Towards a new paradigm for resiliency and security

MI FORUM 2016 | GENEVA

The wholesale settlement systems of the Federal Reserve Banks are by any standard systemically important. They provide the infrastructure by which liquidity circulates through the real economy and the financial system of the United States, the means by which the Federal Reserve settles its monetary policy operations, and the platform through which the United States government issues securities to finance its operations. Given the critical importance of these systems, and the changing nature of the cyber-threats they face, traditional defences against physical attack may not be sufficient. Richard P. Dzina, Executive Vice President and Head of the Wholesale Product Office of the Federal Reserve Bank of New York, argues that systemically important financial market infrastructures may now need to consider greater diversity in a third level of resiliency and security.

The 21st century is going to be a volatile one. Around the world, societies and economies are subject to tectonic shifts with unpredictable implications for cyber-, terror, and geopolitical threats. In this unpredictable environment, one certainty remains: attacks on critical financial market infrastructure are not a matter of “if” but “when.”

This was the message I heard from General Michael Hayden, former Head of the National Security Agency (NSA) and Director of the Central Intelligence Agency (CIA), at a symposium of payments bankers in 2014. As the operator of the wholesale services for the Federal Reserve Banks, this was a sobering message on which to reflect.

The systemic importance of wholesale services

Those wholesale services consist of the Fedwire Funds Service, the Fedwire Securities Service, and the National Settlement Service. Collectively, these services constitute the "franchise" when it comes to the financial market infrastructure of the United States. That may sound like a bold assertion, but it is not an unreasonable one, reflecting at least four considerations.

First, transactional value. In 2015 we processed in excess of $1 quadrillion in Funds, Securities, and National Settlement transactions. That is a one followed by 15 zeros, and is equivalent to the gross domestic product of the United States flowing through our pipes every four days. In other words, the wholesale services represent the central conduit of liquidity – indeed, the circulatory system – of the American economy and financial system.

Secondly, inter-connectedness. In 2012 the Financial Stability Oversight Council, which is empowered under the Dodd-Frank Act to identify and monitor excessive risks to the financial system of the United States, designated eight privately owned financial market utilities as systemically important. They included the Clearing House as operator of CHIPS, a private sector Real Time Gross Settlement (RTGS) system, CLS Bank, the Depository Trust Company, the Chicago Mercantile Exchange, ICE Clear Credit, and the Options Clearing Corporation.

Although the wholesale services operated by the Reserve Banks were not formally designated as systemically important, the Board of Governors of the Federal Reserve committed to hold us to "as high or higher a standard" as it holds these private sector utilities.

This is appropriate as many of these systemically important financial market infrastructures have a critical dependence on the availability of our wholesale services in their daily operations to fund, de-fund and settle positions derived from transactions in other markets. The inverse is not necessarily true. In practice, the wholesale services operated by the Reserve Banks are the base of a pyramid on which all other systemically important infrastructures – and, indeed, the financial system of the United States as a whole - ultimately rest.

Thirdly, our role as central securities depository (CSD) and fiscal agent. As the CSD for over $70 trillion in par value of Fedwire-eligible securities, the Fedwire Securities Service functions as the central repository for the largest, deepest, and most liquid pool of collateral in the world. Moreover, in support of the fiscal agent responsibilities of the Reserve Banks, the Fedwire Securities Service facilitates the issuance, maintenance, and redemption of all Fedwire-eligible securities, performing an indispensable role in financing the operations of the United States government and those of other issuers.

Fourthly, our support for the execution of monetary policy. The wholesale services function as the platform across which the Federal Reserve ultimately settles its monetary policy operations.

Any one of these four elements would likely qualify the wholesale services as "systemic". In the aggregate they represent a staggering portfolio on which the execution of the fiscal and monetary policies of the United States absolutely depend. A wholesale service outage, or even a meaningful disruption that impairs public confidence, represents a risk to the United States with profound, and potentially unpredictable, consequences, for which the only appropriate policy response is "failure is not an option."

Flaws in the historical approach to resiliency and security

Since 9/11, consistent with industry best practice, we have sought to fulfil that resiliency mandate through dispersal of infrastructure and human capital. We have invested considerable resources to ensure operational redundancy through geographic dispersion of data centres and operating sites, real-time data replication, and split operations. These measures have yielded significant resiliency dividends, particularly against physical threats, and deserve to be heralded.

While geographic dispersion of infrastructure and human capital remains an indispensable prerequisite for responding to physical threats, and is likely sufficient for most contingency scenarios we face, it no longer suffices as the central organising paradigm for resiliency in the wake of the escalating cyber-threat. Global realities compel a paradigm shift in how we contemplate the resiliency and security of systemically important infrastructure. To borrow the vernacular of our supervisory colleagues, we must prepare for "extreme but plausible" events.

"The wholesale services operated by the Reserve Banks are the base of a pyramid on which all other systemically important infrastructures – and, indeed, the financial system of the United States as a whole - ultimately rest."

- Richard P. Dzina,
Executive Vice President and
Head of the Wholesale Product
Office of the Federal Reserve
Bank of New York

Consider, for example, a cyber-breach of perimeter security, resulting in the insertion of pernicious malware, a severe data corruption in which confidence in account balances is compromised, or even an application failure that propagates itself almost instantaneously across primary, secondary, and tertiary operating sites. An unfortunate by-product of instantaneous data replication, such a scenario risks rendering a systemic infrastructure functionally inoperable.

Aggravating the cyber-challenge, and in contrast to traditional resiliency scenarios, is the likelihood of facing an adversary that can anticipate and adapt to our contingency response in real-time. Moreover, the nature of the challenge is asymmetric. We must defend across an extended front, while the adversary need only find a single point of entry or vulnerability. These dimensions add a dynamic to resiliency planning we have not previously contemplated.

A new approach to deal with new kinds of cyber threats

In recognition of these escalating threats, the Committee on Payment and Market Infrastructures (CPMI) and the Board of the International Organization of Securities Commissions (IOSCO) recently published a consultative report providing guidance on cyber resilience for financial market infrastructures.[1]

The guidance is designed to supplement the Principles for Financial Market

Infrastructures, published by CPMI-IOSCO in April 2012.[2] It is unequivocal in its expectation that financial market infrastructures (FMIs) establish an objective of resuming critical operations within two hours of disruption, even in the case of extreme events, and regardless of whether they are cyber- or physical attacks.

For most infrastructures, this expectation remains aspirational. However, just as FMIs responded to the post-9/11 supervisory guidance to improve their resilience to physical threats by geographic dispersion of infrastructure and human capital, so will they respond to the current advice on raising their defences against cyber-attacks. There

1 The Committee on Payment and Market Infrastructures (CPMI) and the Board of the International Organization of Securities Commissions (IOSCO), *Consultative report, Guidance on cyber resilience for financial market infrastructures*, November 2015.

2 The Committee on Payment and Settlement Systems and the Technical Committee of the International Organization of Securities Commissions (IOSCO), *Principles for financial market infrastructures*, April 2012.

John Hagon

Head of Global Operations, CLS

Richard Dzina is right. The security methods adopted after 9.11, which focus on real-time data replication and geographical dispersal of people and premises, are effective against physical threats. But the same approach may not be effective against cyber-threats, such as code and data corruption.

For example, if data or code is corrupted, the corruption will likely be replicated at ancillary sites. One way to mitigate that risk is to invest in a separate operating code and database and run them at a third site alongside the existing centres. This requires investment and maintenance costs that are challenging for many institutions.

Our clients rely on the liquidity management, multi-lateral netting and settlement optimisation mechanisms provided by our system, and we cannot ask them to switch at short notice to a platform which offers only some or none of these services, or provides them in a different way.

A possible solution to this dilemma that we are considering would be to host separate versions of our data and code at a third site. The code would always be identical to the version behind the live system, and the data would be replicated at pre-defined intervals, rather than in real-time, allowing us to re-start transaction processing with data drawn from a point prior to its corruption.

One further step we have taken already to address the risk of cyber-threats is to monitor our service for signs of abnormal behaviour by clients. By mapping the current activity of our clients against their past behaviour, we can detect anomalous and potentially malicious conduct, which could be indicative of the presence of a cyber-threat, in real-time, and ensure it is addressed.

As an industry, I believe there is more that can be done. Financial systems are extremely inter-connected, but our systems will be stronger and more likely to maintain the same levels of resilience in the face of a cyber-attack if we work together, where beneficial and appropriate, rather than working in isolation.

While this poses its challenges, the industry recognises the benefits of collaboration. The sharing of ideas and reduced costs are two such examples. Ultimately, our collective goal is to strengthen the international financial system and meet the two-hour recovery time objective (RTO) recommended for financial market infrastructures by the Committee on Payments and Market Infrastructures (CPMI) and the International Organisation of Securities Commissions (IOSCO).

Two hours is an achievable objective in the wake of a physical attack. To maintain the same level of resilience in the face of a cyber-attack, we, as an industry, need to develop technological solutions to be able to identify rapidly the occurrence of a disruptive cyber-event, invoke contingency mechanisms that recover to an acceptable point in time, and resume operations within the two hour RTO.

While challenging, the industry is working towards achieving this goal. For example, a great deal of informal information sharing takes place already and greater collaborative efforts can only serve to improve cyber-security as a whole.

It is crucial to prepare and anticipate potential weaknesses in a system, and address issues as quickly as possible with minimal impact to clients. This is an area CLS is paying careful attention to – particularly in relation to detection and recovery.

Cyber-criminals are smart and becoming increasingly sophisticated. By working together to strengthen the international financial system, the threat of an attack can be reduced. The CLS model demonstrates what the industry can achieve through sound technological investment and industry co-operation.

While the investment and maintenance costs required to protect an institution against cyber-threats are challenging in the current, cost-constrained environment, it is necessary, and we cannot allow complacency to creep in.

A balance has to be struck between the mitigation of risk and expenditure on its management. The best place to strike that balance is by spending on detection and recovery, not the chimera of complete protection from cyber-attacks.

is already considerable collaboration within the industry to identify alternative solutions that can accelerate recovery from attacks, make their deployment more cost-effective, and strengthen not merely the resiliency and security of individual components, but the system as a whole.

All FMIs are making efforts to enhance perimeter security, isolate critical applications, rotate more nimbly across data centres guard against insider threats, and bolster detection and readiness. But the central question for FMIs, as they devise their cyber-security strategies, is third site capacity.

Yves Poullet

Member of the Group Management Committee and Head of Corporate Technology (CTO), Euroclear

Cyber-threats could have just as great a negative impact as the positive potential of the digital revolution. Such threats need to be treated as a strategic issue of the highest priority. Because they are systemically important, financial market infrastructures (FMIs) have a particularly heavy responsibility to maintain a degree of cyber-resilience that reduces the risk of an extreme scenario to infinitesimal proportions.

The first step in defending against cyber-threats is to invest in capabilities that reduce the likelihood of such scenarios happening. In doing so, key infrastructures face the difficulty that the 80:20 rule of management does not apply to cyber-controls. Financial market infrastructures (FMIs) need to prepare for every eventuality.

Investing in the right tools is only half the solution. Ensuring strong awareness of threats, and adherence to policy, and having staff using the tools at their disposal correctly, also minimises cyber-risk. Sophisticated cyber-defence mechanisms can easily be undermined if strong discipline is not applied in standard cyber-controls.

That is why, at Euroclear, we review our cyber-security programmes constantly, provide ongoing employee education on cyber-threats, and now stress test our cyber-security methods and procedures via covert but controlled hacking exercises (so called “Red Team Exercises”). These activities provide a constant evaluation of our standard defence mechanisms, enabling us to strengthen our defences in the face of a continuously evolving threat.

Standard defence mechanisms are used by every FMI. They include safeguarding the perimeter surrounding technology and data assets, early detection of threats, and rapid response and recovery. Such defences need strong governance to ensure measures and counter-measures adapt to constantly changing threats, and that employees, suppliers, customers and business partners maintain a high degree of awareness about the need for cyber-security.

Every market infrastructure also has extremely strong business continuity plans. At Euroclear, for example, we maintain three separate data centres, which gives us the ability to recover from many scenarios by switching production between them. Such measures are effective against physical threats such as fire, flood and terrorist attacks, but counteract a range of cyber-threats too.

However, defences based on instant replication of data can also exacerbate the consequences of a cyber-attack, by reproducing in other systems the malware or breach infecting one. For such “extreme but plausible” scenarios, this risk could be mitigated by maintaining an entirely separate data centre.

But it is not yet clear that this is the right approach. A separate system still requires the original data and the applications to make it useable, and must be fully tested on a regular basis to ensure it can support all the services required when it is activated. An alternative or complementary option to overcome a severe data corruption scenario is to have closer engagement with market participants so that data can be reconstructed from the records of the daily reconciliation process.

Guidance from the Committee on Payments and Market Infrastructures (CPMI) and the International Organisation of Securities Commissions (IOSCO) limits the time allowed to effect such data reconstruction to just two hours. Regulators and market participants expect a systemically important institution to keep maximum down-time within this very short window, and that is perfectly reasonable. The CPMI-IOSCO guidance stipulates that the deadline must be met even in extreme scenarios, but it does recognise the scale of the challenges key infrastructures might face in meeting it.

Chief among those challenges is the potential latency in detecting the cause of the cyber-incident, as in the case of an “advanced persistent threat.” With most traditional operational incidents, it is possible to pinpoint the exact time at which the issue occurred, providing certainty that data processed or applications used prior to the event are not corrupted. An advanced threat that sits inside systems for months, or even years, makes it much harder to determine what damage was inflicted, and when. Without that certainty, it is hard to be confident of the reliability of any data set or application, and so impossible to predict when a service can safely be resumed.

In those circumstances, it may be prudent to take more time to determine when the breach occurred, rather than risk resuming activity with corrupted data by rushing to meet a two-hour deadline. Designing and testing systems and processes to ensure resumption within two hours is an excellent aspiration, but it is important to take the specific circumstances of a breach into account when deciding whether it is safe to do so.

The inter-connectedness of financial markets, which increases the risk of cross-contamination, is a strong argument for putting safety first. It also points to greater collaboration between infrastructures and market participants, to exchange information about how to detect and recover from attacks. That collaboration is happening already, in both formal and informal ways, but FMIs and the authorities should be looking to intensify those efforts.

Trevor Spanner

Chief Operating Officer and Group Risk Officer, Hong Kong Exchanges and Clearing Limited

Physical boundaries are an important but not a sufficient form of defence against rapidly mutating cyber-threats. Today, actionable intelligence about upcoming cyber-attacks, and pooling of techniques to defend data and systems against them, matter a lot more than physical barriers. Intelligence of that kind requires the formal sharing of information not just with the authorities but with other financial market infrastructures (FMIs). In fact, collaboration and agreement between FMIs on cyber-security standards ensures that we do not replicate investments in a wasteful manner.

A great deal can also be gleaned from testing defences with ethical hackers, who track the evolution of threats. Informal communication with other businesses, including those outside financial services, helps too. The attack mechanisms used by cyber-criminals are rarely specific to FMIs, and trust-based collaboration can save a great deal of time and money. Trust is not easy to build, however, because businesses are understandably reluctant to share details of attacks which might expose their vulnerabilities. Nevertheless, a network of trusted relationships is a far more effective defence against cyber-threats than any amount of physical security.

We live in a connected society, so there are by definition digital bridges that will traverse any physical perimeter that surrounds an asset. Cyber-security measures have traditionally concentrated on the gateways to those bridges. However, they have to control the operator of the gateway, know who is entitled to cross the bridge, check the credentials of everybody who wants to cross it and – an issue of increasing importance - monitor their activity once they have crossed the bridge and are inside the perimeter.

There is a reason why the black market price of a social media profile is many multiples that of a credit card holder. A cyber-criminal can do much more damage with a credible social media identity than a stolen credit card. Even a cursory glance at the social media accounts of employees proves that they are more open to sharing information than security specialists would prefer. Software developers, for example, share information about the types of code they are working with, which is extremely useful to cyber-criminals looking for ways to access systems. It follows that ensuring everybody working for an organisation is mindful of the risks they create when posting material on social media is one of the investments FMIs have to make.

Clearly, the question is not whether to spend money, but how much, and in which area. Investment has to be commensurate with the risks to the organisation, but any cost-benefit analysis has also to recognise two important differences from normal return on investment calculations. The first is that the key test of a successful cyber-security investment is negative: nothing untoward happened. In this sense, purchasing security is more like insurance than investment. The second is that cyber-security investments inevitably have a shorter lifecycle than traditional investments, because cyber-threats evolve at least as fast as digital technology. Historically, we have focused our spending on preventative measures, but increasingly we are spending more on detection and response.

The framework we use for assessing cyber-threats aims to ensure any cyber-security investment is proportionate to our risk appetite. A good example of disproportionate investment is a completely separate system and site to meet an artificial deadline of restitution of service within two hours of a denial of service attack. It is simply too difficult to predict the origins and consequences of a cyber-attack to offer that guarantee. But in less unpredictable circumstances, such as loss of premises or power to fire, flood, internal sabotage or a terrorist attack, real-time replication of data means a recovery time of two hours is realistic.

Assuredly, Hong Kong Exchanges and Clearing Limited works to that expectation already. We certainly do at LME Clear, the clearing house for the London Metal Exchange, for example, where the Bank of England has specified a two hour limit on down-time. Achieving it does necessitate an alteration in procedures. When we implement real-time systems, we simultaneously change the way we process, store and grant access to data. By making those procedural changes, we also alter the mindset of the people working for us. Mindfulness – of prevention, detection and response - is definitely our best defence against a successful cyber-attack.

Historically, third site solutions rely on data replication schemes designed to restore critical functionality after primary and secondary data centres are lost. This approach looks increasingly inadequate in the face of an escalating cyber-threat.

Increasingly, FMIs need to contemplate technologically diverse, off-network third site solutions that offer an impregnable firebreak, and a platform for recovery, if the core of an application suite or data set becomes corrupted.

“It may even be preferable for FMIs to develop alternative solutions, to avoid unintended concentration risk or an unhealthy measure of “groupthink”. ”

- Richard P. Dzina,
Executive Vice President and Head of the Wholesale Product Office of the Federal Reserve Bank of New York

A third level of cyber-security instead of a third site

One day perhaps we will refer to these solutions as “third level” rather than “third site,” reflecting the fact that technology is increasingly liberating us from the physical limitations of data centres, and freeing us to consider instead “metaphysical” alternatives, such as cloud or hosted solutions. However compelling the prospect, a technologically diverse third level of resiliency nevertheless raises several important questions.

Where, faced with increasing costs and diminishing returns, should an FMI draw the line on resiliency? How much insurance is enough when the odds of invoking a technologically diverse third level of resiliency may be remote, but the costs of a severe disruption from which recovery is impossible are incomprehensibly large? How can an FMI ensure the integrity of its data and software when it resumes operations after its core components are compromised? For how long should an FMI be prepared to operate in a degraded mode, and how should that assumption inform the business requirements for critical third level functionality?

FMIs will likely respond differently to these questions. They will also likely devise different technical solutions to the two hour resumption challenge set by CPMI-IOSCO, reflecting their unique circumstances and their respective assessments of the likely threats. It may even be preferable for FMIs to develop alternative solutions, to avoid unintended concentration risk or an unhealthy measure of “groupthink”. There is no need to prescribe that a common solution be applied universally across all FMIs, but

Stephen Gilderdale

Head of Customer Security Programme, SWIFT

Dispersal of sites, staff and data are measures typically deployed to ensure an infrastructure remains constantly available. Such measures help, but are not sufficient to ensure robust and comprehensive cyber-security. Provided back-up sites are logically separate, and physically secure, threats become more difficult to introduce across multiple sites. However, financial market infrastructures (FMIs) must give additional thought to cyber-security beyond traditional, availability-led thinking.

That is why best practice, and increasingly regulation, demands more. For example, tight control and authentication of access to facilities and systems (both logical and physical), thoughtful segregation of networks, encryption of data (in-flight and at rest) and measures to enforce integrity of data and software at all levels.

Whilst equipping back-up sites with an alternate technology stack is often considered a strong form of protection against targeted threats, such an approach clearly increases costs and can even degrade the risk outlook; both staff and customers must remain trained and familiar with the operation of an alternative system that is rarely used.

Of course, prioritisation of cyber-security measures remains risk-based. Financial institutions are well-practised at balancing risk versus benefit, and few today judge a wholly separate technology platform as a top priority. Nevertheless, the continually evolving threat landscape will surely drive FMIs to re-evaluate their position and look for ways of further diversifying their technology deployments.

But even the most rigorous preparations and imaginative defences cannot eliminate the risk of a breach. Equally important is the readiness of FMIs to respond fast in the event of a cyber-attack. Effective response testing must engage market participants, so that cyber-security teams can collectively practise their co-ordinated response to an attack.

Better collaboration can help. Cyber-criminals invest in attack mechanisms, and often look to increase the return on those assets by selling them to others. It follows that pooling information and intelligence between institutions will reduce the chances of multiple FMIs succumbing to the same attack vectors. In the United States, for example, information-sharing on cyber-threats between private sector firms is promoted by Executive Order.

However, collaboration can take many forms, and fragmentation makes effective cyber-intelligence management more complicated. Details of threats are disseminated by automated systems as well as by commercial forensics firms. In addition to national and regional Computer Emergency Response Teams (CERTs) and industry-based Information Sharing and Analysis Centres (ISACs), a great deal of informal collaboration takes place between security officials at individual firms. Furthermore, concerns about the distribution and use of information can deter some organisations from submitting valuable data in the first place.

FMIs are well placed to help. Market infrastructures are natural entities with whom participants can share intelligence; they can help create shared solutions and, as a consequence, minimise the associated cost of defence for the industry. Market participants look to FMIs for highly available and resilient shared services – strong cyber-security is key, and perhaps they should also look to FMIs to play a larger role here too.

there is an onus on all FMIs to reflect on how best to respond to an issue of fundamental importance.

The new CPMI-IOSCO cyber-guidance also exhorts FMIs to develop contingency plans for events in which they fail to resume operations within two hours. Both in the Wholesale Product Office and across the Federal Reserve System we are considering remedial actions to mitigate customer and market impacts in the event of a wholesale service disruption from which we cannot recover on a same day basis.

This work proceeds on multiple fronts, including analysing and parsing our transaction flow to identify systemically important activity, and exploring alternative routes to process that activity via other channels and service providers. Later this year, we will be conducting table-top exercises with systemically important customers and FMIs to test our hypotheses and procedures.

But we are not deluding ourselves. No matter how mature our framework for responding to protracted outage scenarios, no matter how sound our procedures, and no matter how tested our protocols, we would never want to rely on such measures. Our real objective is to invest in resiliency and security measures that ensure that we never find ourselves in such a position.

The elements of a new resiliency and security paradigm

What would such a set of measures actually look like? As a former Army officer, I counsel against constructing a Maginot line so inflexible that its rigidities are easily subverted by a creative and nimble adversary. We should aim instead to develop a coherent and integrated system that relies upon all of the classical elements of defence, but depends on none of them exclusively.

We need perimeter security to keep the adversary outside of the environment; defence in depth to safeguard our most critical assets; sophisticated intelligence to understand the tactics of our adversaries; robust surveillance to monitor for intrusion and ensure the integrity of the environment; rapid response to fend off attacks; effective collaboration with allies to enhance collective security; and a strategic reserve to respond deftly in the event of loss.

Combinations of measures of this kind do more than enhance security and resilience. They also provide an extremely effective deterrent by raising the costs our adversaries must bear to perpetrate a successful cyber-attack. In protecting the wholesale services of the Federal Reserve Banks, we aspire not merely to a commercial standard of resiliency, or even to a supervisory standard, but to something approaching national security grade. In this sphere, either intentionally we are progressing or inevitably we are regressing: there is no idleness.

“Cyber-resilience in a changing world” at Sibos

Monday 26 September 2016
09:00 - 10:00
Conference Room 2