Critical Thinking:
Moving from Infrastructure Protection to Infrastructure Resilience

CIPP Discussion Paper Series

Resilience: A Systems Design Imperative

David Arsenault

Arun Sood

January 2007

GEORGE MASON UNIVERSITY
School of Law
CRITICAL INFRASTRUCTURE PROTECTION PROGRAM

Series Outline

The goal of this working paper series is to point out trajectories of the concept of critical infrastructure resilience in theory, policy, and implementation. On the one hand, "resilience" may just be another policy buzzword; but on the other hand, it might indicate a shift in perception and priority of threats, vulnerabilities, and consequences. Indeed, the Critical Infrastructure Task Force (CITF) has recently presented to the Homeland Security Advisory Committee (HSAC) a recommendation to "Promulgate Critical Infrastructure Resilience (CIR) as the top-level strategic objective -the desired outcome- to drive national policy and planning."

Defined as the ability of a system to withstand to and recover from adversity, resilience is increasingly applied to larger social and technical systems. Stress and adversity are experienced not only by individuals and groups, but also by organizations and institutions. In the context of increasing natural and man-made threats and vulnerabilities of modern societies, the concept seems particularly useful to inform policies that mitigate the consequences of such events.

Suggested Citation

Arsenault, D., Sood, A. (2007). *Resilience: A Systems Design Imperative.* CIPP Working Paper 02-07. Arlington, VA: George Mason University.

Resilience: A Systems Design Imperative

David Arsenault

Research Associate
Department of Computer Science
George Mason University
Fairfax, VA

Arun Sood

Professor
Department of Computer Science
George Mason University
Fairfax, VA

Dangerous World, Dependencies, and Fragile Systems

There can be no doubt but that information systems and the networks that connect them have become mission critical to the operations of enterprises, functioning of economies, and defense of nations. Yet these critical information processing systems remain vulnerable to faults, attacks, and their own inherent complexities, despite ongoing global attention to matters of security and availability following the September 11, 2001 attacks on the United States. In fact, The *President's Information Technology Advisory Committee (PITAC)* recently warned that "The IT infrastructure of the United States is highly vulnerable to terrorist and criminal attacks."[2] Other nations' infrastructure systems and those of global enterprises are likely to be equally if not more susceptible to the same phenomenon.

[2] PITAC, Cyber Security: A Crisis of Prioritization, February 2005, www.nitrd.org

Despite decades of concentrated research on computer system and network security, availability, and fault tolerance, our systems remain fragile, brittle, and vulnerable. As modern societies, we depend increasingly on information systems to run financial markets, facilitate international trade, enable global telecommunications, manage transportation networks, and control utilities. The computer hardware and software industries, national research labs, and academic institutions continue with valiant efforts to enable more robust, fault tolerant and secure systems. Some progress is being made indeed; many well-designed technologies exist to address specific systems challenges. However, these solutions tend to be reactive. Security problems become known, they are researched and often viable solutions are found. The resulting new tacit knowledge is then propagated forward into the next generation of hardware and software.

Even a casual survey of information systems literature will reveal a set of core concepts of system design such as availability, fault tolerance, recovery, survivability, reliability, continuity of operation, confidentiality, and integrity. Each of these, when examined more deeply, is a topic in and of itself. Further, many system designs require the combinations of these concept. For example, computer security typically defines “security” as a triad of confidentiality, integrity, and availability—each with volumes of research, methodologies, and commercial products aimed at providing these benefits to information systems. Similarly, the fault-tolerant community deals with issues of reliability, recovery, and survivability, which are vital to systems where life or substantial economic resources are at stake—aircraft control systems, stock markets, banking, or telecommunications, are prime examples.

Fault-tolerance and security, for the most part, are system level concepts. The techniques and technologies are applied in a bottom up fashion starting with the individual data structures, functions, and modules within application programs. Individual application programs are combined either on the same physical machine or through a distributed system approach. Through this layering process we build complex systems and systems of systems. This same approach also build systems that tend to be more brittle than expected with respect to faults or malevolent activities. While typical computer security and fault-tolerance implementations are machine centric concepts, large scale and distributed of systems have design features to ensure continuity of operations, and this requires particular attention to the human dimension. In the face of large-scale destruction, it is not sufficient for the systems to survive, but to maintain continuity of operations a properly trained work force is also essential. Resilience seeks to address these problems by taking a holistic top-down approach—a system of systems architectural approach.

A Systems Approach: Resilience

Why are our systems so brittle on a large scale? Why are they still susceptible to both internal and external faults despite active and diligent research? Why do many types of faults often cause unpredictable effects that cascade beyond the initial point of impact?

The answers to these questions lie in the intersection of three critical characteristics of large scale systems: connectivity, complexity, and interdependence. When these factors are balanced in both the macro scale and the micro scale, it is possible to achieve systems that exhibit what we will define as resilience.

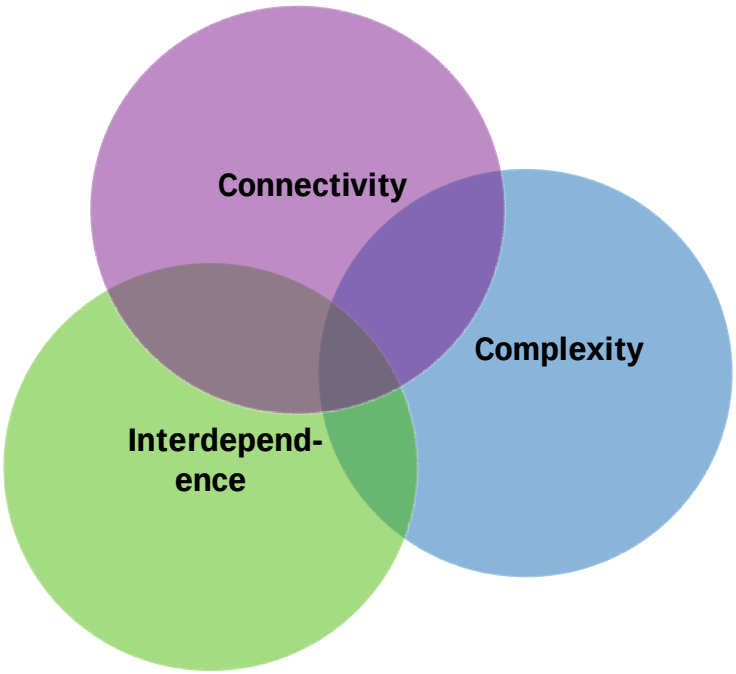

Fig. 1. Resilience is found in the intersection of connectivity, complexity, and interdependence.

Connectivity refers to fact that many individual systems are interconnected via networks to form larger systems of systems. The Internet, the air traffic control system, and power grids all demonstrate this principle. The most important systems used in today's society tend to be systems of systems. It is the interconnections between systems that add to their vulnerabilities. PITAC eloquently captured the importance of system interconnectivity with a simple equation:

Ubiquitous Interconnectivity = Widespread Vulnerability. [3]

[3] PITAC, Cyber Security: A Crisis of Prioritization, February 2005, www.nitrd.org

Connectivity in complex distributed systems, if structured correctly, can also streghten systems and provide additional assurances of robustness under adverse conditions.

Complexity invades our systems at all levels—in the systems of systems (the entire banking "system" or even a single large bank), in individual systems (a foreign exchange trading platform), and in the components within individual systems (operating system, database, application server).

Interdependence stems from how complex our systems have become and the methods we use to deal with this complexity. To build complex systems of systems we break the system down into services and link them together using networks. For complex individual systems such as a transition server or database engine, we separate the functionality within the system into modules and modules into classes and classes into functions and data structures. The common thread here is the interdependence of all of the parts within a system; only when the interdependent parts work correctly will the system produce the intended results such as providing a set of services. Interdependence is both a positive and a negative.

A Multidiscipline View of Resilience

When we speak of resilience we focus on the functions that a system is designed to fulfill—clearing financial transactions, managing airspace, controlling power grids—not the individual components of the system or network. Physics and engineering disciplines define *resilience* as a physical property of materials: *the capacity of a material to absorb energy when it is deformed elastically, and then upon unloading, return this energy.* Ecologists have a more complex view of resilience in natural systems and thus two completing definitions have emerged each emphasizing a different aspect of resilience. One definition, known as engineering resilience, focuses on resistance to disturbance and the rate of return to equilibrium: *resilience is the rate at which a system returns to a single steady or cyclic state following a perturbation.*[4] The other view of resilience, specifically ecological resilience, focuses on state changes in complex systems: *resilience is measured by the magnitude of disturbance that can be absorbed before the system changes its structure by changing the variables and processes that control behavior within the system.* From a human perspective, resilience can be thought of as how well an organization can absorb unex-

[4] For more on definitions of resilience, see http://en.wikipedia.org/wiki/Resilience

pected challenges such as human error, malicious acts (insider threats), or the loss of key human assets.

These different domains each offer something of value when we consider what resilience means in terms of information systems. A multidiscipline composite interpretation of resilience is most useful in our domain as it can and *should* inform us about how to architect systems that are better able to survive and indeed thrive in a dangerous world. Surviving means the system continues to provide the services it was designed to provide. Thriving means the system provides these services at acceptable levels of performance even in the presence of negative conditions.

Domain Models and Events

To further the concept of resilience as a multidiscipline holistic approach to systems architecture, it is useful to examine the models from well-developed domains, such as security and fault tolerance, as well as the types of events that drive these models. The following table provides a mapping between various systems domains and the characteristics we propose for resilient systems.

Table 1. Existing Domain Models Can Support the Resilient Systems Approach

Domain	Domain Models	Resilience Relationships
Fault Tolerance	Fault Models	Complexity, Interdependence
Availability	High Availability Models	Interdependence, Connectivity
Integrity	Integrity Models	Interdependence
Security	Vulnerability Models	Complexity, Connectivity
Resilience (ecological)	Perturbation Models	All Three Resilience Factors
Resilience (materials)	Shock Models	All Three Resilience Factors

Resilience in Nature

When seeking guidance on designing robust, large-scale systems that exhibit resiliency, we must seriously consider the critical lessons that nature has to teach us. There is a promising new science known as *biomimicry* which seeks to "examine nature's models and then imitates or takes inspiration from these designs and processes to solve human problems, e.g., a solar cell inspired by a leaf." [5] Nature has much to tell us about how to build complex systems that work under forbidding conditions.

Interestingly, most systems in nature are systems of systems and are built from smaller components (i.e., cells and organs in mammals) that can exhibit surprising amounts of failure yet the overall system maintains a minimum level of functionality—survival!

> The core idea [of biomimicry] is that nature, imaginative by necessity, has already solved many of the problems we are grappling with. Animals, plants, and microbes are the consummate engineers. They have found what works, what is appropriate, and most important, what *lasts* here on Earth. This is the real news of biomimicry: After 3.8 billion years of research and development, failures are fossils, and what surrounds us is the secret to survival.
>
> – Janine Benyus. [6]

Perturbation models from ecology and shock models from physical sciences can be applied to systems resilience most directly. Perturbation models provide the means to examine complex system behavior over time during periods of stress as sell as during periods of nominal operation. The key question here is how much can a system—in the macro sense, a system of systems—adapt to changes in operating conditions and still function nominally. By comparison, the shock model from materials science provides the framework for measuring and predicting the amount of energy a material can absorb as it is defomed (usually stretched, bent, or compressed) before it fails physically which entails loss of one or more critical physical characteristics.

Central to most system perturbation models is the state of nominal operation of the system and abnormal operation (or failure) of the system with some triggering event or events that cause the system perturbation which at some level of intensity creates a phase change in the system as it transitions from a state of nominal operation to that of the failure condition.

[5] For more on biomimicry see: www.biomimicry.net and *Biomimicry: Innovation Inspired by Nature.* Janine Benyus (HarperCollings. New York. 1997).
[6] www.biomimicry.net/faq.html

Table 2. System Perturbation Triggering Events

Event Type	Cause
Fault-driven	System fault (flaw or failure)
Availability-driven	Network outage, system outage
Security-driven	Exploited vulnerability (known or unknown prior to event)
Human-driven	Innocent error, malicious actions

Clearly other causes for these types of events exist, but the idea here is to classify and characterize the main types of events that can adversely impact a system. It is these types of events rather than specific details of a set of events that a resilient system must be architected to handle.

Resilience by Design

Functionality-Performance Tradeoff Curves

We define a *functionality-performance tradeoff curve* (FPT Curve) to describe the behavior of a resilient system over a range of operating conditions as stresses or perturbation factors are applied. The Functionality-Performance Tradeoff Curve captures the relationship between functional richness and system performance. In terms of abstraction are two types of systems to consider when developing a FPT curve: performance-oriented systems and functionally-oriented systems.

Performance-oriented systems, as the name implies, place a premium on operational performance characteristics such as end-to-end delay, transaction processing time, data throughput, or update speeds.

Functionality-oriented systems operate at the other end of the spectrum; for these systems timing is less important than maintaining a full spectrum of system functionality. The FPT curve for functionality-oriented systems will trade performance for richness of functionality when the system experiences stresses or perturbations.

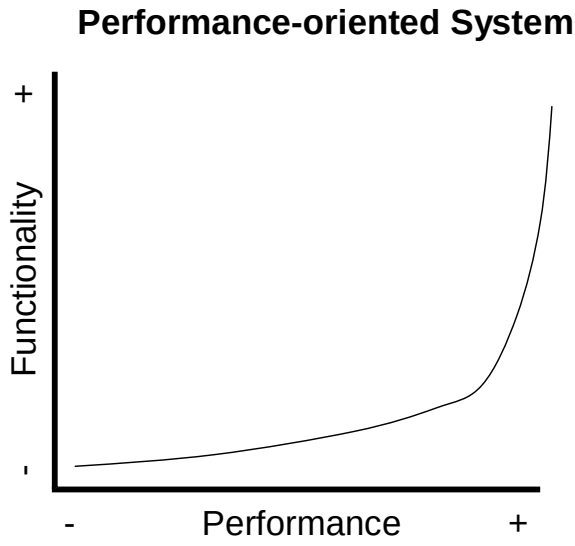

Fig. 2. Performance-oriented System

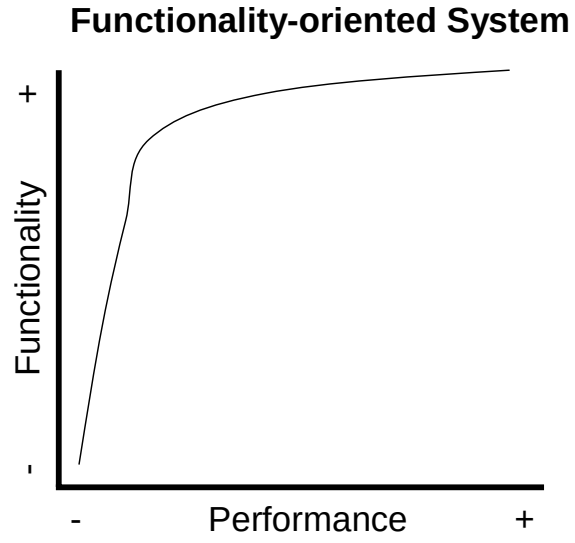

Fig. 3. Functionality-oriented System

Service Guarantees

Graceful degradation of services in resilient systems borrows from Service Level Agreements common in the application service provider domain. Formally defining how services can degrade to prevent an overall system failure is critical to maintaining system performance, functionality, or perhaps a blend of the two. Preventative actions such as shutting down certain services or

throttling the performance of other services when systems experience stress can often make a significant difference in system stability.

Another, perhaps complementary, way to look at service guarantees is assured *minimum service performance.* Borrowing from quality of serive (QoS) concepts in data networks, the idea of assured minimum service performance involves issues of admission control (who can get into the system and when) scheduling (what actions to take when) and differential treatment of services (what requests are most important).

Prioritized Recovery

In an ideal world—which we know does not exist—systems would never encounter a situation where recovery is needed. In our formulation for systems resilience we need to provide for orderly and predictable recovery of services as systems attempt to return to nominal operating conditions following a significant perturbation.

Recalling the notion of, prioritized revovery of services can be thought of as the micro-level recovery behavior of individual system services, servers, or applications which enmass produce the desired overall or macro system recovery behavior characterized using FPT Curves.

Resilience at Multiple Scales

To design resilience into our systems we must look at the scale of such systems from two perspectives: top-down (macro-to-micro) and bottom-up (micro-to-macro). As such we can view any complex system as a system of systems where each system (in the micro sense) is composed of several interdependent components. Some research challenges in the area of scale involve constructing resilient systems from essential unreliable individual systems which are in turn comprised of unreliable components. Adding emergent and unpredicable behaviors of large complex distributed systems including the human factor make this a rich area for research.

Summary

This paper introduced the concept of resilience as the primary characteristic we must architect into our systems at all levels due to the risks and challenges posed by three convergent forces—connectivity, complexity, and interdependence. Taking a widely multidisciplinary approach to

characterize what resilience can and should be, we have borrowed from network science, physical science, systems engineering, data networking, and social science.

Ongoing research is needed in resilience, specifically how complex systems are created and how they behave under stress. Much can be learned and leveraged from the growing body of knowledge in network science. Our future research will seek to advance the notion of resilient systems using network science as a basis for these efforts.