Federal Reserve Bank of New York
Staff Reports

Cyber Risk and the U.S. Financial System: A Pre-Mortem Analysis

Thomas M. Eisenbach
Anna Kovner
Michael Junho Lee

Staff Report No. 909
January 2020
Revised June 2020

Cyber Risk and the U.S. Financial System: A Pre-Mortem Analysis

Thomas M. Eisenbach, Anna Kovner, and Michael Junho Lee

Federal Reserve Bank of New York Staff Reports, no. 909
January 2020; revised June 2020
JEL classification: G12, G21, G28

Abstract

We model how a cyber attack may be amplified through the U.S. financial system, focusing on the wholesale payments network. We estimate that the impairment of any of the five most active U.S. banks will result in significant spillovers to other banks, with 38 percent of the network affected on average. The impact varies and can be larger on particular days and in geographies with concentrated banking markets. When banks respond to uncertainty by liquidity hoarding, the potential impact in forgone payment activity is dramatic, reaching more than 2.5 times daily GDP. In a reverse stress test, interruptions originating from banks with less than $10 billion in assets are sufficient to impair a significant amount of the system. Additional risk emerges from third-party providers, which connect otherwise unrelated banks.

Key words: cyber, banks, networks, payments

Eisenbach, Kovner, Lee: Federal Reserve Bank of New York (emails: thomas.eisenbach@ny.frb.org, anna.kovner@ny.frb.org, michael.j.lee@ny.frb.org). The authors thank Danny Brando, Darrell Duffie, Beverly Hirtle, João Santos, and Jason Tarnowski for comments. They also thank Aaron Plesset, Helene Hall, and Montgomery Fischer for outstanding research assistance.

1 Introduction

Cyber attacks are an increasing concern especially for financial service firms which may experience up to 300 times more cyber attacks per year than other firms (Boston Consulting Group, 2019). Almost every financial stability survey includes cyber attacks among the top risks.[1] Yet, there is still no universal consensus on the taxonomy and definition of cyber events, let alone comprehensive data collection on the frequency and nature of cyber attacks. In this paper, we seek to understand the risk presented by cyber attacks to the U.S. financial system and to quantify how a cyber attack may be amplified through the system.

We begin by integrating cyber risk into the theoretical literature on bank runs, highlighting similarities and differences between cyber and traditional shocks. In some ways, losses related to cyber attacks are similar to other operational loss events that can trigger liquidity runs and lead to solvency issues. Such runs can be panic based (Diamond and Dybvig, 1983) or fundamentals based (Goldstein and Pauzner, 2005). As in a standard run, a cyber related run would likely result in direct costs to the affected bank and spillovers to counterparties within the financial sector and to the real economy.

The classification used by the cyber community suggests a taxonomy of how cyber attacks may differ from traditional operational risks (Curti et al., 2019). In a cyber attack, confidentiality of data may be compromised, as may be the availability and integrity of data or systems. Further, a cyber event may be a deliberate attack to damage the financial system. Technological linkages through which cyber attacks can spread are likely to be different from solvency and linkages arising from business interactions. Finally, cyber attacks are likely to be associated with significant uncertainty.

When a cyber attack compromises the availability or integrity of an institution's system and/or data — unlike in a traditional run — cyber attacks may impair the ability of the bank to service running creditors. In such a scenario, the classic first-mover advantage would be weakened, potentially limiting the incentive to run. On the other hand, uncertainty regarding the nature and extent of the attack could prompt runs to occur in segments of banks' operations that are otherwise unaffected.

We use data on payment activity in Fedwire Funds which represents the majority of wholesale payments between financial institutions in the U.S. to evaluate how a cyber

[1]Survey results come from both industry and regulatory reports. To name a few, see the DTCC's 2018 Systemic Risk Barometer Survey (http://www.dtcc.com/systemic-risk/systemic-risk-barometer-surveys), the Bank of England's Systemic Risk Survey (http://www.bankofengland.co.uk/systemic-risk-survey/2019/2019-h1), and the Bank of Canada's Financial System Survey (http://www.bankofcanada.ca/2019/05/fss-highlights).

attack could be amplified throughout the financial system. In general, rare events are difficult to analyze, as we have seen with financial crises (Gorton, 2014) and macroeconomic disasters (Barro and Ursúa, 2012). Rather than a post-mortem analysis of a crippling cyber attack, we instead conduct a pre-mortem analysis to uncover ways that attacks may be amplified into a disaster. By doing so, we hope to inform the ex ante response by mapping out how, in a variety of plausible scenarios, an attack on a single institution or collection of related institutions could be transmitted throughout the financial system. This analysis informs our understanding of transmission mechanisms and the evaluation of vulnerabilities in the financial system, as well as highlights potential policy responses that would be effective to mitigate amplification. In addition, we reverse the analysis to understand different channels through which the system could become impaired.

The wholesale payment network has several characteristics that make it an attractive environment to study. Wholesale payments are one of the most critical networks for resiliency. Indeed, high-value payment and settlement systems may be a natural candidate for a malicious attacker intent on inflicting the largest possible damage to the financial system and the broader economy. From a practical standpoint, data on wholesale payments provides detailed information on the accounts and flows between a diverse set of financial institutions. This enables us to simulate a broad set of scenarios using actual data and examine the result on the network without relying on too many assumptions. This exercise is similar in spirit to that of Duffie and Younger (2019), but calibrated in the post-Dodd Frank Act environment of ample reserves and heightened liquidity and capital regulation.

We begin by establishing a framework to estimate cyber vulnerability, making simple assumptions about an attack and bank responses, and tracing out the impact on payments and liquidity of other banks in the network. Our framework assumes that an attacked institution receives payments, but is unable to send payments for a single full day. By accumulating payments from its counterparties, the attacked institution soaks up liquidity, effectively acting as a liquidity black hole. We estimate how and when the lack of payments leads to liquidity dislocation within the network, and estimate the impact on the liquidity position of other banks that fail to receive large quantities of payments from the attacked institution. A bank is considered *impaired* if its end of day reserves fall sufficiently below the bank's past daily reserve average. We design the framework to be flexible in order to explore various scenarios and variations on assumptions, which illustrate how results may be different depending on the impact of an attack, the responsiveness of banks and the responsiveness of policy.

In our baseline scenario, we begin by analyzing the impact of a cyber attack on the largest five participants in the network. We estimate that, on average, if any of these large

banks stops making payments, 6 percent of institutions breach their end-of-day reserves threshold. While the number of affected banks is relatively small relative to the number of banks in the network, the affected banks are often very large. Weighting the impact of a cyber attack by the assets of affected banks dramatically increases the average spillover impact to about 38 percent of bank assets (excluding the attacked institution). This reflects the high concentration of payments between large institutions, and the large liquidity imbalances that follow if even one large institution fails to remit payments to its counterparties.

Disruptions can be even larger within a local area when the impairment of banks impacts liquidity at local bank branches. For instance, almost 10 percent of U.S. metropolitan statistical areas (MSAs) would experience a severe disruption (more than 40 percent of MSA deposits at impaired banks) if one of the largest banks are successfully attacked. Markets with less bank competition are inherently more vulnerable to disruptions, as a direct impact on a large bank is likely to directly impair a large share of local deposits. Additional fragility in a geography arises when a large share of local deposits are held by banks which are more connected in the payment system.

One distinguishing feature of cyber attacks is that they may be designed for maximum disruption. The extent to which an attacker is informed with respect to the payment system, the targeted institution, and its relation to the payment network, may dictate the magnitude of systemic risk arising in an attack. For example, past studies highlight that total payment activity is often heightened at predictable, regular days over the course of the year. Attacks on seasonal days associated with greater payment activity are more disruptive relative to a typical day, with average impacts that are about 11 percent greater. Similarly, we find that a cyber attacker with specific knowledge of a targeted institution that targets the attack to a particular date could produce impairments that are another 14 percent larger than seasonal day disruptions. We estimate that, on average, attacking on the worst date for a particular large institution adds an additional 38 percent in impairment relative to the case of no specific knowledge.

This initial analysis assumes no response by the banks indirectly affected by the attack. An added impact arises if banks respond strategically to improve their liquidity positions by proactively forgoing payments and hoarding liquidity (a cascade scenario). Theoretically, the effect of liquidity hoarding is unclear depending on payment patterns, as hoarding can mitigate disruption by bypassing affected banks or amplify disruption by creating new impairments. We find that liquidity hoarding amplifies the network impact of the cyber attack, both increasing the average impact on the system and increasing the maximal risk. Under the cascade scenario, the average amplification impact of banks with impaired

liquidity is relatively modest, about 1 percent larger. While this is modest, a large quantity of payments fail to settle as a consequence of banks' strategic hoarding behavior. Excluding the attacked institution, banks' endogenously foregone payments in our one-day scenario are close to average one-day U.S. GDP. Even without accounting for additional macro spillovers such as firms or consumers forgoing spending, a cyber event can therefore have a sizable impact on economic activity.

Traditional interconnectedness research examines the network impact of links formed by banks through interbank borrowing and other counterparty relationships. Yet cyber vulnerabilities may reveal themselves through connections that are not actively captured through traditional counterparty risk management. We explore such vulnerabilities by making use of regulatory data on client lists of significant technology service providers, with a scenario where all bank clients of a service provider are shocked at the same time. For a provider with multiple large and medium size bank clients, an average of 60 percent of banks by assets become impaired. In addition to highlighting the direct impact that a cyber attack may have on its clients, this exercise reveals the importance of understanding operational linkages presented by third party service providers.

Finally, we do a reverse stress test exercise. Since we observe that any one of the five most active institutions can pose systemic risk, we ask how many smaller institutions it would take to impair any of the most active ones. For 9 percent of days, we an identify a subset of small domestic banks (under $10 billion in assets) that can impair at least one of the largest institutions. For larger domestic banks (under $50 billion), such a subset exists on 40 percent of days. Furthermore, we find that including similarly sized branches of foreign bank organizations (FBOs) dramatically increases the potential of impairment. After including FBO branches, we find that the largest institutions may be impaired due to a successful cyber attack on small institutions (under $10 billion) on 72 percent of days. Not only that, the number of institutions required is surprisingly small when including FBO branches. We estimate that, on average, it would take a successful attack on just six banks below $10 billion or just one institution between $10 and $50 billion. This exercise highlights how the vulnerability of many smaller institutions or foreign banks also presents systemic risk.

The paper proceeds as follows. We begin by providing a discussion on cyber risk in the context of the theoretical literature on financial fragility in Section 2. In Section 3, we outline the data and a framework for testing cyber vulnerability. Section 4 summarizes our baseline scenario and main analysis. In Section 5, we augment the baseline scenario by introducing strategic liquidity hoarding behavior of banks. In Section 6, we analyze the systemic implications of cyber risk originating from other non-bank sources, as well as

results from a reverse stress test. Section 7 concludes.

2 Theory on Cyber Risk and Threats to Financial Stability

We begin by integrating cyber risk into the broader theoretical literature on bank runs, highlighting similarities and differences between cyber and other shocks modeled in the theoretical literature.

2.1 Classification of Cyber Events

Cyber events are conventionally classified into three categories based on the technical nature of the event, introduced in the Federal Information Security Management Act. However, the classification is also useful for thinking about the economic consequences of a cyber event.

Confidentiality. The confidentiality of data is compromised; for example, customer social security numbers or proprietary trading records are hacked and publicized. Such events cause direct losses to the affected bank – for example, the compromise of customer social security numbers would incur a loss due to liability for damages while an attack on proprietary trading records might incur a loss due to competitors learning about the bank's strategies. In addition, as for all types of confidentiality events, the bank would suffer reputational costs.

Availability. The availability of data or systems is compromised; for example, a bank's computer systems are shut down. Similar to confidentiality events, availability events can cause direct losses and reputational costs. In addition, availability events can immobilize capital and liquidity, and affect the ability of the bank to perform its core activities. Such events can therefore have considerable spillovers to the banks customers and counterparties, within and outside the financial sector.

Integrity. The integrity of data is compromised; for example, customer account balances or proprietary trading records are impaired. Similar to confidentiality and availability events, integrity events can have direct costs such as the cost of restoring the integrity of the data and legal costs of resolving issues where integrity cannot be restored. Similar to availability events, integrity events can have severe spillover costs if they impair the bank's ability to perform its core activities. In particular, integrity events and the resulting

legal uncertainty could significantly extend the time required to recover full functioning of the affected bank and the system more broadly.

Based on this classification, cyber events threaten financial stability through their direct costs as well as through the spillovers they cause. While this classification shows similarity to other types of shocks, cyber events are different along specific dimensions (for a related discussion, see Healey et al., 2018, Curti et al., 2019, Duffie and Younger, 2019, Kashyap and Wetherilt, 2019, and European Systemic Cyber Group (2020)).

Intent. A cyber event is typically a deliberate act which may be intended to produce financial gain for the attacker or, more importantly, to create damage to the attacked bank and the financial system. Unlike risk events that impair operations, which may be expected to occur randomly or seasonally, the malicious intent of cyber means that these events may be more likely to happen at times and at banks such that the impact is largest.

Technology. Due to its technological nature, a cyber event can spread through technological linkages, e.g. communication networks, as opposed to traditional economic linkages. In addition, due to the commonality of technology across banks, a cyber event can have significantly larger scale than other types of shocks, affecting a large number of banks at the same time. Combined with malicious intent, attacks with larger potential scale are more likely.

Uncertainty. A cyber event may remain hidden for a considerable time before being detected.[2] This increases the potential for damage while undetected and the problems of recovery, especially for integrity events. Even when detected by one bank, other banks may remain uncertain about whether they are affected as well. Again, this characteristic interacts with intent since an attacker may have an incentive to remain hidden as long as possible.

2.2 Impact of a Cyber Attack on a Financial Institution

Like any operational risk event, a cyber attack can trigger a liquidity run and lead to solvency issues. Such a run can be panic based (Diamond and Dybvig, 1983) or fundamentals based (Goldstein and Pauzner, 2005). As in a standard run, there are direct costs to the af-

[2]Median dwell time — the time from first evidence of compromise to detection — has been close to 100 days in recent years (FireEye Mandiant Services, 2019).

fected bank, e.g. due to inefficient liquidation of assets, and spillovers within the financial sector (counterparties) and to the real economy (borrowers).

A detail specific to cyber attacks is that they may impair the ability of the bank to service running creditors, e.g. if payments or accounts are not available. In such a scenario, the classic first-mover advantage would be weakened, potentially limiting the incentive to run. This channel through which inaccessibility attenuates the run incentives in the aftermath of a cyber attack, however, relies on a partial equilibrium argument. As banks and financial institutions operate over a diverse set of markets and perform multiple functions on behalf of clients, uncertainty regarding the magnitude and potential spillover to other segments of the attacked institution may spark runs from clients of other uncompromised operations of the respective bank. This issue is further complicated by asymmetric information problems that arise as attacked institutions formulate their disclosure strategy for communicating with clients.

More generally, cyber attacks have the potential to immobilize capital and liquidity which can impose costs irrespective of a run or other changes in behavior of the affected bank and its creditors and counterparties.

2.3 Amplification through Network Structure and Uncertainty

It is well established that in addition to accounting-based effects (Eisenberg and Noe, 2001), network structure can amplify and propagate shocks (Allen and Gale, 2000). While these standard models consider the propagation of solvency and liquidity shocks, their insights also apply to cyber shocks. For example, if a cyber shock impacts the distribution of or access to liquidity, it can trigger contagion events as in Allen and Gale (2000). The technological impact of a cyber attack may travel through a different network from interbank lending or payments, as a virus or technical exploit may propagate through data and communications networks, through shared service providers or technological similarities.

In addition, there are recent theoretical insights derived mainly for solvency shocks that also apply to cyber shocks. For example, Erol and Vohra (2018) study the systemic risk in financial networks arising from both correlated defaults of peripheral banks, and defaults of systemically important core banks. Cyber attacks may result in correlated impairments, for example through technological connections, leading to similar systemic risk.

Asymmetric information plays an important amplifying role in standard theories of systemic risk and also applies to cyber shocks. For example, Caballero and Simsek (2013) show how incomplete information about the location of a solvency shock in a financial network greatly increases individual banks' incentive to engage in individually prudent

but systemically harmful actions. This naturally applies to cyber shocks and banks' individually optimal response in financial networks. Uncertainty about the location of a cyber shock can significantly amplify its disruptive effects. News or rumors regarding impairment of some (possibly unknown) banks or financial institutions could lead to preemptive withdrawals from institutions who are not affected.

Strategic complementarities, coordination problems and lack of common knowledge have been studied extensively (for a review, see Morris and Shin, 2003). Afonso and Shin (2011) show specifically for shocks to the payment system — which naturally include cyber shocks — how coordination motives can lead to precautionary demand for liquidity. The analysis of Corsetti, Dasgupta, Morris, and Shin (2004) examines the impact of strategic complementarities with large players, a model that represents well many financial markets and applies to cyber shocks just as it does to fundamental shocks.

A final consideration in cyber is that financial networks structured for resiliency to market risk and for liquidity sharing may be vulnerable to cyber shocks, which connects firms through different networks. If redundancies are not available, networks with a core-periphery structure can see rapid spread when the core is compromised, and may struggle to identify their connections in the network.

2.4 Policy Responses to a Cyber Event

When thinking about policy responses, it is helpful to distinguish solvency and liquidity effects that are similar regardless of the initial shock, and those effects that are specific to a cyber-attack related shock. In this way we can evaluate how ex ante policy measures to reduce the vulnerability to or the impact from a cyber event as well as ex post policy measures once a cyber event has occurred may be necessary. See Kashyap and Wetherilt (2019) for a discussion of ex ante regulatory and supervisory measures.

Similar to a traditional shock, a cyber event may require ex post liquidity injections via the discount window, open market operations or market-wide liquidity facilities. Similarly, ex ante regulation and supervision can emphasize resiliency to and contingency planning for cyber events. However, due to the unique properties of cyber events, traditional policy tools such as ex ante capital requirements or ex post liquidity provision may not be as effective. For example, a cyber event that falls into the availability or integrity category may require additional policy responses such as a bank holiday to allow for the time necessary to recover the affected systems. In addition, regulatory requirements such as liquidity or reserve requirements could be temporarily suspended if banks are technologically unable to address violations, limiting the knock-on effects of perceived impairment.

More generally, it is possible that cyber attacks can be optimally mitigated through additional roles of the Federal Reserve or other agencies. For example, the provision of dedicated back-up facilities in core markets could reduce the impact of availability and integrity events (Duffie and Younger, 2019). Due to the systemic externalities, the individual cost of such facilities would likely outweigh the individual benefit, requiring public provision. In addition, individual banks may not anticipate that conditional on receiving a cyber shock themselves, there is a high probability of other banks being affected at the same time. This is analogous to the issue of "crowded trades", where individual banks overestimate the liquidity that will be available when they want to unwind a position. Pre-positioning of such facilities, and ensuring that a cyber attack would not prevent access to such facilities would be critical, as the reduction in uncertainty can be as valuable as the actual facilities.

The asymmetric information resulting from a cyber event also creates the potential for beneficial policy intervention. Ex ante, requirements to disclose to regulators even minor cyber events or to share with other banks information on threat assessments and contingency plans could increase resilience by reducing uncertainty and improving collective learning. This could also boost credibility of government certification of the information environment. The catalyst for coordination failures is the lack of common knowledge, i.e. strategic uncertainty. Ex post, uncertainty could be reduced, for example, by identifying the affected banks or, instead, certifying key banks as unaffected. However, as Angeletos, Hellwig, and Pavan (2006) illustrate, policies surrounding information alone may not suffice to alleviate coordination failures when policies are endogenously determined.

3 Evidence from the U.S. Wholesale Payment Network

In this section, we investigate quantitatively the impact that a cyber attack can have on the financial system, examining how a cyber attack on a set of banks impairs payment activity in Fedwire Funds Service ("Fedwire"). Fedwire represents the majority of wholesale payments between financial institutions in the U.S. and provides detailed information on the accounts and flows between a diverse set of financial institutions. This enables us to simulate a broad set of scenarios and examine the results on the network without relying on heavy parametric assumptions.

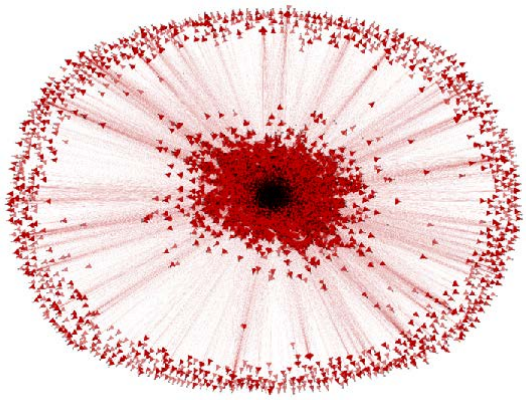


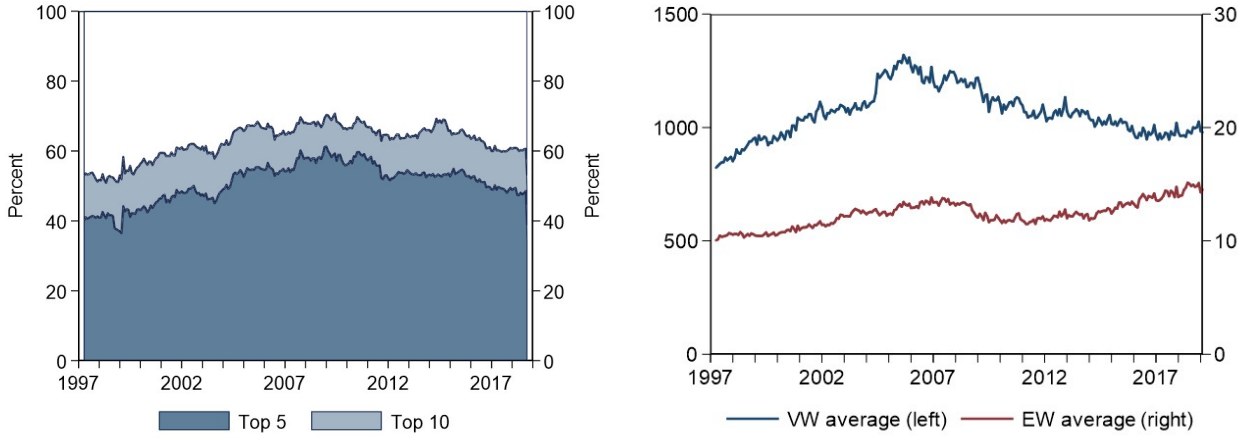
Figure 1: Fedwire payment network. Network representation of the Fedwire payment system. Lines represent payment flows between participant on Fedwire. Distance between pairs represents value of payment flows.

3.1 Overview of the Wholesale Payment Network

The U.S. payments network has a core-periphery structure (Soramäki, Bech, Arnold, Glass, and Beyeler, 2007). Consistent with this, payments are highly concentrated among a small set of institutions (Figure 1). The top five most active banks in the payment system account for close to 50 percent of total payments and the top ten for over 60 percent (Figure 2a). Activity is concentrated not only in terms of payments value but also in terms of network connections, with the most active banks' connections outnumbering the average banks' by several orders of magnitude (Figure 2b).

It is worth noting the dramatic change in the underlying reserve environment after 2008. Past studies that highlight strategic behavior in liquidity management, both empirically and theoretically, assume a system with scarce reserves. Following the financial crisis, the Federal Reserve's asset purchasing programs expanded its balance sheet, and with it the aggregate quantity of reserves. As shown in Figure 3, the ratio of total payments to total reserves has fallen a hundred fold. While a complete analysis of the changes to the underlying payment environment is outside the scope of this paper, we provide some context and relevant dimensions to understand the analysis of cyber attacks.

Existing empirical literature documents strategic behavior and complementarities in the wholesale payment system. When liquidity is scarce, banks manage intraday liquidity by delaying payments and closely matching inflows and outflows to avoid liquidity shortages (McAndrews and Rajan, 2000). Over the normal course of the day, this creates both clusters and delays in the aggregate value of payments. Klee (2010) shows that disruptions due to operational problems increase aggregate uncertainty on market functioning and are



(**a**) Concentration of payments

(**b**) Network dependencies

Figure 2: Concentration and network dependencies. The left panel shows the share of payments sent by the top 5 and top 10 institutions by payments activity. The right panel shows value weighted (VW) and equal weighted (EW) averages across participants of the number of other participants receiving payments from a participant.

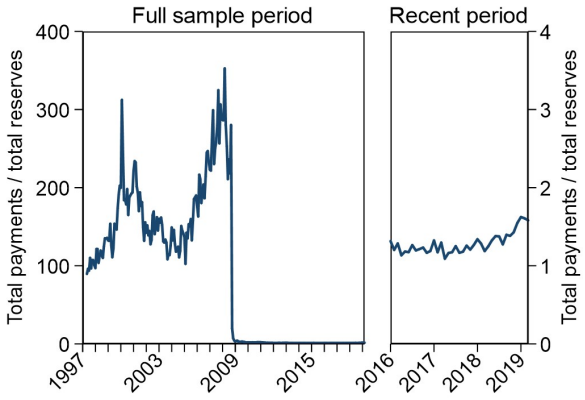


Figure 3: Total payments as a fraction of reserves. This figure provides a time-series plot of the total daily payments in Fedwire, scaled by total reserves.

realized in prices, volume, and discount window activity. Afonso and Shin (2011) simulate intraday liquidity dynamics and show that network externalities in liquidity hoarding behavior can spark illiquidity spirals in payment systems.

Past shocks offer a glimpse into how the payment network behaves in times of crisis. On September 11, 2001, many banks based in lower Manhattan experienced direct damage to property and communication systems (Lacker, 2004) which were followed by payment coordination breakdowns (McAndrews and Potter, 2002). Heightened uncertainty can also amplify private incentives to hoard liquidity. Ashcraft, McAndrews, and Skeie (2011) find evidence of strategic liquidity management in interbank markets during the 2007–2009 financial crisis. In recent work, Allen (2020) considers the impact of operational risk events more broadly on the payments system.

Some measures of banks' strategic management of intraday liquidity have fallen with the abundant supply of reserves. For example, delays in payments and turnover, metrics which proxy for strategic complementarities in payment activity have dramatically decreased (McAndrews and Kroeger, 2016). However, other post-financial crisis changes such as liquidity regulation, liquidity stress testing and improved liquidity risk management mean that banks continue to closely manage their liquidity. More recently, these proxies for strategic complementarities have shown signs of increasing again, suggesting that even in an environment with an ample quantity of reserves, banks' payment activities continue to be carefully managed.

3.2 A Framework for Cyber Vulnerability

Our main strategy is to analyze scenarios in which a cyber attack impairs a subset of institutions, and to evaluate the systemic implications by examining the overall impact on the payment network. Throughout the paper, we focus our analysis on a single-day shock to the payment network. Multi-day impairments necessarily result in broader impact. To fix ideas, we outline a framework used to understand how hypothetical cyber attacks would be amplified through the system.

As highlighted in Section 2, a cyber incident can compromise normal functioning of an institution through affecting a bank's confidentiality, availability, and/or integrity. We focus on scenarios involving availability or integrity. A cyber attack may temporarily impair the availability of relevant data or communication and messaging systems of a target institution. For the duration of the issue, the institution may not be able to process or transmit payments. Alternatively, a cyber attack may compromise the integrity of a target institution's data, either by manipulating or corrupting the data. While the target institution

could, in principle, perform payments on behalf of its clients and/or its key operations to avoid reputational costs, and/or delay announcing its compromised position, this would be unlikely, given the large notional size of payments and because payments in Fedwire are settled with finality. The alternative may be that the institution undergoes a process of reconciliation, in which it rebuilds its data both internally and externally. In either case, a successful cyber attack can have an immediate and direct impact on an institution's ability to carry out day-to-day operations.

A hypothetical scenario starts with a set of target institutions that experience a cyber attack, and are able to receive but unable to remit any payments over the Fedwire operating day. The assumption that only the sending of payments is impacted reflects the institutional practices of Fedwire. In a nutshell, a payment is actualized when Fedwire receives a payment request from the sender. As a result, an institution's balance, as represented in the Fedwire system, can immediately reflect changes due to incoming payments, even if the institution is unable to observe or interact with the Fedwire network due to a cyber incident. Attacked institutions can therefore become liquidity black holes, soaking up liquidity but unable to send out payments (in the form of reserves). This can restrict the effective flow of liquidity, similar to the situation after September 11, 2001 (Lacker, 2004).

Let U be the universe of financial institutions, indexed i, in the network. A scenario S specifies three primitives (Z, R, B), where $Z \subset U$ is the set of financial institutions that are directly affected by the cyber attack (referred to as *shocked* institutions); R is the set of reaction functions of institution $i \in Z^c = U \setminus Z$ that characterize its payment behavior within a given day; and $B = \{b^i\}$ is the set of *impairment* conditions**,** whereby an institution i is considered adversely affected by the scenario if condition(s) b^i holds. With slight abuse of notation, we also use B is to denote the set of institutions $i \in U$ that are impaired. By default, we assume that any directly affected institution $i \in Z$ is also impaired. As such, an outcome used to characterize the network impact includes both that of the *primary* impact (via institutions $i \in Z$) and the *derivative* impact (via institutions $i \in Z^c$).

A scenario provides us with a baseline to describe the behavior of the constituents of the network and evaluate whether an institution is impaired by the scenario or not. Fixing a scenario, we first simulate the scenario for the entire sample period, and construct a dataset with counterfactual outcomes, including the list of impaired institutions. We then assess the effect of the shock by evaluating and analyzing the properties and characteristics of impaired institutions implied by the counterfactual outcomes. We focus on three different aspects. First, what is the magnitude of the network impact caused by the scenario? Second, what is the set of institutions that are vulnerable to the shock? Finally, how does the level of information available to the attacker relate to the systemic implications

of a cyber attack?

3.3 Data Description

Our primary data sources consist of intraday payment data between banks in the Fedwire Funds Service payment system and the individual banks' end-of-day reserve balances over the year 2018. During this period of time, the aggregate supply of reserves gradually shrinks, with depository institutions collectively holding about $2.3 trillion in reserves in January 2018, and about about $1.8 trillion in reserves at the end of 2018.[3]

Table 1: Summary statistics. The table shows summary statistics of payments in Fedwire.

	Avg.	St. dev.	p1	p25	p50	p75	p99
Total sent$_{i,t}$ (millions)	519.94	9588.86	0.00	0.00	0.37	3.16	7242.90
Avg. by institution i (millions)	508.75	9318.75	0.00	0.15	0.86	4.29	7818.85
Total on day t (trillions)	2.85	0.32	2.32	2.63	2.81	3.02	3.73

Table 1 summarizes statistics on daily payment value. The average total daily payments in Fedwire in 2018 is about $2.8 trillion, with a standard deviation of $320 billion and a maximum of over $3.7 trillion. The cross-sectional variation is considerably larger with a highly skewed distribution of total payments sent at the institution level. The median bank sends less than $1 million per day while the average bank sends about $500 million per day and the most active banks send over $7 billion per day.

Fedwire payments data are organized by "ABAs" — account numbers instituted by the American Bankers Association to facilitate check payments. Because some banks have multiple ABAs (typically due to mergers), we aggregate entities to the parent depository institution level.[4] We use data from Call Reports, Y-9C and FFIEC 002 filings as well as Summary of Deposits data to complement Fedwire data with detailed information on financial institutions, including asset size, branch location, and other relevant information. As will be shown, taking into account institutions' size is integral to understanding and evaluating the overall impact of a cyber attack.

Thus, our main universe U consists of financial institutions for which information on asset size is available, which still permits a wide class of U.S. chartered financial institu-

[3] While the aggregate quantity of reserves and distribution of reserves throughout the system will affect this analysis, there is sufficient daily variability to inform the analysis.

[4] Our analysis is at the depository institution level for two reasons. First, financial institutions may face liquidity demands and regulations, including reserve requirements, at the depository institution level. Second, impairments in the payment system could make it difficult to shift liquidity between accounts, even within the same bank holding company. Our main results continue to hold if we aggregate accounts to the bank holding company level.

tions, including national banks, state member and non-member banks, federal and state savings banks, and branches of foreign banks. These institutions together represent about half of the participants in the Fedwire payment system, and account for about three quarters of total daily payments and reserves. All reported results on the network impact are based on this subset of institutions. We exclude all other institutions from the analysis of network impact. By doing so, our results on both unweighted and weighted impact of cyber attacks are based on the same subset of potentially impaired institutions. In addition to excluding a number of non-depository institutions, a large number of institutions excluded from our analysis are credit unions, which together represent about 18 percent of participants, but collectively account for less than 0.5 percent of payments by value. By excluding these institutions from U, we are potentially underestimating the network impact. This is because excluded institutions may be heavily impacted but are not included as potential firms in B. Second, as will be seen in Section 5, excluded institutions are implicitly assumed to continue to send out payments unconditionally, which underestimates the cascade effect.

4 Baseline Scenario

In this section, we outline our baseline scenario and summarize the results. In the baseline scenario, we consider an attack on a single institution on day t and assume that the attacked institution does not send any payments on that day. All other institutions are assumed to make payments as observed in the payments data for day t. We calculate the counterfactual end-of-day-t reserve balance for each institution and analyze the set of institutions B that drop below an impairment threshold (Section 4.1). Since only the shocked institution's payment activity is assumed to be disrupted, the impact in the baseline scenario arises exclusively from other institutions failing to receive payments from the shocked institution. We consider the impact of other institutions endogenously suspending payments in Section 5. There are several rationales for this simple baseline scenario. First, some banks may not have the systems or protocols in place to identify and respond immediately to the failure of a counterparty to make payments. Second, banks may not be sensitive to intraday liquidity conditions, due to the ample reserves environment, and thus may not react to abnormal conditions in a timely manner. Finally, the baseline scenario may also be interpreted as payment network participants continuing to operate as usual in expectation of a regulatory intervention.

We focus attention on attacks on the most active institutions in the payments system. As described in Section 3.1, a well-documented and stable feature of U.S. wholesale payments

systems is its high concentration of activity among a small set of financial institutions. The core banks in the network are among the largest banks in the U.S., with a rank correlation between assets and payments of over 80 percent, and therefore play vital roles in various other markets, including Treasury, equity, and derivatives markets. This strongly suggests that a cyber attack at a single institution alone could materially impair normal functioning in the payment network. We therefore present results for a hypothetical situation in which Z consists of one of the largest five institutions in the payment network ("top-5"), by value in 2018.

4.1 Defining Impairment

An important unknown in evaluating the severity of a cyber event is under which conditions an institution that is not itself attacked becomes materially impaired. For instance, even though we observe each individual bank's reserve levels at a high frequency, we do not directly observe the bank's desired reserve holdings, nor do we observe levels at which the bank would view its liquidity position as severely compromised. For institutions subject to regulatory requirements such as the liquidity coverage ratio (LCR), high frequency and detailed data on balance sheet is needed to assess how binding liquidity regulation is for an institution. Furthermore, banks' internal liquidity targets may incorporate a buffer above the regulatory minimum. Additional liquidity may be required within entities of the bank holding company to support recovery and resolution planning. As individual reserve balances are endogenously determined by both market-wide liquidity conditions, and each bank's idiosyncratic liquidity demand, a bank may manage its day-to-day reserve balances to accommodate shocks to its balance sheet positions.

One way to approximate this is to use empirical variation in banks' reserves as a basis for evaluating when they would be materially impaired. Our main measure uses individual banks' end-of-day reserve dynamics to identify when a cyber event has a significant effect on a bank's liquidity position. In our baseline definition, we classify an institution as *impaired* if the counterfactual end-of-day reserve balance r_t^i is more than two standard deviations below its average, over a 30 day window:[5]

$$b_t^i = \bar{r}_{t-1,t-30}^i - 2\sigma_{t-1,t-30}^i.$$

[5]Since the impairment threshold b_t^i is a moving average based on past reserves, it is possible that a bank's end-of-day reserves drops below b_t^i on certain days, even before accounting for missing payments from our scenario. We show in Appendix A that excluding from the analysis any banks that are impaired but do not receive any payments from the shocked institution on that day does not materially affect our results.

Table 2: Distance from impairment threshold. This table summarizes statistics on banks' average z-score for reserves over 2018, weighted by asset size. The z-score measures banks' distance, in standard deviations, from their 30-day rolling means in the sample period.

p1	p5	p25	p50	p75	p95	p99	Avg.	St. dev.
-.4996	-.4565	-.1570	-.0342	.0588	.2422	.6300	.4852	63.5251

A key advantage to our reserve-based impairment measure is that it provides a time-varying cutoff that adjusts as individual banks optimize their reserve balances to their liabilities and liquidity positions as well as to changes in aggregate reserves.

How binding is b_t^i? One way to test this is to examine how close banks come to their impairment threshold during normal times. Consider the normalized distance to the past 30-day average given by a time-varying z-score measure

$$z_t^i = \frac{r_t - \bar{r}_{t-1,t-30}}{\sigma_{t-1,t-30}}.$$

Table 2 summarizes the average z-scores for banks, weighted by size. The table shows that banks on average stay within roughly half a standard deviation above and below their trailing average balance. This suggests that banks, adjusting for size, tend to actively manage a relatively stable end-of-day reserve balance.

While our current impairment measure is chosen for its intuitive, parsimonious definition, the framework is designed to be flexible to different definitions of impairment b_t^i. The impairment definition can be adjusted to be more or less stringent, simulating possible policy responses to an attack For example if reserve balances reflect compliance with liquidity regulation or supervisory expectations, a policy response could be to change those regulations or expectations in order to mitigate the amplification of an attack.

4.2 Distribution of Impact across Institutions and Time

As a starting point, we examine the share of institutions that become impaired, i.e. end-of-day reserves fall below their impairment threshhold, as a consequence of a successful cyber attack on one of the top-5 institutions. Figure 4 outlines the average daily impact of an attack on each of the five institutions in 2018, where unweighted share is the fraction of impaired institutions, i.e. $\frac{|B|}{|U|}$. Across all days in 2018, the average share of institutions impaired ranges from 4.8 percent to 8.5 percent, depending on which of the top-5 institutions is attacked (blue bars). There is substantial variation on different days of the extent

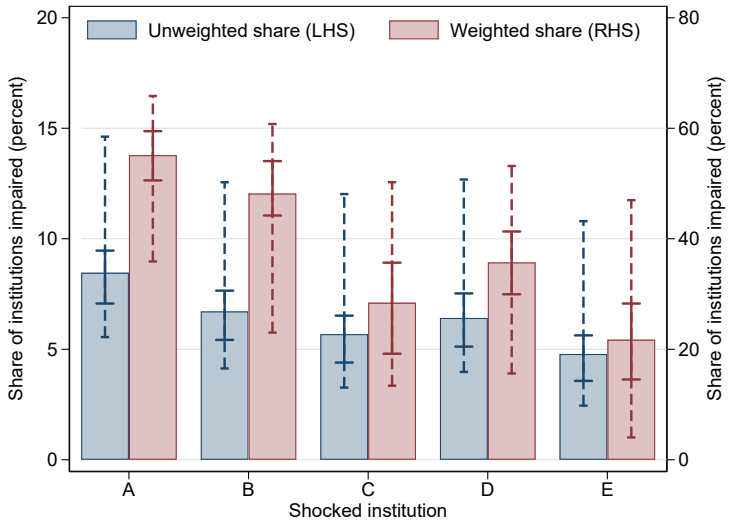


Figure 4: Distributions of impact of a shock to the top-5 institutions. The figure shows the distribution of the unweighted share (blue) and share weighted by assets (red) of institutions impaired by a shock to each of the top-5 institutions. Bars represent the average impact; solid whiskers represent the p25/p75 range; dashed whiskers the p1/p99 range.

of amplification of an attack on a single large institution. The daily variation in the share of impaired institutions is larger within each shocked institution than it is across the shocked institutions. The share impaired on the 1 percent worst days is roughly three times larger than the share on the 1 percent best possible days (dashed whiskers).

By comparison, we also examine a size-weighted share of the impaired institutions, which is computed using $\frac{\omega_q^i |B|}{|U|}$, where $\omega_q^i = \frac{A_q^i}{\sum_q A_q^i}$ and A_q^i is the size of the institution measured in total assets in the quarter q (red bars in Figure 4). Across the top-5 institutions, weighting impaired institutions by asset size leads to between 22 percent and 55 percent of bank assets being impaired, magnifying the average impact by over four times relative to the numerical share of affected institutions.

Figure 5 shows histograms of the distribution across days in 2018, averaged across the shocked top-5 institutions. Compared to the unweighted distribution (blue), the weighted distribution is shifted to the right and exhibits a more symmetric form, whether the shocked top-5 institution is counted among the impaired or not (red or gray, respectively). Since the directly shocked institutions are large in assets, the direct impact of the shocked institution has a nontrivial contribution to the total network impact. However this direct effect does not explain the variation in the simulated network impact. Even excluding the shocked institution, on the days of maximum impact, an attack results in close to half of the network being impaired, in terms of bank assets. This distribution is shown in the gray bars of the histogram, which nets out the direct impact on the top-5 institution. The

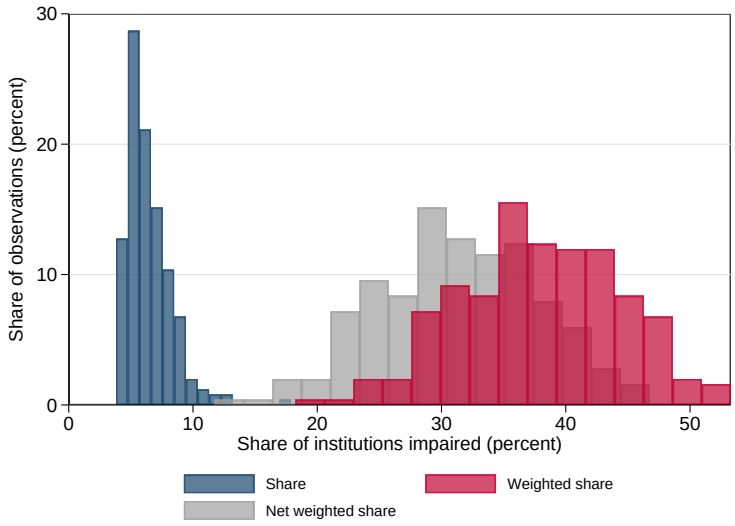


Figure 5: Distribution of average network impact. The figure shows the distribution of the network impact from the baseline scenario, averaged across the top-5 institutions. "Share" represents the unweighted fraction of institutions that become impaired. "Weighted Share" represents the fraction of institutions that become impaired, weighted by asset size. "Net Weighted Share" refers to the fraction of institutions that become impaired, net of the shocked institution.

large amount of assets impacted reflects the high concentration of payments between relatively large institutions, and the large liquidity imbalances that follow if even one large institution fails to remit payments to its counterparties. Because of the high volume and size of payments among core institutions, large and highly central institutions are more commonly impacted by an attack on any of the top-5, and represent a large share of the affected banks. In other words, the large share of affected assets generally reflects transmission from one top-5 bank to another.

Of course, different assumptions would produce different estimates of the extent of amplification. In terms of comparative statics, any ability of the attacked institution to send out some payments would result in less amplification. Similarly, lowering the impairment threshhold would also reduce amplification. Changes to the information environment also affect amplification. For example, if banks were informed about the impairment and stopped sending payments to the affected insitution, their liquidity would alos be improved. We explore the impact of allowing for different responses from other banks in Section 5.

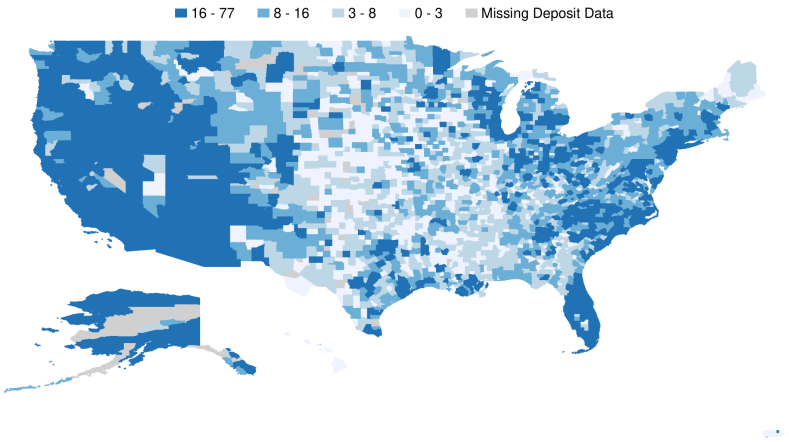


Figure 6: Geographic impact at the county level. The figure shows the average impact of the baseline scenario, where impact is measured by the deposit share (percent) of impaired institutions within each state. Impact excludes the shocked institution.

4.3 Geographic Impact

We began by estimating the number and assets of affected banks on a national basis. A key concern in a cyber attack is that disruptions in the wholesale payment network can lead to widespread spillovers to the real economy. In this section, we enrich our analysis of amplification from a single direct attack in our main scenario, by understanding how the physical structure of the U.S. banking system could transmit shocks to bank depositors and customers. Given the high concentration of activity in key financial centers in the U.S., a natural question is, to what extent a successful cyber attack on a large systemic institution has an impact on particular geographic areas.

To examine the geographic impact on the banking sector, we evaluate the overall market share of impaired institutions by geography. We achieve this in two steps. First, for each institution $i \in U$, we make use of Summary of Deposits data to measure deposit market share in each respective state, Metropolitan Statistical Area (MSA), and county, which proxies for banking market presence. Then, for each scenario, we aggregate the overall impact by total market share of the set of impaired institutions B.

To the extent that banking markets are concentrated, an impairment of a small number of institutions may have a large effect on depositors in that market, particularly when those institutions are highly connected in the wholesale payments network. We find a significant amount of variation in the extent of an impact. Figure 6 depicts the geographical impact aggregated at the county level. A cyber attack on a top-5 institution results in a broad geographic impact, well beyond states with financial centers, such as New York, California, or Illinois. On average, about 23 percent of each state's deposits are at impaired banks.

Five states have over 40 percent of deposits impaired, and over half of all states have at least 20 percent impaired. Some of these high levels reflect the high market shares of the largest banks, which are directly impacted, but also the impact of transmission to other banks. Geographic impact varies because the extent to which the shock affects banking markets depends not only on how connected relevant institutions are to the network, but also on how concentrated markets are. For instance, Ohio sees, on average, 53 percent of its banking sector impaired under the main scenario because it has a high concentration of banking activity by banks within the core of the payment network.

A similar story emerges when looking at the level of metropolitan statistical areas (MSAs). Examining the top ten MSAs with the largest shares of deposits held by impaired institutions, in the baseline scenario, we find that all ten MSAs average at least 40 percent of the local banking market impaired. While the geographic impact is correlated with MSAs with greater financial activity, the impact is also large for MSAs where the financial sector does not play an outsize role in the local economy. This can be attributed to a combination of two factors: concentration in the banking market, and the deposit market share in the MSA of core institutions in the payment network.[6]

This illustrates that disruptions in wholesale payments and subsequent liquidity dislocations can potentially prompt widespread illiquidity of banks that can be very concentrated in certain geographies. Notably, a cyber attack can lead to uncertainty regarding the security and safety of the banking system that can be felt by depositors and financial market participants, both at the institutional and individual household level. This points to a broader concern: a cyber attack has the potential to trigger rapid degradation in the trust towards financial institutions and can therefore quickly lead to traditional forms of financial fragility. At the micro level, this also shows an additional possible risk of highly concentrated banking markets. Disruption in a geographic area is likely to be higher if a large share of customers and businesses cannot access their deposits. While a lack of reserves may not immediately lead to a lack of cash in ATMs, if affected banks delay payments or reduce lending, the impact is likely to be magnified in some geographies.

4.4 Timed Attacks

A defining feature of cyber risk is that attacks may be targeted, with the intent to achieve maximum disruption rather than financial gain. The extent to which an attacker is informed with respect to the payment system overall and the targeted institution specifically

[6]Note that the Summary of Deposits data have not been adjusted to account for the way in which some corporate and online deposits are accounted for at bank headquarters locations, meaning that measures of concentration in MSAs where the large banks are headquartered may be overstated.

can affect the magnitude of systemic risk associated with a cyber attack.

An important dimension of private information entails the timing of a cyber attack. While payment volumes are high throughout the year, utilization varies depending on the demand for transactions by both the real and financial sector. The most active days see over 60 percent more total payments than the least active days (see Table 1). Perhaps unsurprisingly, these large swings in payment value also correspond to large variation in the network impact of a cyber attack. Even within the top-5 institutions, shares of total payments are variable, and result in large swings in the transmission of shocks to other banks over the course of the year.

This indicates that the timing of the cyber attack can heavily influence the magnitude of impact. In this section, we evaluate how progressively higher levels of knowledge that a cyber attacker may possess increase the attacker's ability to inflict damage to the payment network, and to what extent information can be effectively used to disrupt financial activity.

Public information. The first level is public information. Past studies document strong seasonality in wholesale payment activity, with large, periodic increases in payment activity over the course of a year (Klee, 2010; Furfine, 2000; Soramäki et al., 2007). A cyber attack may therefore be timed to coincide with high-activity periods, so as to increase the network impact. To illustrate this, we examine the counterfactual impact of a subset of days coinciding with calendar periods documented by existing studies to have higher than usual payment activity (referred to as "seasonal" days), and compare it an average Fedwire day in 2018. In Figure 7, we provide the percentage increase in the average weighted network impact on seasonal days relative to the average day for the baseline scenario. Timed attacks on seasonal days increase weighted impact by about 11 percent on average.

Private information on a target institution. The second level is private information, specifically pertaining to a target institution. A common form of cyber attack is an "advanced persistent threat" (APT), during which the attacker is undetected and may extract data on an institution's activities and obtain information on the internal mechanisms within the institution. For example, Mexico's Bancomext and Chile's Bank of Chile fell victim to malicious actors who gained access into their systems but remained in stealth mode for extended periods of time before exploiting system weaknesses.[7] These instances primarily involved thefts on the order of millions of dollars. However, given a different objective, such as inflicting the largest damage, an attacker could attempt to analyze the

[7] https://www.wired.com/story/mexico-bank-hack/

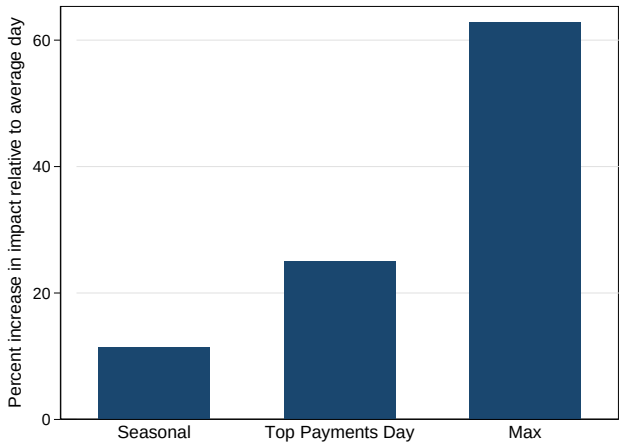


Figure 7: Network impact conditional on attacker's information set. The figure shows how much higher the network impact of an attack on a top-5 institution is on particular days relative to an average day. From left to right, bars show the percent increase in the average weighted share of institutions that become impaired if the attack occurs on a seasonal day, on a top payment day, and on the day of maximum impact, respectively.

compromised institution's financial and payment activity. In the same vein, private information and insider knowledge may become available via direct and indirect forms of insider threat (Randazzo et al., 2005).

In the context of payment networks, a valuable type of private information is understanding patterns in the individual institution's payments. As hinted earlier, days in which an institution conducts large payments are correlated with days in which the scenario yields a large share of impaired institutions. Figure 8 plots the weighted share of impaired institutions on a particular day against one top-5 institution's total payments on that day. As expected, we see that there is a positive correlation between the two measures. Days with a high impact include many days that are seasonal days as well as days where the attacked institution has high payments value.

For each of the five institutions, we identify days in which it individually has high payment value in 2018. Figure 7 shows the network impact for days with the top one percent in payments sent for the average top-5 institution relative to an average day. Attacks on days with high payment value translate into greater network impact relative to seasonal days, and increase the network impact relative to the average day by about 25 percent.

Private information on network interconnectedness. What is the largest possible shock that can be inflicted on the network? This ex-post identification of worst-case days proxies for circumstances under which an attacker is privy not only to private information regarding a target institution's payment activity, but also the underlying network and counter-

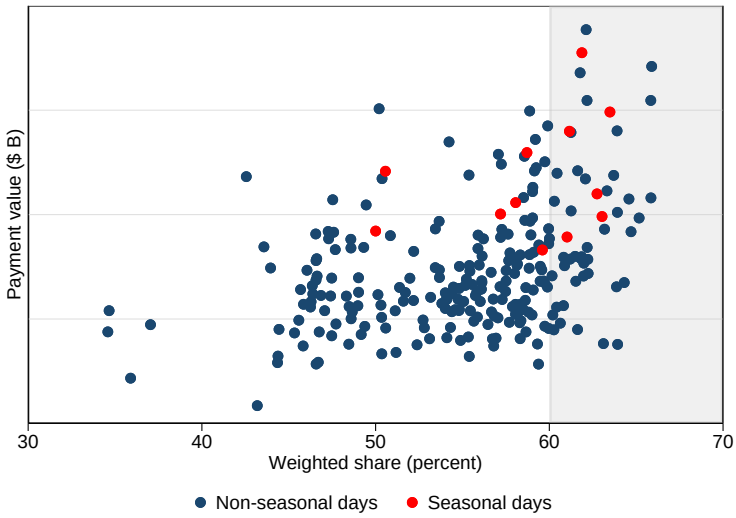


Figure 8: Relation between payment value and network impact. The figure shows a scatterplot of daily payments sent by one top-5 institution against the weighted share of impaired institutions from the baseline scenario of the same top-5 institution.

parties' activities. As shown in Figure 7, the average of the maximum impact for a top-5 institution is considerably large relative to an average day, with an increase in network impact by about 63 percent. While access to such detailed information may be unlikely, it provides us a way to estimate how much access to private information can magnify the systemic consequences of a cyber attack, as well as serving as a reverse stress analysis.

In all of these examples, we show that the extent to which the attacker is privy to detailed institutional and insider knowledge on the payment network, banks' relations, and banks' operations can substantially heighten the systemic risk associated with a cyber attack. One limitation of the payment network data is that we are not able to directly observe the purpose of these payments. It is possible that an attacker that is able to extract further information on payment characteristics, such as the purpose or underlying bank customer could achieve greater disruption to the economy, for example, by timing disruptions to coincide with payroll dates or political events such as elections.

5 Liquidity Hoarding and Network Cascades

The baseline scenario assumes that all institutions other than the directly attacked institution continue to make payments as usual, even after they fail to receive expected payments from a core institution. To study a richer strategic interplay, we now consider an augmented scenario, which we refer to as the *cascade scenario*, in which institutions react

to abnormal imbalances in their intraday reserve positions.

The absence of payments from an affected core institution is likely not only to create large liquidity dislocations at the end of the day, but also to have a significant impact on the intraday balances of counterparties. Relative to an operational outage, a suspected cyber attack may be accompanied by more uncertainty and lack of common knowledge regarding the source, magnitude, and implications of the attack. Given the malicious intent of a cyber attack, market participants may also be concerned about extended disruption and delayed resolution. Uncertainty regarding the nature of a suspected cyber attack may also be sustained endogenously — target institutions may be reluctant to disclose to counterparties and clients the exact state of their internal systems, and may delay communications until they acquire enough information. Such delays can also contribute to asymmetric information, prompting institutions to stall payments and attempt to retain as much liquidity as possible (in risk of operational default). Finally, even if a particular bank is happy to send payments, its customers may hoard balances or try to effect payments through other means (Duffie and Younger, 2019).

5.1 Intraday Reaction Functions

An institution may take notice of abnormal payment activity if it observes unusually large deficits in its intraday liquidity position. Let $\eta^i_{\tau,t}$ be the net payment deficit of institution $i \in U$ on day t up to minute τ (total outflows minus total inflows). We assume that each institution i uses a trigger strategy $r^i \in R$ that characterizes its intraday reaction to changes in its net payment deficit: if $\eta^i_{\tau,t}$ surpasses some threshold, institution i stops sending payments for the rest of the day. In our main specification, we take this threshold value to be the maximum realized net payment deficit of the institution in the entire year of 2018. This threshold, denoted $\eta^i_{\max}$, corresponds to the maximum difference that an institution had between its payment inflow and payment outflow under normal operations, at the minute level. Bank i's reaction function is then given by

$$r^i(\{\eta^i_{\tau,t} | \tau \leq k\}) = \begin{cases} \text{send payments at minute } k & \text{if } \eta^i_{\tau,t} < \eta^i_{\max} \text{ for all } \tau \leq k \text{ on day } t \\ \text{hoard liquidity} & \text{otherwise.} \end{cases}$$

There are two different channels that may trigger a bank to hoard liquidity. The primary channel is due to deficits that grow as the institution affected by the cyber attack fails to remit payments that it would have under the normal course of the day. When the volume of such payments to institution $i \in Z^c$ becomes sufficiently large, institution i may halt its own payments in order to preserve and absorb incoming liquidity, in anticipation of

aggregate illiquidity at the end-of-day. Thus, the trigger strategy r^i can be interpreted as the reaction to an informative signal (i.e. $\eta^i_{\tau,t}$) through which institution i becomes aware of the possibility of a cyber attack on a large systemic institution, or as an optimal liquidity management strategy to maximize the probability that its end-of-day reserve balance does not fall below its internal target, i.e. $r^i_t > b^i_t$.[8]

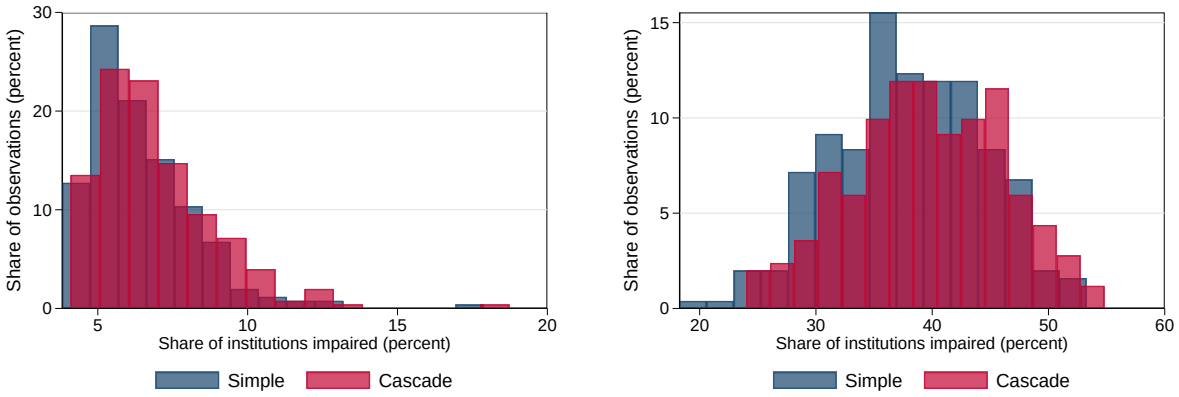
A secondary channel that would trigger a bank to hoard liquidity arises from illiquidity spillovers that hoarding behavior of institutions inflicts on other institutions. Relative to a model without hoarding or strategic reactions to intraday reserves imbalances, the liquidity hoarding behavior of other banks collectively contributes to additional non-payments. An individual bank, by halting payments, can retain liquidity, leading to *endogenous* liquidity traps. This can, in turn, exacerbate the reduced flow of payments, as other banks, which may not have been significantly impacted by the first wave, also pass their trigger strategy and amplify the initial liquidity shortage. Importantly, from a counterparty's perspective, the primary institution that is impaired due to the cyber attack, and secondary institutions that hoard liquidity may be indistinguishable. As a result, even if the network distance between an institution and the primary target institution is large, the institution trigger strategy is aptly described as before – either due to concern of a cyber attack or prompted to liquidity concerns.

Of course, there are other ways to model bank's endogenous reactions to intraday reserves imbalances as well as their reaction to the failure to receive payments from counterparties. We believe that this approach is parsimonious and captures the dimensions through which banks may respond to inform the scale of the cascade. It is also possible to envision a scenario where a malicious attacker publicizes the attack, and even banks that were not exposed to any missed payments begin to hoard liquidity, in which case the cascades would be even larger than those we model here.

5.2 Impairment in Cascades

When banks respond to intraday payment activity, banks, responding to abnormal liquidity shortages from deficits in payment inflows, may individually seek to improve their liquidity positions by forgoing payments. While this imposes network externalities, the overall impact on liquidity dislocations are unclear. Institutions may collectively resolve uncertainty about their end-of-day liquidity conditions by hoarding liquidity. Setting aside the large quantity of unsettled payment demand, this would imply that fewer banks may

[8]Note, this implies that for each institution, there is exactly one day in which it reaches $\eta^i_{\max}$ at some moment within the day. To avoid mechanical liquidity hoarding, we exclude the institution from reactions on the day that $\eta^i_{\max}$ is actually reached, though doing such does not significantly affect the results.



(**a**) Unweighted share

(**b**) Weighted share

Figure 9: Empirical distribution of network impact for cascade scenario. The figure shows the distribution of the network impact for the baseline scenario (blue) and the cascade scenario (red), averaged across the top-5 institutions. The left panel shows the unweighted share of impaired institutions. The right panel shows the share of impaired institutions weighted by assets.

be impaired, as defined in Section 4.1, as a result of the cyber attack.

Our results suggest that this is not the case. Figure 9 shows the empirical distribution of the cascade scenario and compares it to the baseline scenario. In both we compare averages over each of the days for an attack on each of the top-5. We find that strategic liquidity hoarding, if anything, amplifies the network impact of the cyber event. This indicates that taking into account strategic behavior of payment network participants, a cyber attack has even greater systemic consequences.

The cascade scenario provides us with a high-frequency counterfactual illustrating how liquidity hoarding evolves over the course of a given day (Fedwire payments for day t occur between 9:00pm EST on day $t-1$ and 6:30pm EST on day t). Figure 10 illustrates results on intraday dynamics of liquidity hoarding behavior from two cascade scenarios (attacks on top-5 banks A and B), averaged across days in 2018. Over the course of the day, more and more counterparties of the attacked institution pass their trigger threshold, and start to hoard liquidity. Eventually, their suspended payments also cause their recipients to start hoarding and so on. The pace at which hoarding accelerates over the day is driven by a combination of the primary and secondary channels. First, in a normal day, payment activity intensifies towards the afternoon. As a result, the original attacked institution's failure to remit payments affects a broader set of institutions later in the day. As the day unfolds, the number of institutions that switch to liquidity hoarding increases, even more institutions pass their intraday trigger threshold, and revert to self preservation as well.

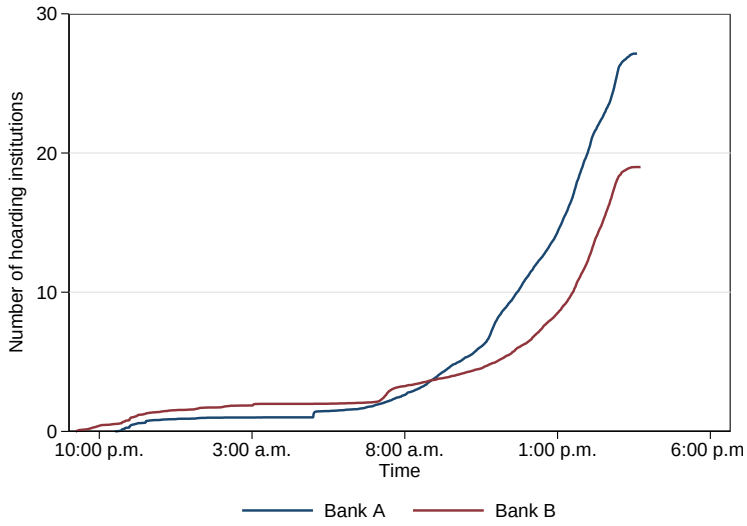


Figure 10: Intraday dynamics of liquidity hoarding. The figure shows, over the course of the Fedwire day, the average number of institutions triggered to hoard liquidity for the cascade scenario of two top-5 institutions.

5.3 Foregone Payments in Cascades

A stark difference, relative to the baseline scenario, is in the value of payments that are not made as a result of hoarding behavior. Relative to before, the systemic implications are driven not only from the compromised liquidity positions of financial institutions, but from the failure to facilitate vital financial services and operations necessary for financial markets and the broader economy. Under the cascade scenario, forgone transactions, in terms of value, represent from five to 35 percent of total daily payment value, and amount from one to eleven times daily GDP. Figure 11 provides a a distribution of the average forgone transactions by payment value associated with cascade scenarios on the top-5 institutions. The range of the distribution illustrate the variability of the vulnerability in calendar time as discussed in Section 4.4. As expected, a significant fraction of forgone value can attributed to the shocked institution, which in our baseline scenario is a top-5 institution by payment value. The net forgone volume represents forgone payments by value for hoarding institutions, which still reaches 0.75 times daily GDP on average.

Wholesale payments such as those in Fedwire can be broadly categorized as payments related to real economic activity, e.g. firms paying their suppliers, and payments related to financial activity, e.g. interbank loans. We attempt to identify these two categories using a business function code that institutions can apply to each of their Fedwire payments. We proxy for real-activity payments by using the code "Customer Transfer" and for financial-activity payments using the code "Bank Transfer". Figure 12 shows the distribution of

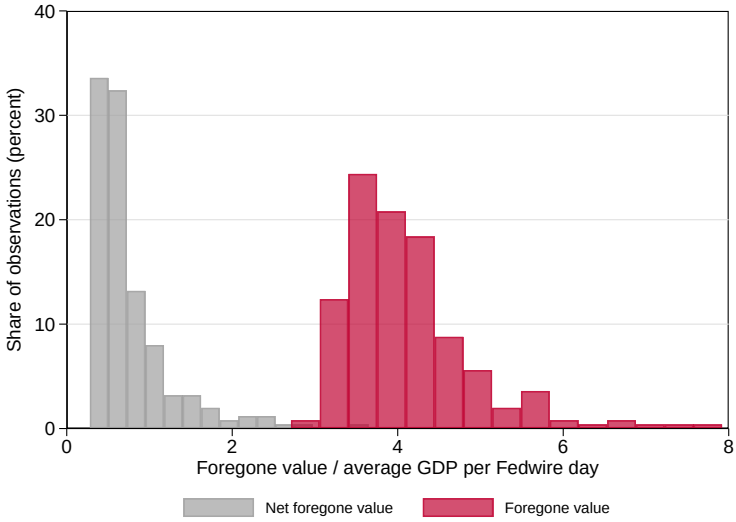
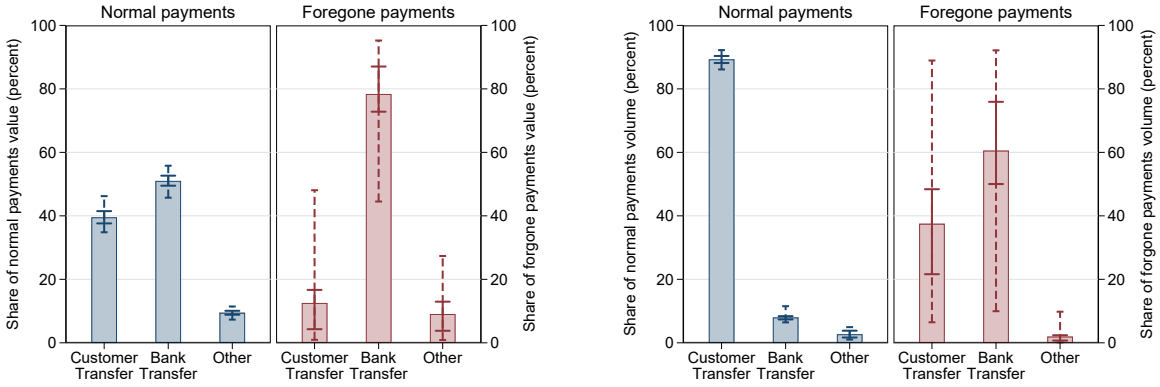


Figure 11: Distribution of total value of forgone transactions. The figure shows the distribution of the value of transactions forgone as a consequence of banks' liquidity hoarding behavior in the cascade scenario, averaged across the top-5 institutions. Payment value is scaled by daily GDP, where the number of days is given by Fedwire days. The net forgone value excludes payments by the shocked institution.



(a) **Distribution of payments value** (b) **Distribution of payments volume**

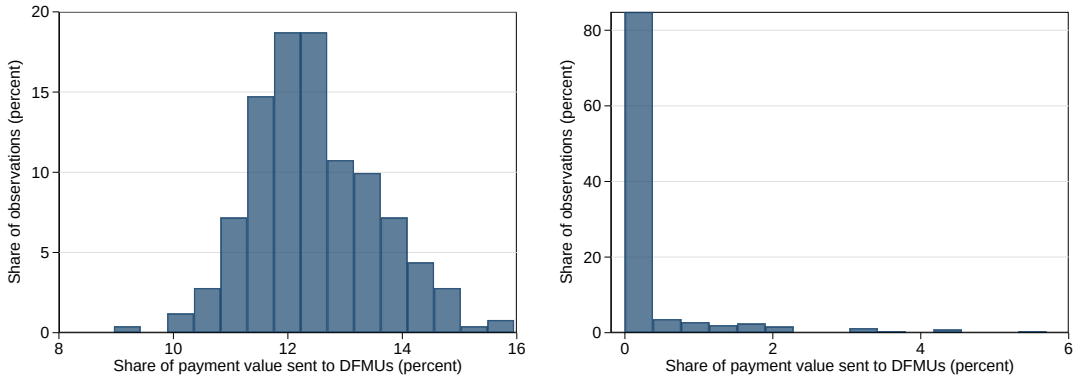
Figure 12: Distribution of transfer types in normal payments and in the cascade scenario's foregone payments. The figure shows the distribution of Fedwire business function codes used among payments in normal times (blue), and the distribution used among foregone payments in the cascade scenario (red, averaged across the top-5 institutions). Figure 12a shows the distribution in terms of payments value and Figure 12b in terms of payments volume. Bars represent the average impact across days; solid whiskers represent the p25/p75 range and dashed whiskers the p1/p99 range across days. All statistics are "net" (excluding payments by the shocked institution).

payments across real and financial activity in normal times (blue bars) and compares it to the distribution of foregone payments in the cascade scenario (red bars). We see, first, that the distribution of foregone payments varies dramatically more across days than the distribution of normal payments (whiskers). This is consistent with the high variance of network impact across days and the implied tail risk for cyber events (Section 4.4). In addition, we see that transactions with other financial intermediaries ("Bank Transfers") represent the majority of foregone payments and are overrepresented compared to the distribution of normal payments, both in terms of value (Figure 12a) as well as in terms of volume (Figure 12b). This may be viewed as attenuating the impact of the cyber event if we think that such payments can more easily be handled via back-up systems, e.g. because of high value but low volume or because of a high potential for netting. However, financial-activity payments could also be assigned a high multiplier in terms of real activity they support indirectly.

A special sub-category of financial transactions are payments to financial market utilities (FMUs) such as exchanges, central counterparties, and other payments systems. As such, the impact of a cyber attack that disrupts payments in Fedwire can be magnified through its effect on other FMUs, through two channels. First, FMUs may serve as a contagion channel for payment disruptions in Fedwire to propagate and trigger settlement failures and illiquidity in financial markets more broadly. Second, illiquidity may be magnified through FMUs that offer liquidity-saving mechanisms, such as netting.

To understand the potential amplifications through FMUs, we examine the disruption in payments sent to Designated Financial Market Utilities (DFMUs) in the cascade scenario.[9] Figure 13 shows the results, broken down into forgone payments by directly shocked institutions and those made by banks that strategically hoard liquidity as a result of the cascade. As expected, we find that a top-5 institution represents a sizable fraction of total daily payments received by DFMUs, reaching up to 16 percent (left panel). This suggests that the concentration of payment activity in Fedwire translates similarly to other interconnected FMUs, and could have similar ramifications for a broader set of financial activity. However, the amplification beyond the direct shock is relatively limited as the payments forgone by banks hoarding liquidity in the cascade account for a smaller share of payments to DFMUs (right panel).

[9]For a list of DFMUs, see https://www.federalreserve.gov/paymentsystems/designated_fmu_about.htm.



(**a**) Shocked institution (**b**) Liquidity hoarding banks

Figure 13: Distribution of forgone payments to DFMUs. The figure shows the distribution of the share of payments to DFMUs forgone in the cascade scenario, averaged across the top-5 institutions. The left panel shows the share of payments to DFMUs from the shocked top-5 institution. The right panel shows the share of payments to DFMUs forgone due to liquidity hoarding behavior.

6 Correlated Vulnerabilities

In this section, we explore scenarios in which multiple institutions are directly affected due to technological or other commonalities. As in the baseline scenario, we assume in the analysis in this section that banks do not respond strategically throughout the day.

6.1 Technological Commonality

Existing research on financial networks examines transmissions between banks linked either through interbank borrowing and other counterparty relationships. In the context of cyber, a salient concern is risk associated with technological commonality between financial institutions. As linkages between institutions in a network arise from financial motives, a financial network may not be structured to be resilient to shocks arising from technological similarities.

Vulnerabilities arising from third-party service providers is viewed as a prominent sources of cyber risk especially when a provider is common to many institutions. With third-party service providers, a classical free-rider problem can arise because agents may individually have weaker incentives to monitor the common provider which erodes governance and oversight (Shleifer and Vishny, 1997). Further, it may be practically challenging for banks to effectively monitor the resilience of their third (and fourth and fifth) party providers. The Federal Reserve has highlighted the importance of managing risk from financial institutions' service providers including activities beyond traditional core bank

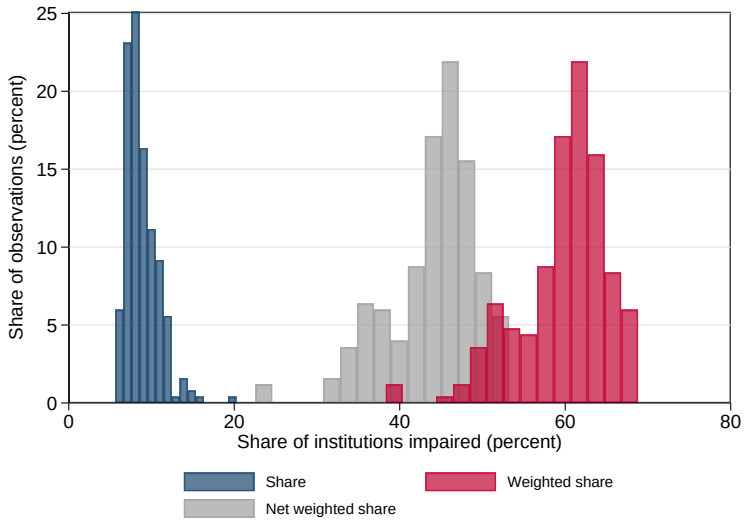


Figure 14: Cyber vulnerability through a third-party service provider. The figure shows the distribution of the network impact for the scenario with a disruption originating from a third-party service provider. "Share" represents the unweighted percent of institutions that become impaired. "Weighted share" represents the percent of institutions that become impaired, weighted by asset size. "Net weighted share" refers to the percent of institutions that become impaired, net of the shocked institutions.

processing and information technology services.[10] While the Federal Reserve supervises some key third party service providers, the connections between institutions from shared service providers are not likely to be widely understood by the network.

We therefore analyze a scenario in which multiple institutions share a service provider, making use of regulatory data on customer lists of significant service providers. The service provider specializes in data and systems management and serves a set of large and medium-sized banks. We assume that a successful cyber attack on the service provider disrupts client institutions' payment operations as if they were directly attacked.

Figure 14 shows the empirical distribution of the scenario and shows that a shock to the set of large and medium banks can have an outsize impact on the payment network. Similar to the baseline scenario, the weighted share of impaired institutions is several times greater than the unweighted share. However, the overall magnitude is greater than the effect of an attack on a single top-5 institution with an average weighted impact of 60 percent. This points to the potential for a third-party service provider, by simultaneously impairing multiple institutions, to cause a systemic event that is amplified through the payment network and propagates through the U.S. financial system.

[10] See, for example, SR-13-19 Guidance on Managing Outsourcing Risk (https://www.federalreserve.gov/supervisionreg/srletters/sr1319.htm).

6.2 Reverse Scenario

A clear implication of the baseline scenario is that a successful cyber attack on any single one of the top-5 institutions has the potential to have significant spillovers to other banks, amplifying the attack through the financial system. However, these large institutions can take advantage of economies of scale in investment in cyber defenses and, as systemically important institutions, they are also subject to heightened supervisory standards and stringent capital and liquidity regulation. They also dedicate large amounts of resources to cyber resiliency, meaning that from the perspective of a malicious attacker, a cyber attack on the largest institutions would be costly. A malicious actor may therefore instead target institutions that may be relatively easier to successfully infiltrate. In particular, smaller institutions, may have fewer resources to defend against a sophisticated attack. With this in mind, we run a *reverse scenario* — estimating what is the minimum set of smaller institutions necessary to impair a top-5 institution in the payment network.

With a large amount of time and computing power, it would be possible to estimate all the different permutations of institutions that would succeed in impairing one of the top-5 institutions, or resulting in the impairment of a given share of payments or banking assets. In order to implement this approach more parsimoniously, we group institutions by asset size, since a number of regulations and supervisory standards are implemented with cut-offs based on asset size. Implicitly, this analysis highlights that supervisory and regulatory standards based on bank size may not take into account the network externalities of smaller banks with respect to a cyber attack.

Our reverse scenario is implemented as follows. Two sets of potentially shocked institutions, $U_N \subset U$ are considered: (i) "mid-sized" entities with asset size between \$10 billion and \$50 billion, denoted U_{50}; (ii) "small-sized" entities with asset size below \$10 billion, denoted U_{10}. These cutoffs are chosen to identify subsets of banks under meaningfully different and progressively relaxed regulatory and supervisory conditions.[11] In setting these cut-offs, we are motivated by Federal Reserve Supervisory guidance such as that laid out in SR 12-17[12] which makes asset size based distinctions in a supervisory framework which differs at the \$50 billion in assets cut-off. This framework specifically does not apply to community banking organizations, defined as banks under \$10 billion in size. Also, the Dodd-Frank Act only requires banks with more than \$10 billion in total assets to be subject to regulatory oversight by the CFPB and to be required to perform and report the results of annual firm-run stress tests. In order to focus on small banking institutions, we exclude any accounts associated with government-sponsored entities (including

[11] Eisenbach, Lucca, and Townsend (2019) provides a detailed study on bank supervision based on size.
[12] https://www.federalreserve.gov/supervisionreg/srletters/sr1217.htm

Table 3: Summary statistics of reverse scenario. The table reports the distribution for the reverse scenario of $|U_{Nt}^{\min}|$ for entities under \$10 billion in asset size, $N = 10$, and entities between \$10 and \$50 billion in asset size, $N = 50$. "Days with Impairment" indicates the number of days for which there existed at least one set of banks in U_N for which a top-5 institution would become impaired.

Impairment	p1	p25	p50	p75	p99	Mean	SD	Days with Impairment
U_{10}	1	1	5	24	221	24	50	23 of 250
U_{50}	1	1	3	8	111	10	25	101 of 250

federal home loan banks), FMUs, and branches of FBOs.

For each set of potentially shocked institutions $U_N \in \{U_{10}, U_{50}\}$, for each top-5 institution j, and for each day t, we identify the smallest subset of U_N which, if successfully attacked, would lead to the top-5 institution becoming impaired, according to the impairment definition in Section 4.1. Formally, we identify the set $U_{Njt}^{\min} \subset U_N$ which is given by

$$U_{Njt}^{\min} = \arg \min_{U' \subset U_N} |U'| \text{ s.t. } j \in B \text{ for } (U', R, B) \text{ on date } t.$$

We then identify for each day the minimum number of entities required to impair any one of the top-5 institutions, i.e. the set $U_{Nt}^{\min} = \arg\min_j \{|U_{Njt}^{\min}|\}_j$. We outline the results of the reverse scenario in Table 3. The table outlines both the number of days in which a top-5 institution is impaired or not, along with summary statistics of the number of institutions in $U_{Nt}^{\min}$ for t in 2018.

Our results suggest that a successful attack on small institutions could be amplified dramatically through its effect on top-5 institutions. On 23 of the 250 Fedwire days in 2018, a coordinated cyber attack on a set of small institutions, if judiciously selected, could result in at least one of the top-5 institutions' reserves falling below its liquidity threshold. Among those days where a potential impairment arises, an average of 24 small institutions successfully attacked were sufficient. Raising the size to mid-sized entities suggests even more amplification; we find a greater number of days with potential impairments, and a smaller number of successfully attacked institutions required to impair a top-5 institution.

The amplification from small banks can be attributed to two factors. First, though small in size, these institutions may collectively have large payments by value. Second, on certain days, end-of-day reserves of a top-5 institution may be close to its impairment threshold. On such a day, the top-5 institution's end-of-day reserve buffers may be insufficient to cover unexpected payment deficits even from small institutions. Combined, these factors suggest that vulnarability can be reduced both through strengthening the cyber resilience

Table 4: Summary statistics of reverse scenario including FBOs. The table reports the distribution for the reverse scenario of $|U_{Nt}^{\min}|$ for entities under $10 billion in asset size, $N = 10$, and entities between $10 and $50 billion in asset size, $N = 50$, including branches of FBOs. "Days with Impairment" indicates the number of days for which there existed at least one set of banks in U_N for which a top-5 institution would become impaired.

Impairment	p1	p25	p50	p75	p99	Mean	SD	Days with Impairment
U_{10}	1	1	2	6	381	11	43	180 of 250
U_{50}	1	1	1	1	6	1	1	250 of 250

of small institutions as well as through robust liquidity buffers for the largest institutions.

To the extent that supervisory scrutiny can help ensure cyber resilience, we also explore the impact of a potential attack on branches and agencies of FBOs. These firms have access to Fedwire payments, but, much like small U.S. banks, are not subject to the heightened supervisory scrutiny at large domestic institutions. Furthermore, many FBO branches are "small" by assets, but exhibit outsized payment activity.

Table 4 shows results of our reverse scenario when we allow the sets U_{10} and U_{50} to include FBO branches. After including FBO branches, we find that there exists a subset of banks that results in an impairment of a top-5 institution for 72 percent (U_{10}) and 100 percent of days (U_{50}). For the majority of Fedwire days in 2018, two or fewer institutions successfully attacked are sufficient to impair at least one of the top-5 institutions. Strikingly, the reverse scenario also indicates that, on an average day in our sample, only one mid-sized institution in U_{50} is necessary to impair a top-5 institution. Several factors drive these results. First, FBO branches are responsible for significantly larger value of payments relative to their asset size. Second, our results indicate greater payment flow between FBO branches and core institutions relative to a domestic institution of similar size. The large role that these institutions play in the payments network suggests that it is important to ensure their cyber resiliency in the context of the U.S. payment system.

7 Conclusion

Our analysis demonstrates how cyber attacks on a single large bank, a group of smaller banks or a common service provider can be transmitted through the payments system. A cyber attack on any of the most active U.S. banks that impairs any of those banks' abilities to send payments would likely be amplified to affect the liquidity of many other banks in the system. The extent of the amplification would be even greater if banks respond strategically, which they are likely to do if there is uncertainty about the attack. The impact on geographies with concentrated banks may be even larger. We also identify other ways

that the system may become impaired that highlight the importance of all banks in the network, not just the largest banks. First, if a number of small or midsize banks are connected through a shared vulnerability, such as a significant service provider, this could result in the transmission of a shock throughout the network. Similarly, banks with a relatively small amount of assets but large payment flows also have the potential to impair the system.

While the shock we assume is extreme in some ways — a complete inability to send payments — it is conservative in others. We currently do not model spillovers outside the payment network such as to short term creditors that provide liquidity to impaired banks. Allowing creditors and customers to run without the ability to realize additional liquidity through the sale of liquid assets would surely make our results even larger. We also focus primarily on the impact that a cyber attack may have within a single day. However, if a cyber attack were to compromise the integrity of banks' systems, the reconciliation and recuperation process would be an unprecedented task. This could have severe implications on the stability of the broader financial system vis-à-vis spillovers to investors, creditors, and other financial market participants. One shortcoming of the analysis is that while we see payment flows among banks, we do not know the purpose of the flows. Therefore we cannot measure how these flows will affect customers and borrowers and the real economy. In future work, we hope to explore how borrowers and other counterparties would respond.

References

Afonso, G. and H. S. Shin (2011). Precautionary demand and liquidity in payment systems. *Journal of Money, Credit and Banking 43*(s2), 589–619.

Allen, F. and D. Gale (2000). Financial contagion. *Journal of Political Economy 108*(1), 1–33.

Allen, H. J. (2020). Payments failure. Working Paper, American University.

Angeletos, G.-M., C. Hellwig, and A. Pavan (2006). Signaling in a global game: Coordination and policy traps. *Journal of Political Economy 114*(3), 452–484.

Ashcraft, A., J. McAndrews, and D. Skeie (2011). Precautionary reserves and the interbank market. *Journal of Money, Credit and Banking 43*, 311–348.

Barro, R. J. and J. F. Ursúa (2012). Rare macroeconomic disasters. *Annu. Rev. Econ. 4*(1), 83–109.

Boston Consulting Group (2019). *Global Wealth 2019: Reigniting Radical Growth*.

Caballero, R. J. and A. Simsek (2013). Fire sales in a model of complexity. *Journal of Finance 68*(6), 2549–2587.

Corsetti, G., A. Dasgupta, S. Morris, and H. S. Shin (2004). Does one soros make a difference? a theory of currency crises with large and small traders. *Review of Economic Studies 71*(1), 87–113.

Curti, F., J. Gerlach, S. Kazinnik, M. J. Lee, and A. Mihov (2019). Cyber risk definition and classification for financial risk management.

Diamond, D. W. and P. H. Dybvig (1983). Bank runs, deposit insurance, and liquidity. *Journal of Political Economy 91*(3), 401–419.

Duffie, D. and J. Younger (2019). Cyber runs. Hutchins Center Working Paper 51, Brookings Institution.

Eisenbach, T. M., D. O. Lucca, and R. M. Townsend (2019). The economics of bank supervision. Working Paper.

Eisenberg, L. and T. H. Noe (2001). Systemic risk in financial systems. *Management Science 47*(2), 236–249.

Erol, S. and R. Vohra (2018). Network formation and systemic risk. Working Paper, Carnegie Mellon University.

European Systemic Cyber Group (2020). Systemic cyber risk. Technical report, European Systemic Risk Board.

FireEye Mandiant Services (2019). *M-Trends 2019*.

Furfine, C. H. (2000). Interbank payments and the daily federal funds rate. *Journal of Monetary Economics 46*(2), 535–553.

Goldstein, I. and A. Pauzner (2005). Demand–deposit contracts and the probability of bank runs. *Journal of Finance 60*(3), 1293–1327.

Gorton, G. (2014). Some reflections on the recent financial crisis. In *Trade, Globalization and Development*, pp. 161–184. Springer.

Healey, J., P. Mosser, K. Rosen, and A. Tache (2018). The future of financial stability and cyber risk. Cyber Security Project at Brookings, Brookings Institution.

Kashyap, A. K. and A. Wetherilt (2019). Some principles for regulating cyber risk. *AEA Papers and Proceedings 109*, 482–87.

Klee, E. (2010). Operational outages and aggregate uncertainty in the federal funds market. *Journal of Banking & Finance 34*(10), 2386–2402.

Lacker, J. M. (2004). Payment system disruptions and the federal reserve following september 11, 2001. *Journal of Monetary Economics 51*(5), 935 – 965.

McAndrews, J. and A. Kroeger (2016). The payment system benefits of high reserve balances. *Journal of Payments Strategy & Systems 10*(1), 72–83.

McAndrews, J. and S. Potter (2002). Liquidity effects of the events of september 11, 2001. *Economic Policy Review 8*(2).

McAndrews, J. and S. Rajan (2000). The timing and funding of fedwire funds transfers. *Economic Policy Review 6*(2).

Morris, S. and H. S. Shin (2003). *Global Games: Theory and Applications*, Volume 1 of *Econometric Society Monographs*, pp. 56–114. Cambridge University Press.

Randazzo, M. R., M. Keeney, E. Kowalski, D. M. Cappelli, and A. P. Moore (2005). Insider threat study: Illicit cyber activity in the banking and finance sector.

Shleifer, A. and R. W. Vishny (1997). A survey of corporate governance. *The journal of finance 52*(2), 737–783.

Soramäki, K., M. L. Bech, J. Arnold, R. J. Glass, and W. E. Beyeler (2007). The topology of interbank payment flows. *Physica A: Statistical Mechanics and its Applications 379*(1), 317–333.

A Classifying Impaired Institutions

As a robustness exercise, we examine a variation that excludes from the set of impaired institutions any banks that are impaired but do not receive any payments from the shocked institution on that day. Figure 15 shows results for the average network impact, after excluding the the subset of banks that satisfy the above criteria. The weighted share of impaired institutions remain large, while we see a noticeable decrease in the unweighted share of impaired banks. These results indicate that, consistent with the main analysis, that network impact is primarily driven by large, core institutions.

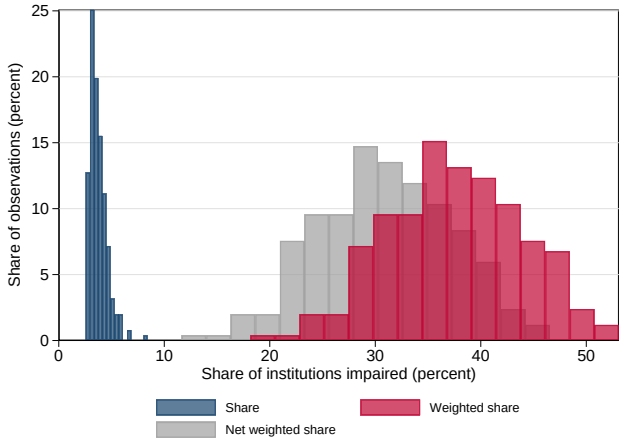


Figure 15: Average network impact for alternative impairment classification. The figure shows the distribution of network impact from the baseline scenario where we exclude from the set of impaired institutions any banks that are impaired but do not receive any payments from the shocked institution on that day.