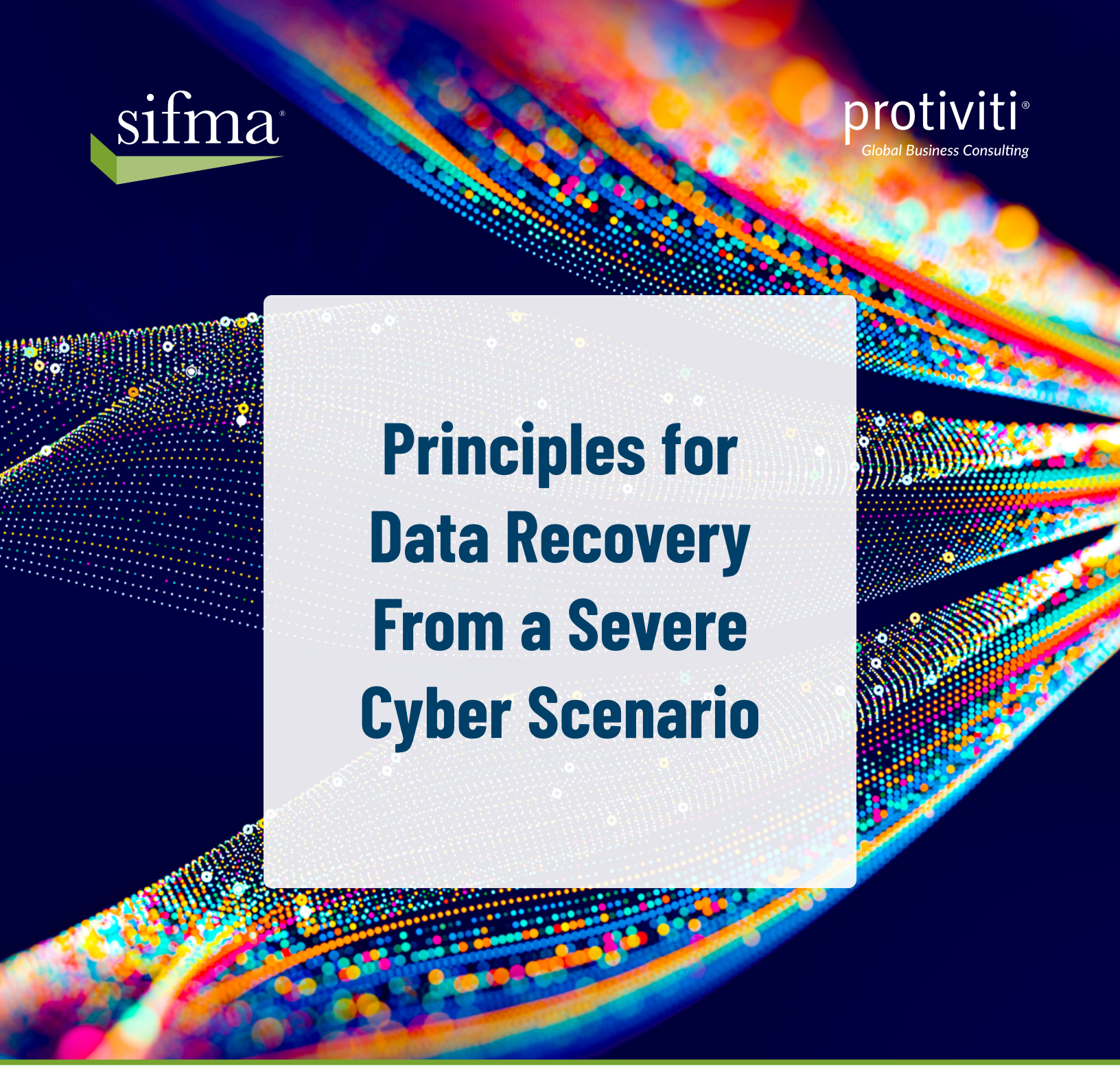

Principles for Data Recovery From a Severe Cyber Scenario

Table of Contents

Overview

Financial institutions build and sustain capabilities to mitigate the impact of events that may compromise the confidentiality, integrity or availability of firm and customer data. As part of this process, financial institutions plan and exercise how they would respond to an extreme-tail event such as a highly destructive cybersecurity incident so as to mitigate harm to financial markets, counterparties, customers and the investing public. Regulatory agencies around the world are similarly focused on the resilience of an institution's critical operations during and recovering from a potential disruptive event.

This paper is intended to prompt increased dialogue between financial institutions, trade associations and regulatory authorities on a rapidly evolving topic. It lays out a set of principles that could align regulators, the financial sector and all three lines of defense within an organization to a cohesive view of resilience. A key objective of this paper is to highlight the challenges in meeting regulatory obligations during extreme cyber events that result in data corruption.

Meeting today's regulatory mandates may be aspirational, and the goal of the financial institution is to ensure that firm and customer information is not at risk. If not implemented safely, rapid recovery based on mandated regulatory guidelines could harm investors, a firm's ability to service their customers and, potentially, financial stability across the sector.

Regulators should support industry resiliency and recovery practices that strive for a safe but rapid recovery, recognizing that firms and regulators have a shared interest in recovering critical operations as quickly as possible, but only if done in such a way that will not result in further harm to the firm or financial markets.

Regulatory Drivers

The introduction of operational resilience regulations in multiple jurisdictions has prioritized the ability of institutions to recover from severe but plausible events. In particular, the concepts of impact tolerance and maximum tolerable period of disruption (i.e., the point beyond which the impact of an outage is unacceptable) have renewed the industry conversation around the feasibility of meeting, for example, a two-hour recovery time objective (RTO) under certain scenarios. This expectation remains current among regulators. For instance, the International Organization of Securities Commissions (IOSCO) recently emphasized the expectation that financial market infrastructure (FMI) institutions resume operations within two hours of a disruption, including an extreme cyber attack.[1]

Institutions should strive for a safe but rapid recovery rather than a mandated RTO that may ultimately harm the institution, its customers and the financial sector. The Bank of England recently published the results of its inaugural cyber stress test in which it acknowledged that "there might be instances where the disruption caused by an incident was such that, despite prior planning, attempting to recover by the end of the value date could have a more adverse impact on financial stability than failing to do so."[2]

Mandated recovery times that do not contemplate recovery feasibility or practicality under a range of disruptions may drive significant time and investment into aspirational rather than achievable results or may force institutions to consider meeting regulatory RTOs versus addressing disruptions safely and effectively.

RTOs mandated by regulation play a significant role in influencing how financial institutions prioritize their mitigation investments and resource allocations. For example, some institutions may be driven to prioritize investment in recovery capabilities that result in little practical improvement over investments in their security and control environment that could more meaningfully reduce the probability of a disruptive event. Alternatively, the industry may be directed toward technical recovery solutions when collaborative actions to improve coordination and management in the event of a major incident would be more effective mitigants. Consequently, the principles set forth below along with a risk-based approach to data resilience and recovery may ultimately address regulatory concerns and best serve to support the continued resilience of the financial services industry.

1 Implementation Monitoring of the PFMI: Level 3 Assessment on Financial Market Infrastructures' Cyber Resilience, Bank for International Settlements and IOSCO, November 2022, https://www.iosco.org/library/pubdocs/pdf/IOSCOPD723.pdf.
2 Thematic Findings From the 2022 Cyber Stress Test, BoE and Prudential Regulation Authority, March 29, 2023, https://www.bankofengland.co.uk/-/media/boe/files/prudential-regulation/letter/2023/thematic-findings-2022-cyber-stress-test.pdf.

Deterministic Versus Nondeterministic Recovery

Recovery from a significant cyber incident, in particular an incident that renders data corrupt or unavailable, will have significant variability given that the recovery process typically includes deterministic (i.e., fixed) and nondeterministic (i.e., variable and event-driven) dependencies. While some elements of recovery can be predetermined, tested and improved (e.g., restoring from bare metal), other elements are determined by the incident (e.g., the time needed to identify and remove a malicious actor from an institution's environment). Moreover, the time needed to restore data following a destructive data event will always vary based on the extent of data loss and unknowable details about the status of the environment to which the data is being restored.

RECOVERY TYPES

Deterministic	Nondeterministic
Definition	
Recovery time fixed and not dependent on event type and/or event severity	Recovery time variable and dependent on event type and severity
Example	
System rebuild	Ensuring that systems are clear from a cyber event
Testing Needs	
Effective understanding of recovery time	Relative understanding of recovery time
Recovery Time Impact	
Tested recovery times essential	Tested recovery times aspirational

In a scenario that involves deterministic and nondeterministic elements, a generic recovery-time mandate cannot effectively be proven, as the breadth of impact can vary greatly. The appropriate emphasis in recovery, therefore, needs to be on capabilities that continuously mature, effectively accelerating recovery against a variety of scenarios. Arbitrary and prescriptive goals could result in a rush to premature recovery and inflict even greater market damage than would have been inflicted absent those goals.

Institutions can partly account for the nondeterministic elements of recovery by improving and shortening the time required for the deterministic elements. The objective is to create an ever-larger delta between the impact tolerance and the time required for the deterministic elements of recovery, thereby increasing the time available to address the nondeterministic elements such as incident assessment and data reconstruction. In addition, institutions can use scenario-based testing to exercise the skills and capabilities needed for recovery, allowing them to validate assumptions, develop muscle memory and identify areas of improvement.

However, while assumptions can allow for application of testing deterministic elements of recovery, testing of nondeterministic components cannot replicate recovery times experienced in an actual event. As a result, certain assumptions are necessary when setting recovery time frames (i.e., RTOs and RPOs).

Therefore, when developing recovery capabilities for an extreme cybersecurity incident, impact tolerances take on the role of aspirational guides to drive investment and risk-decisioning, rather than prescriptive recovery times that cannot currently be met during certain events.

Resilience Principles

The following principles represent a prudent approach to managing the impact of a significant data event caused by an extreme cybersecurity incident, in alignment with operational resilience regulatory expectations and an institution's risk appetite.

1. Risk Management

Resilience is the outcome of effective risk management, and the two cannot be decoupled. Regulators should encourage financial institutions to maintain a holistic view of risk that appropriately balances resilience decisions between prevention and recovery.

Decisions and investment in resilience should be risk-based, as it is not feasible to ensure recovery in a prescribed time frame for every scenario. Institutions strive to prevent, detect, respond to and recover from critical service disruptions in a risk-based fashion. Opportunities to mitigate, transfer and accept risk should be continually evaluated.

- Institutions' resilience and risk-tolerance decisions should be based on business-service criticality and the projected investment necessary to mitigate risk within the institution's risk appetite.

- While resilience efforts have historically focused on the recovery of a service and the realization of a resilience event, they should not be decoupled from an institution's overall risk and control frameworks. Creation of new governance or frameworks is likely to add complexity and result in siloed resilience risk management.

  Example: *At some point, the marginal gain in technology investment may be better spent on risk prevention. While this prevention spending may not be clearly articulated as enhancing the institution's resilience, it will ultimately increase the overall risk posture of the institution.*

- We have serious doubts that zero data loss is achievable in any data recovery event. Institutions may be advised to invest in other areas that can reduce the systemic nature of an event and minimize customer harm. Institutions should have the freedom to recognize the trade-offs of their approach and reasonably invest in this reduction of both risk and resilience toward this goal.

  Example: *In any data-center placement, there is a trade-off between latency, cyber risk reduction (a consequence of increased latency), environmental risk and other factors. Institutions recognize these factors and make appropriate risk-adjusted decisions based on their corporate profile and risk tolerance.*

- The process of moving from a challenged state to business as usual (BAU) is implemented in stages. Recovery of a service should be based on an institution's ability to deliver that service with a reduced set of capability, rather than a full return to BAU in a prescribed time frame. Expectations for a full BAU recovery within a regulatory mandated RTO may create incentives to prioritize fast rather than safe recovery.

2. Recovery Objectives

As there are an infinite number of scenarios that may impact a firm's critical business services, financial institutions should take a risk-based approach to setting impact tolerances and recovery targets that account for the risk profile of the business service, its underlying applications and the controls it has in place.

Impact tolerances should be applied consistently across disruption event types regardless of cause and outcome. Distinct measures based on event type or at system levels add complexity that may not strengthen the ability or capacity to recover or report on recovery. Recovery metrics should be managed at the business-service level, rather than at the application level, with references to asset-level metrics as appropriate.

- Traditional recovery point objectives (RPOs) and RTOs are not effective measures to use during a data-corruption event. During a significant cybersecurity incident, impact tolerances also become aspirational and should be used as guides for recovery.
- Alternative procedures, contingency plans and minimum viable products and services, and by extension minimum viable datasets, should be considered when assessing the recovery of a service. Service capability, rather than a full return to BAU, should be the basis for recovery prioritization.
- It is not possible to delineate all variables in severe scenarios or definitively draw lines between plausible and what's theoretically possible. This divergence means that precision in estimating recovery times should be deemphasized for adaptive planning and targeted maturation of recovery capabilities. A rational risk-adjusted approach, including threat-vector analysis, should be used in determining investments and priorities.

3. Product, Asset and Capability Types

Financial institutions should use reasonable threat vectors to determine the plausibility of extreme scenarios for the purposes of planning and testing resilience.

Threat profiles for applications, internal critical infrastructure, data and third parties vary for a number of reasons. Asset threat profiles and their deterministic and nondeterministic recovery profiles should be considered when assessing and regulating an institution's resilience posture.

- Applying threat-vector analysis to specific infrastructures can further calibrate risk, prioritize investment and drive risk-based segregation.

  Example: *While mainframes are equally vulnerable to compromise in institutions with weak credentialing, security and encryption practices, they are not robust targets for ransomware attacks and other operating-level system exploits, which more routinely impact distributed systems. Similarly, external-facing systems inherently offer an attack vector that is not present in closed internal systems.*

- Threat profile considerations are appropriate as institutions continue to mature their operational resilience programs, which should be used to distinguish extreme but plausible scenarios from scenarios that are possible but are not supported by precedent or reasonable probability.

4. Testing

Financial institutions should test outcomes, rather than an exhaustive list of hypothetical scenarios, and then extrapolate the findings across a broad range of scenarios.

Firms today use a risk-based approach to resilience testing to determine an institution's ability to address likely disruption-event impacts (e.g., data corruption). However, there are limitations to all types of testing; no single test can comprehensively or perfectly validate recovery capabilities against a live threat. Therefore, testing should be aligned to the unique characteristics of an institution's critical operations' threats and risks.

- Testing programs should account for the full range of threats facing the institution and take a risk-based approach to choosing an impact to test and how frequently a given scenario is used. The approach could include systematic sampling to ensure that critical components are tested. Testing should be initially focused on recovery from various impacts and then test against a maturing range of scenarios.
- While the number of event scenarios that an institution could test is infinite (e.g., every cybersecurity incident or ransomware event will unfold in a different way), the impacts of events are finite (e.g., data unavailability). As such, impacts are more efficient and effective to test, as they can be applied to multiple scenarios. It follows that the most effective approach for institutions to build and measure recovery capabilities is impact-driven, rather than scenario-driven.
- Testing by corrupting live data or production environments is contradictory to control processes. Separate environments (e.g., development, quality assurance) may be preferred; however, they are neither scalable nor sustainable. Simulation exercises involving all relevant parties (e.g., operations, business, technology and control functions) provide better learning opportunities versus precise scenario-configuration recoveries, which at best can emulate a very narrow set of circumstances.
- Third- and fourth-party vendors may approach resilience testing differently than the institution they support. However, third- and fourth-party vendor testing may be relied on provided there is parity in the design and outputs of an institution's tests and those of its respective third- and fourth-party vendors.
- Proxy testing, where the results from the test of a system, application, impact or scenario are used across scenarios to assess an institution's capabilities and reduce testing complexity, should be an acceptable form of testing reduction. For example, it may not be necessary to test every database if each of them utilizes the same build. Instead, an institution would be better off testing a sample set on a rotating basis and reallocating resources to test other assets.
- Even for the largest institutions, resources are not infinite and other risks must be accounted for in testing and exercising programs. Impact and likelihood risk must be the primary lens through which the selection of tests, and the assets or services being tested, are chosen. Supervisors and first-, second- and third-line teams should seek to understand and verify the methodology for selection. An approach that advocates for an ever-larger set of tests covering a more extreme set of scenarios is likely to create significant inefficiencies in the institution's overall resilience efforts and possibly detract from more beneficial activities.

5. Reporting

Financial institutions should use a risk-adjusted set of metrics to report resilience that addresses the need for the board to understand probable recovery capabilities of critical business services under extreme but plausible scenarios.

Recovery-capability reporting should be risk-based and consistent across event and recovery types. Each institution should be able to articulate clearly how its internal taxonomy aligns with the varying regulatory terminology to which it will have to comply.

- Management and board reporting should be differentiated, in alignment with broader corporate-governance principles. It is unsustainable and unrealistic to provide and review service-level detail across management layers. Reporting should be multidimensional, including relevant facts (e.g., physical capabilities) and details of recovery capability at the service level.
- Representation of risk and the level of detail required by regulatory bodies should be harmonized or consistent and be based on the importance of the service.
- While the importance of operational resilience is recognized, institutions need to be able to balance the reporting they provide to their board with a wide range of other operational and financial risks to resilience. Only by allowing flexibility to the institution to determine the suitable level of detail to provide directors and executives can resilience subject matter experts be sure that boards and senior management are able to make appropriate risk decisions. These decisions should reflect an understanding of the materiality of the service and an appreciation for the broader risk landscape in which their institution operates.

Conclusion

Financial firms and regulators have a shared interest in recovering critical operations in a safe and effective manner. Setting aspirational recovery-time objectives and impact tolerances that do not balance safety and speed in recovery may, in some instances, create more risks to financial institutions, the investors they serve and the sector at large.

The principles outlined above will help firms and regulators determine what is achievable during extreme events and set appropriate risk-based expectations for testing, reporting, resiliency and recovery from extreme events.

SIFMA is the leading trade association for broker-dealers, investment banks and asset managers operating in the U.S. and global capital markets. On behalf of our industry's nearly 1 million employees, we advocate for legislation, regulation and business policy, affecting retail and institutional investors, equity and fixed income markets, and related products and services. We serve as an industry coordinating body to promote fair and orderly markets, informed regulatory compliance, and efficient market operations and resiliency. We also provide a forum for industry policy and professional development. SIFMA, with offices in New York and Washington, D.C., is the U.S. regional member of the Global Financial Markets Association (GFMA).

sifma.org

Protiviti is a global consulting firm that delivers deep expertise, objective insights, a tailored approach and unparalleled collaboration to help leaders confidently face the future. Through our network of more than 85 offices in over 25 countries, Protiviti and its independent and locally owned Member Firms provide clients with consulting solutions in finance, technology, operations, data, analytics, governance, risk and internal audit.

Named to the 2023 *Fortune* 100 Best Companies to Work For® list, Protiviti has served more than 80 percent of Fortune 100 and nearly 80 percent of Fortune 500 companies. The firm also works with smaller, growing companies, including those looking to go public, as well as with government agencies. Protiviti is a wholly owned subsidiary of Robert Half (NYSE: RHI). Founded in 1948, Robert Half is a member of the S&P 500 index.

protiviti.com

Contacts

Thomas Wagner
Managing Director
Financial Services Operations
SIFMA
+1.212.313.1161
twagner@sifma.org

Tom Price
Managing Director
Technology, Operations and Business Continuity
SIFMA
+1.212.313.1260
tprice@sifma.org

Douglas Wilbert
Managing Director
Protiviti
+1.917.697.1572
douglas.wilbert@protiviti.com

Andrew Retrum
Managing Director
Protiviti
+1.312.476.6353
andrew.retrum@protiviti.com