sun et al., “Routing Attacks on Internet Services,” Freedom to Tinker, April 11, 2018, https://freedom-to-tinker.com/2018/04/11/routing-attacks-on-internet-services/.


Carnegie Endowment for International Peace

The Carnegie Endowment for International Peace is a unique global network of policy research centers in Russia, China, Europe, the Middle East, India, and the United States. Our mission, dating back more than a century, is to advance peace through analysis and development of fresh policy ideas and direct engagement and collaboration with decision-makers in government, business, and civil society. Working together, our centers bring the inestimable benefit of multiple national viewpoints to bilateral, regional, and global issues.

Technology and International Affairs Program

The Carnegie Technology and International Affairs Program (TIA) helps governments and industries reduce large-scale international risks of new technologies and related services. Recognizing that commercial actors control many of the most germane technologies, TIA identifies best practices and incentives that can motivate industry stakeholders to pursue growth by enhancing rather than undermining international relations.

TIA’s work informs and is informed by direct dialogues among thought-leaders, senior officials, and executives in key countries. We share the data, insights, and policy recommendations that result in reports, commentaries, and web tools. Carnegie’s regional centers and networks in the United States, China, Europe, India, and Russia provide a widely respected international platform for promoting our policy proposals.
Aspen Institute

The Aspen Institute is a global nonprofit organization committed to realizing a free, just, and equitable society. Founded in 1949, the Institute drives change through dialogue, leadership, and action to help solve the most important challenges facing the United States and the world. Headquartered in Washington, DC, the Institute has a campus in Aspen, Colorado, and an international network of partners.

Aspen Digital Program

Aspen Digital empowers policy-makers, civic organizations, companies, and the public to be responsible stewards of technology and media in the service of an informed, just, and equitable world. This Aspen Institute program shines a light on urgent global issues across cybersecurity, the information ecosystem, emerging technology, the industry talent pipeline, tech and communications policy, and innovation. It then turns ideas to action and develops human solutions to these digital challenges.