2 Demand, Capacity, Fragility, and the Emergence of Networks

2.1 INTRODUCTION

This chapter is more theoretical than most in that it identifies an emerging trend in thinking rather than describing some of the changes in the strategic infrastructure that have taken place since the first edition. At the time of the first edition, much of the focus on critical infrastructure protection (CIP) efforts was directed at the very local level—how to protect key facilities, etc.

This has evolved over time. Recent infrastructure issues have highlighted the fact that this infrastructure is subject to impacts that can flow along interdependencies and also disruptions within its networked environment.

2.2 WHAT ARE WE TRYING TO PROTECT? THE CONCEPT OF CAPACITY

If critical infrastructure is really about the infrastructure necessary to preserve the safety, security, and economic well-being of citizens, then shouldn’t the focus necessarily be on protecting infrastructure or assuring that a given service continues to be delivered as required? Although the former is certainly important, the latter aligns much more closely with the stated goals of CIP.

The fact is that a given infrastructure at the local level is there to provide some level of contribution into the system. The sum of these contributions, the ability to coordinate how those services are delivered, and the means of delivering them to their intended recipients may be best described as the capacity of the system.

These three elements (safety, security, and economic well-being) are important because they operate similarly to the fire triad (heat, oxygen, and chemical reaction). If the infrastructure can generate a significant amount of the service but cannot identify where it is useful or deliver it to those points, then the system has essentially failed. At the same time, a well-coordinated and well-maintained grid that does not have anything sent through it is still failing to meet the final goal. The ability of the system to produce, distribute, and deliver can be described as the system’s capacity.
2.3 DEMAND: THE REASON FOR CAPACITY

Demand and capacity exist in a constant balancing act. This is not to say that they are always in equilibrium—they rarely are. It simply means that where there is a demand, capacity will attempt to fill that demand. Where there is surplus capacity, there is likely going to be a demand attempting to exploit that capacity. Those with a background in a supply-and-demand economics will find this concept very familiar.

2.3.1 THE CONCEPT OF PERFORMANCE

The concept of performance basically describes whether the system works with sufficient capacity to meet its demand. For example, if there is a demand for 500 units of something, then the system would be considered in balance when it delivers those 500 units, and otherwise out of balance.

Because of the nature of critical infrastructure, it can be reasonably argued that three imbalances have to be considered. The most serious of these involves a situation where the capacity does not meet the demand. This may be represented by a situation in which some portion of the population does not receive an expected level of the critical service—such as occurs during a power failure. The second most serious condition occurs when the capacity exceeds the demand, but leads to a response where the capacity is reduced, leaving the system vulnerable to a spike in demand. This might be exhibited in situations where the private sector is primarily involved in the delivery of the service, but due to a surplus of supply, businesses leave the market because they become intolerably unprofitable. The final imbalance is a sustained surplus of capacity.

2.3.2 LOCAL IMPACT AND THE INFLUENCE ON CAPACITY

When infrastructure is disrupted at the local level, that disruption loses its ability to provide the expected level of capacity into the overall system. At the local level, the clearest understanding regarding the loss of capacity will flow from activities less associated with physical security, but rather with business continuity planning (BCP). Within BCP, thresholds are communicated that are used to determine the severity of impacts or losses of key resources, etc. Although BCP generally ends at the edge of the organization’s responsibility or mandate, the concept of CIP urges this approach to be carried on throughout the organization and into progressively larger systems.

Some care has to be taken here to ensure that the quality of service is maintained at a manageable level. What if the final product (e.g., a fuel) fails to reach that level of quality for it to be usable in the system? This aspect of integrity is somewhat different than the traditional “nothing added, nothing deleted, and only authorized changes made through well-formed or defined processes” and more closely in line with the traditional views of quality assurance and quality management.
2.3.3 Results of a Local Impact in the Immediate Sense

When something is disrupted, we return to the concept that the availability of the critical service has been reduced. This leads to three important events that are worthy of study. The first event involves what the loss or reduction of that service means to the overall system. This revolves around the concept of what consequences arise should the organization fail to meet its goals—again, a power failure, loss of transportation, etc. The second event involves what the loss or reduction of that service means to the internal use or management of inputs that would normally be used to maintain that level of service delivery. How do the unused inputs survive the impact? Are they perishable or must they be used within a certain time frame before they are no longer of value? Are they persistent in that they can be stored nearly indefinitely without a loss of value? These factors should generally be included in the basic impact analysis—often in consultation with operations or material management personnel. The third event involves how the organization manages the fact that it is no longer consuming those inputs at the same rate. Does this mean that it will stop purchases of future inputs or that it will simply delay the delivery of some? These upstream impacts are also important factors to be considered both in the local impact analysis and later in the understanding of the impact on the overall system. For those seeking parallels, concepts defined in supply chain management and logistics provide some input.

Generally, at the local level, four classes of impacts are observed. The first are delaying impacts that essentially slow the inward flow of something into the system. This concept is seen when warehouses are filled—at some point, the warehouse is full but we still need to store the material. The second involves the concept of lag. This category of disruption describes the condition where something else is slowed down because the necessary amount of inputs is not being received. Finally, at the other end of the spectrum, the system will attempt to balance itself through either the third class, push (seeking to find new demand), or the fourth class, pull (seeking to find surplus capacity that can be aligned against unmet demand).

2.3.4 Relevance to CIP

The concepts of push, pull, lag, and delay are becoming increasingly understood at the local level. This was initially established through bodies of knowledge associated with supply chains and logistics; it then moved into the realm of BCP, and has now become more understood in the realm of CIP. Where the divide currently resides is between the local and regional (small system) levels when you look at the CIP services that have stemmed from such concepts as force protection, infrastructure protection, etc. For the reader, having an awareness that this bridge is likely to be built in the near future has a significant impact on how organizations can prepare and integrate their corporate security, information security (including information technology (IT) security and network security), BCP, incident response (in both the physical and informatics senses), business resumption planning, and disaster recovery planning programs.
2.3.5 Push, Pull, Lag, and Delay in the Network Environment

The concept of push, pull, lag, and delay is not unique to the transportation system—but is rather characteristic of any system that involves exchanges. The concept is familiar within the energy sector in pipelines and across transmission lines. These topics are also familiar in water systems and networks, and although they illustrate applications in physical networks, the concept is similar to various other concepts, such as bandwidth, throughput, and buffering within the logical realm.

For the CIP professional, an understanding of how these four elements operate is of vital importance. Fortunately, these sectors have already carried out significant research with respect to each of these elements as they work on understanding and refining their understanding of their own risks.

2.4 At the Regional (Small System) Level

Returning to the first principle, we find that the core values associated with critical infrastructure services can be prioritized in order of availability, integrity, and confidentiality, as we migrate to a smaller system, usually at a regional level. Understanding what the concept of push, pull, lag, and delay means within that small system plays a vital role in the ability to assure delivery of critical services—now being thought of in some limited circles as critical infrastructure assurance.

2.4.1 Influence at the Small System Level

When attempting to assure these services, it is most important to understand how these concepts operate at the small system (regional) level. Consider that each node (or intersection) and each channel (or conduit) can only handle a certain capacity. If there is no release to the surplus demand (e.g., through a release of pressure), then the system simply operates as it is best able to. Beyond that, however, the system begins to clog as the surplus demand that cannot be handled attempts to find other options and, if this is not possible, remains in place.

This concept can be seen in most metropolitan automotive traffic congestion conditions quite clearly. A route can handle a certain number of cars in a certain amount of time. When that level is exceeded, the route begins to fill. When the space between intersections is full, cars cannot pass through the intersection or will block the intersection, thus compounding the problem, and the system begins to fill.

What becomes important at this point is the ability to identify that a disruption has taken place, find alternatives that can release the pressure, and then route or reroute the demand onto those alternatives. The release of pressure, if balanced correctly, allows the system to break the cycle of cascading and expanding failure, and regain that delicate operating balance between capacity and demand.

There are two factors that have a serious influence on this. First, what if there is no surplus capacity available in the system? In this instance, the system fills. It is also important to note that where the system is full, it, too, denies further movement through the system. The second factor may be whether surplus capacity within the system can actually be reached from the disruption. The routes between nodes fill
as a result of the surplus demand; here again we have a situation where the impact is cascading and expanding.

2.4.2 Current Efforts and Research

As the reader will soon see, the legislation, regulation, and other forms of oversight regarding the local layer associated with critical infrastructure have evolved somewhat since the first edition of this book.

The first significant line of research has focused on the concept of interdependencies. Interdependency is where the level of one system's product is reduced, and this reduction causes an impact in another system. For example, a loss of fuel production impacts the transportation sector or a loss of electrical power affects telecommunications. For those involved in BCP, the concept of interdependency may appear to be complicated from an operational viewpoint, but is relatively simple to accept at a theoretical level.

The challenge here is that the concept of interdependencies is approaching a situation much like cancer research. Most of us understand that the term cancer actually represents a significant number of different diseases. As a result, one might fund "cancer research" (and we would certainly not discourage you from doing so) but not have a clear sense as to what form of cancer is being researched. The same might be argued for interdependencies (see the example reports within the footnotes, especially the one from Idaho National Laboratories).

The second challenge is associated with the concept of network fragmentation and dissolution. Since the first edition, significant work has been carried out in the transportation and energy sectors to try to understand how the disruption in one part of the system impacts the rest of the system. For some, it is simply akin to the butterfly effect—an assumption that may hold true for nearly inconsequential parts of the system, such as a terminal or isolated node. On the other hand, disruptions at major infrastructure points may be apparent rather quickly, as the impacts flow throughout the various connections and begin to influence the capacity at other locations. Documentation has been published through the U.S. Department of Homeland Security's Transportation Security Administration fairly early on, and recently, entire works have been dedicated to the concept of using technology such as intelligent transportation systems as a safeguard against this type of issue.

2.4.3 The Interdependency Hydra

We have alluded to the concept that the term interdependency is becoming used to describe a number of states within and between networks and their interrelation between other systems. A number of projects have recently attempted to quantify and qualify the relationship between sectors. In Canada, these studies have resulted in Defence Research and Development Canada putting forward a call for proposals and papers as late as November 2008. Similar projects have existed within the United States for a number of years.

Several centers have been involved in this type of research over the past few years. In 2006, the Idaho National Laboratory conducted a survey to identify the
major works associated with "critical infrastructure and interdependency modeling." Private researchers have also published since the publication of the first work (including the authors), often through Web portals and university publication portals.

When considering interdependencies, one might argue that there needs to be a basic understanding of how the impact flows between or across sectors. These might include, as a basic system of categorization, the following:

Interdependencies flowing out of one system (host) and impacting an independent system in that the impact does not cycle back onto itself (the host)—henceforth, this would be more of a dependency rather than an interdependency.
Interdependencies flowing out of one system (host) and impacting a system that provides a direct good or service back into the host system, leading to an elevated rate of deterioration attributable to the initial disruption.
Interdependencies flowing out of one system (host) and impacting a system that then provides a service to another sector that then has an influence on the host.

There is still a significant amount of work to be done with respect to the proper categorization and definition of these types of events. The questions that persist across a number of blogs and discussions where researchers tend to communicate continue to center on a general acceptance that some of the underpinning principles appear to be common, but are still difficult to quantify.

2.4.4 Network Fragmentation and Dissolution

The concept of interdependencies and cascading impacts has also worked in parallel and even contributed to a growing amount of research into the fragmentation and dissolution of networks. Much of this still focuses on the concept of mathematical models and translations of informatics systems into physical infrastructure. Again, there is merit to this research: Where such concepts translate gracefully into the physical domain, they are worth keeping. Where a concept is discounted, the results of the research still have value in that they can narrow the focus of other research based not on pet theories or whimsical intuition, but rather on sound scientific bases.

For those entering the arena of critical infrastructure assurance, the concept of network fragmentation and dissolution is relatively simple to explain—if one does not get bogged down in the complexity. Consider capacity and demand. Where there is a surplus of demand, it will seek out spare capacity (or where there is an ability to meet the demand, surplus capacity will be sought). This goes back to what was discussed in terms of how impacts affect the small system level.

What has become increasingly important to researchers is the ability to predict how that system will collapse and break apart. This is important for two reasons. First, a predictive model enables effective preventative measures (focusing on the robustness of the system) and also pre-position mitigation and response strategies (focusing on the recovery aspects) to be established. To return to the traffic congestion example, this is somewhat akin to being able to identify where the traffic jam is most likely to appear next.
It is perhaps fortuitous that this research has coincided with difficult economic times. This is because both U.S. and Canadian administrations appear to have committed to working on significant infrastructure upgrades as part of their economic recovery packages. Prudent planning would involve a forward-looking approach that identifies what capacity will be needed, rather than simply restoring overburdened infrastructure to its original design.

2.5 CYBERTERRORISM

In addition to the physical and operational safeguards, the concept of cyberterrorism has approached the forefront of many critical infrastructure issues. Outside of Hollywood’s extrapolation of potential events, the world has seen clear examples of the results of coordinated cyberattacks in Estonia and Georgia as part of political and military campaigns. We have also seen the specter of groups of cybercriminals operating out of Asia with potential ties to state actors. These issues are of significant concern.

The Gilmore Commission noted that the “cyberattacks incident” to conflicts in the Middle East “emphasized the potentially disastrous effects that such concentrated attacks can have on information and other critical government and private sector electronic systems.” The commission concluded that although not “mass destructive,” attacks on critical infrastructure would certainly be “mass disruptive.” It also concluded that likely perpetrators of cyberattacks on critical infrastructures are terrorists and criminal groups rather than nation-states. As a result, the commission predicted that detection of these attacks would fall primarily to the private sector and to local law enforcement authorities.

This statement, however, has taken on additional importance. There is now increased recognition within industry and government that if key resources (this term is chosen specifically to align with BCP approaches) are connected through Internet-enabled technology, cyberthreats to those key resources need to be recognized and addressed.

2.5.1 THE PENDULUM OF CONVERGENCE

Convergence, simply put, is the gradual integration of physical and logical infrastructure. For those without degrees in architecture (logical or physical), it may be described as the gradual march onto the network-enabled system.

Convergence is really being driven by two interrelated variables. The first variable involves the need for increased efficiency and situational awareness. This is a direct result from the need to be increasingly competitive on a global stage. Where North American markets used to be serviced by North American companies, one might argue that the past 25 years have essentially destroyed that concept, particularly when considering issues associated with supply chains and offshore production. As a result, there is an increasing intolerance for isolated or stand-alone systems that cannot be expanded as operationally required or as per the will of management. The second variable involves changes in technology. The current generation of engineers and developers is not particularly enamored of the concept of working with single-purpose, analog systems. The generation that did have to work with that
technology is gradually moving on into its retirement years—something that some consider a crisis and others a blessing. The end result, however, is the deployment of key resources using a type of technology that may, if not treated carefully, be subject to the same types of attacks that were present within the context of cyberterrorism. These two factors, the increasing pressure toward network-enabled systems and the decreasing supply of those able to work in past logical environments, will likely change the face of physical security and enterprise security.

The concept of convergence does not simply mean a change in the application of technology; it also requires a change in organizational culture and personal approaches to the issue of security. Some of the basic concepts will, of course, be consistent. As we look at how issues are identified, problems and issues are scoped, challenges are met, and solutions are applied, however, the traditionally diverging IT and physical and personnel security communities will be forced back to the same tables. Although the security industry is in for some interesting years as the various elements in these communities go through the normal processes associated with storming and finally norming, the end result for industry may well be worth the effort if both sides remember the primacy of operations.

2.5.2 **Convergence and the Understanding of Threat**

Convergence is also going to impact on how threats are considered within an organization. The all-hazards approach has been front and center in the past—but its application has largely looked at the surface layers of threat. For example, keeping a cybercriminal at bay was a matter of installing a firewall, whereas keeping a prowler out was a matter of locking the door. Today’s criminal, however, has access to both tools, and with the change of technology, it may be that the prowler is unknowingly working for the cybercriminal and has access to complex tools specifically conceptualized, designed, and used for defeating security infrastructure through logical means.

Perhaps this example can be described by providing three divergent threat scenarios—each of which intersects with the critical infrastructure and key resources domain. First, consider the Federal Bureau of Investigation (FBI) report on December 3, 2008, that identified another threat to U.S. infrastructure—the theft of copper that is being fed by an increasing demand for the metal, including in overseas markets. Although it may be argued that all network infrastructure runs on fiber optics (it does not), it should be noted that this is not the type of traditional threat that might appear on a technical vulnerability analysis. The second threat involves personnel. Again, the FBI published (December 16, 2008) a report describing how certain elements of organized crime were able to infiltrate, through financial means, seaports along the East Coast. In this case, the threat vector was not asset based, but rather personnel based as the gradual trapping of individuals who had been given access. Finally, in a similar report in *CSO* magazine, a network administrator was able to establish himself as the key source of control over much of the city’s network infrastructure. These three events show how a potential adversary or attacker could gain access over key resources using indirect methods.
Of significant concern today is the concept of the hybrid attack. As noted earlier, changes in technology and the availability of more and more processing power (an ongoing challenge) are leading to situations where adversaries have a wider array of tools at their disposal. Thus organizations that tend to focus their security activities in such a way that any one or the personnel, information, or physical safeguards are often left more exposed may be at risk of an attacker identifying, examining, and finally exploiting that vulnerability.

Consider, for example, a meeting room in a public area. On one hand, the fact that it is intended for public access and resides outside of the more sensitive work areas is good; on the other hand, one has to examine whether the IT infrastructure installed in that boardroom is sufficiently hardened. This should be done so that an attacker does not simply bypass the physical security infrastructure by using the network connectivity to pass through the barrier in a way similar to crawling over a dropped ceiling or defeating a weak lock. Although hard connections are reasonably simple to address, the propensity of several organizations to work toward wireless access points or capabilities means that the physical security expert will have to look not only at the physical design, but also at how to establish the necessary levels of shielding and standoff—particularly if the adversary can simply sit in public areas where it is difficult to control his or her activities.

2.5.3 Setting the Stage for Fragility

Given a basic understanding of some of the pressures within the system and some of the upcoming challenges associated with understanding how the infrastructure is protected, we can begin to look at a concept called fragility. Fragility, although reasonably new in this context, is not mystical—it has been inferred in such fields as reliability engineering for some time.

Fragility can be described in terms of the propensity of something to fail. At the local level, this aligns reasonably closely with the concept of the risk of loss associated with availability and, as a secondary factor, integrity. In reality, this can be divided into three major categories.

The first of these categories refers to the design of objects. When an engineer designs something, he or she indicates some level of assurance with respect to the design actually performing as intended. This is largely tied to the amount of effort spent in design, implementation, and other aspects of quality assurance. Aircraft manufacturers and other entities are subject to strict safety regimes; for example, they may have remarkably low tolerances for failure. Other industries, where the impacts are not so grave, may have considerably lower thresholds. Thus, given that an engineer may ensure that an aircraft design will work 99.5% of the time, whereas another engineer may only have to ensure that his product will work 75% of the time, we have our first significant difference in fragility.

This fragility, however, is often based on averages, norms, or set ranges of conditions. These norms or averages are used to provide that final calculated value that gives us that assurance regarding the design. We know, however, that as conditions change, they may have an impact. Personnel may be less able to perform tasks in extreme heat or cold, assets may be susceptible to certain conditions (e.g., low
humidity leading to static electricity near computers), facilities may require that certain environments be maintained. Information may require certain systems for handling in order for it to be considered trusted, etc. As we look at the item being examined (similar to a target), we may find that certain inputs into the system do not perform as well under certain conditions—for example, workers in high-heat areas may not be able to exert themselves the same way. This leads to the second type of fragility, natural fragility, so named because it is based on how the target would perform within the immediate environment.

As has been noted by scientists and poets alike, change is a constant within our environment. Cyclical fragility brings together the major elements discussed earlier—the concept of systems being sustained by the efforts of various inputs (persons, objects, facilities, information, and activities), the concept of capacity and demand attempting to maintain a level of equilibrium, and finally the fragility that is intrinsic at each point of time within the system.

For personnel, there are a number of cycles that one must remain aware of. In the longer-term view, we have the current issues associated with an aging population and the impact this will have on corporate knowledge. At the same time, there is the time involved in developing bodies of knowledge and communicating it to people, such as what we are seeing with the convergence issue. Within the medium term, the life cycle associated with business and with labor contracts provides another example. At the very short and immediate end, one might even argue that the various cycles are associated with fatigue and attention spans. As one will quickly realize, many of these do not impact the security realm—but have a significant bearing on the concept of critical infrastructure assurance when looking at potential sources of disruption.

Assets face similar challenges. When engineers design things, they generally include a life cycle based on adherence to a specific maintenance cycle and without certain constraints. We see this with our cars. They are anticipated to last a certain period—but only when you do not abuse them and keep the necessary maintenance up to date. Perhaps the most advanced bodies of knowledge in this regard involve life cycle management and safety programs. These programs track the use, maintenance, and age of assets as part of a means of reducing the risk of failures that can lead to either loss or accidents. Again, however, this type of approach has a significant bearing on critical infrastructure.

Facilities provide a nexus between two types of cycles. The first cycle, the age and usefulness of the facility, can be linked back to the same issues associated with assets. Materials deteriorate and require replacement. Structures become outdated in terms of the infrastructure they can provide. Another variable, weather, plays a significant role. Again, in the longer term, seasonal changes can affect the ability of persons, assets, or activities to perform as intended. Although some of these may be reasonably innocuous (e.g., a slight change in temperature), others may be profound, such as periodic flooding or dry spells. In the short term, the simple change between day and night may lead to different levels of risk.

Information and data, however, are somewhat different. In this context, the cycles are not attached so much to natural conditions, but rather to operations and what the information describes. Consider, for example, a table associated with the movement
of a container—the value associated with the movement of the container shifts as one
moves across the planning stage, through coordination and monitoring, and finally
into audit and review. The cycles associated with information, one might argue, are
inextricably linked to the timeliness and relevance of the data and what they represent.

Finally, consider the concept of activities. Cycles play a factor here. Looking at
activities, we cannot ignore the effects of fatigue associated with persons and their
assets. Thus we cannot ignore that various activities are more relevant at some points
than others. These are generally associated with operations and coordination—
points that permeate various systems and processes.

So how do these factors impact critical infrastructure? The answer lies in the need
to ensure that the critical services are, in fact, available on demand and can be relied
on from a quality assurance perspective. Although this approach argues that these
five categories (persons, objects, facilities, information, and activities) cover signif-
ICant aspects of a process, it is still incumbent on those conducting assessments to
examine each process thoroughly. The challenges associated with convergence and
new ways of thinking play a significant role given that those deficiencies in current
and forecasted ways of doing things lead to gaps in understanding and knowledge
with respect to the risk to critical infrastructure.

2.5.4 Fragility and Destabilization of Systems

At the regional or small system level, all the factors cited above have an influ-
ence on the capacity of the system. These influences can skew the balance between
demand and capacity, shifting it in ways that lead to situations involving push, pull,
lag, and delay. Where these influences stem from single infrastructure points, the
effect manifests itself first in that local area. Depending on the nature of the effect,
it will then influence those areas around it until the system is able to naturally
restore balance.

The immediately impacted area will often depend on the level of capacity deliv-
ered by the infrastructure into the overall system. A relatively inconsequential or
insignificant piece of infrastructure may cause some destabilization within an area
that can be corrected reasonably quickly. On the other hand, where a key piece of
infrastructure is disrupted, the immediately impacted area may be much larger and
the system more destabilized; for example, the removal of a central hub within a
transportation system or a key power production facility. How these disruptions ca-
cade through the system will again depend on the system resiliency and redundancy.

The second consideration is where the full small system is impacted. The fac-
tors involved in this case are those that span the full system—in any of the catego-
ries of personnel, assets, facilities, information, or activities. We see this illustrated
in terms of weather, labor disruptions, certain types of cyberattacks, and similar
types of threat vectors. Where these are involved, the capacity in the overall system
becomes diminished, again leading to disruption of the equilibrium between demand
and capacity.
2.5.5 **Fragmentation and Dissolution of Networks**

Fragmentation and, catastrophically, dissolution of systems occurs when elements of the system are no longer able to communicate and coordinate their activities, essentially becoming individual entities encapsulated within the system. Although the local level looks at this in terms of a disruption, the same can be said at the small system or regional level.

Following a disruption at an infrastructure point (e.g., a facility), the next phase involves fragmentation. Although the concepts of push, pull, lag, and delay provide a mechanical description of the system-level impact, one can also divide the impacts into two broad categories.

The first category concerns disruptions involving the loss of infrastructure. This category involves situations where the infrastructure essentially fails, resulting in the contribution of capacity being lost to the system. We have seen these types of events in a range of bridge collapses, failures in the surface system, manufacturing sectors, and so on. When that infrastructure is lost, the capacity is lost, and the loss of capacity skews the equilibrium between demand and capacity in such a way that the system suffers disruption until it can rearrange itself by determining new options on how to meet the demands placed on it.

Fragmentation occurs under two conditions. First, the disruption may be at a key point that severs two parts of the system, essentially creating a number of smaller systems until it can be reestablished. In this context, the concept of fragmentation comes from the loss of what could be considered a key resource. It may be characterized by a single event, and because of its localized nature, there may be a significant focus on building up the robustness in terms of its ability to withstand impacts.

The second category involves conditions where the system essentially fills due to a surplus of demand combined with a reduction in capacity. If the capacity of the system involves a rate, then that rate becomes a very important factor, particularly where the system is working at near capacity. A reduction in the rate of being able to handle demand is essentially a reduction in capacity—and when demand is approaching the limit of capacity, then the system will begin to fill. Once the system is filled, and if the demand continues to try to exploit the capacity, it will not be able to do so. As a result, the system will gradually slow down and, if the reduction in capacity is serious enough, the operations will abruptly halt. This essentially fragments the affected area from the system, although recovery can be attained once the rate at which demand is handled allows for the system to clear itself.

Dissolution of the network involves fragmentation at a catastrophic level. In this case, the impacts are adequately severe so that the various components within the system can no longer communicate with each other. The end result is that the network becomes a community of individualities, unable to coordinate its activities.

Dissolution of the network becomes a real risk where the network relies on a single service, a single type of service, or is subject to a sector-wide vulnerability. Again, we have to return to the concept of persons, objects, facilities, information, and activities at this point. As with BCP, the concept of having single points of failure at the network’s strategic level should be anathema.
2.6 DISSOLUTION AND CONVERGENCE: AN EMERGING RISK

Where all aspects of an infrastructure share common characteristics or where systems rely on a common point of service, the overall infrastructure or system becomes vulnerable. Given the trend toward convergence and network-enabled infrastructure, it would be worthwhile to include some brief and introductory comments regarding the risk of dissolution and the concept of convergence.

2.6.1 CONVERGENCE, NETWORK EXPANSION, OPEN ARCHITECTURE, AND COMMON CRITERIA

The increasing reliance on networks illustrates one potential avenue that could become a sector- or system-wide vulnerability if not handled diligently. First, convergence describes a condition by which the physical and asset protection infrastructure is becoming increasingly network enabled. Similarly, operational networks are expanding with new partnerships being formed to streamline delivery of services.

The concepts of open architecture and common criteria pose a significant challenge at this point. By publishing the open architecture and common criteria, a determined attack planner can identify certain characteristics that are common across the system and can attempt to reverse-engineer those characteristics or criteria in an attempt to determine a weakness. Given that both of these concepts communicate on a global scale, one can only assume that this potential threat vector is likely to occur.

However, the same concepts, in their development, also allow for a wide community challenge to the architecture and criteria, thereby reducing a number of the vulnerabilities in the system. The key here is in allowing for that broad and diligent consultative period before moving the overall structures into the public domain.

This concept has been a challenge for cryptographers and cryptoanalysts for quite some time. The result is reasonably simple. The encryption process uses a similar process in the development of new methods of protecting data against unauthorized disclosure or modification. For the critical infrastructure sectors, the principle may be similar. On one hand, the common criteria and open architecture may be globally available, but the specific and detailed information necessary to exploit something at the local level remains hidden from view. Again, however, the concept hinges on an approach that was rigorously challenged at the front and then monitored in terms of its effectiveness.

2.7 MARKING THE JOURNEY

Up to this point, we have looked at a theoretical situation that describes critical infrastructure assurance as an overarching, system-based, mission-focused view of how critical infrastructure sectors function. One might ask why this appears at the relative start of the work. The answer lies in three important steps.

First, we understand the nature of the changes that occurred as a result of 9/11. Many of us do not realize that the changes at that time were a fusion of the changes that followed the Y2K concerns of 1999.
Second, by providing two benchmarks based on the first and second editions of this work, we can define some of the changes to date. The first edition marks a point where one might argue that Western society was in a significant mitigation phase with respect to the 9/11 attacks—the term mitigation being used in its emergency preparedness context of those immediate steps used to contain damage. The second edition came at what might arguably be at the end of that phase.

Third, we understand that the business of CIP, homeland security, and similar programs is likely to undergo a significant change. We have looked briefly at convergence and the challenges of an evolving hybrid (physical and cyber) threat. There is also a new pressure to address economic and social issues—it may be argued that the old way of doing business (massive contributions and spending) is about to dry up as we look at renewed fiscal responsibility and restraint. Finally, the people involved are changing—from the retirement of many of the Cold War era security specialists through the process and systems engineers who defined what we now call critical infrastructure.

Thus this final step is about looking back into the past, and then looking at what is in place today and understanding and questioning the changes that have taken place. Finally, it is about extrapolating that information into the future to best estimate what is likely to be.

2.7.1 Overview

The period following September 2001 was a period of treaties, legislation, and regulation—much of it done quickly to respond to perceived needs. The marine industry is the clearest example of this, where entirely new regimes were constructed from an amendment to a safety code (Safety of Life at Sea [SOLAS]) in response to the events.


As we look at the legislation that has been passed since the 107th Congress (2001–2002) and up to the 110th Congress (2007–2008), we see what may be described as the response to one block of activity. In the 107th Congress, one might argue that the focus was on immediate responses to events—particularly within the transportation sector (particularly aviation, seaports, and pipelines), border security, dams, and the beginning of the push into cybersecurity. One must also note the legislation that gave rise to the Department of Homeland Security.

2.7.3 Legislation: 108th Congress to 109th Congress

We see, from the 107th Congress through to the 109th Congress ( inclusively), a shift from the very general legislation to issues that are more granular. For example, in the transportation sector, we see the initial marine security bills being general in nature, whereas in 2006, marine security was beginning to legislate the approach of layered defenses (HR4954) and increasing granularity up to 2008.

We also need to understand that legislation, where none exists, can be challenging in democracies. As a result, we also see, particularly in Canada and the United
States, a gradual shift from legislation to regulation through rule and measure making. There are challenges with this approach—the oversight mechanisms for each of these are fundamentally different. Although not fully examined, one might wonder if there is a research opportunity for someone to examine the nature of this shift with respect to the oversight mechanisms used to create it.

2.7.4 The State Today: A Recap

Today’s situation puts the critical infrastructure protection and critical infrastructure assurance domains at a crossroads.

First, there are significant resource issues. The collapses within the financial and automobile sectors put a significant strain on the economy. Concurrently, the costs associated with foreign wars and activities have put strains onto existing economies in terms of debt. Finally, there is the challenge of the reducing tax base caused by an aging population that packs two punches—reduced income through management of fixed incomes (pensions, etc.) and lost earnings (in terms of income tax), as well as the costs to social security and social support programs. In essence, one might argue that although the cupboard is not bare, we have certainly seen the back of the cupboard and realized that things cannot carry on as they have.

Second, there are significant knowledge base issues. The aging population noted earlier also carries within it a significant portion of the corporate knowledge associated with the infrastructure involved. We are also finally beginning to make the shift from an asset protection mode (some would theorize this being a mitigation response to 9/11) back to an infrastructure assurance mode where a new level of understanding has to be overlaid on top of traditional security and protection approaches. The issues of convergence mean that traditionally separate communities will now be forced, through necessity (even if to survive in the economic sense), to interact and cross-pollinate, and as we address these issues within our own community, our adversaries and competitors are working on harnessing opportunities provided in the new structure and our readiness to accept it.

Finally, there is the concept of public-private partnerships. With significant portions of the infrastructure operating outside of federal, or even government, ownership, governments have had to adjust their thinking from a span of control approach (e.g., through the dictating of plans) to a span of influence approach (indicated by the shift toward frameworks).

2.7.5 Research and Understanding

For those involved in CIP, homeland security, and emergency preparedness, these are also exciting times in the research community. The concepts described here are those that are openly researched in the community—concepts that are driving new technologies and approaches to infrastructure management. These concepts also cause friction between communities, particularly those that are firmly attached to rigid doctrines and dogmas and are unwilling to expand the breadth and depth of those approaches into new areas.
2.8 AUTHORS' NOTES

Flexibility in thinking for the professional can be challenging—particularly when the profession demands that the professional remain compliant with a certain approach or in line with a certain body of knowledge. Today, the infrastructure and asset protection communities have a plethora of certifications that can be applied to certain aspects or sectors.

Within organizations, there is a certain inertia that resists change. Pride in the organization's culture, heritage, or traditions can be either an anchor in the storm or an anchor in the race to reach new destinations. The key is to see the anchor for what it is—a tool that needs to be applied correctly. Will organizational issues be a help or a hindrance—and can the conflicts and frictions that inevitably come with changes of culture be managed effectively, or will they pose greater vulnerability to the infrastructure in question?

Finally, where does the leadership in the critical infrastructure assurance question lie? Does it lie with society and its duly elected representatives, or does it lie in the private sector? If you have a specific answer to the question, what does this mean to the role of government and oversight? What if times change and new approaches are needed? Is our own organizational structure resilient enough to meet those challenges? These are some of the questions within this field—making it awesome in its scope and more than a little exciting in terms of its application.

NOTES


2. The call for proposals can be found at www.css.drdc-rddc.gc.ca/program/index-eng.asp. Although the proposals technically closed in November 2008, this kind of research is often ongoing and will require significant studies in the future (alt URL: http://cipbook.infracritical.com/book5/chapter2/ch2ref11.pdf).


4. The Gilmore Commission is an advisory panel that assesses the capabilities for responding to terrorist incidents in the United States involving weapons of mass destruction. Response capabilities at the federal, state, and local levels are examined, with a particular emphasis on the latter; the secretary of defense, in consultation with the attorney general, the secretary of energy, the secretary of health and human services, and the director of the Federal Emergency Management Agency, entered into a contract with the RAND National Defense Research Institute, a federally funded research and development center, to establish the advisory panel, which released its fifth and final report in December 2003.


9. www.csoonline.com/article/437873/IT_Admin_Locks_up_San_Francisco_s_Network.


