Journal of Strategic Security

Volume 2
Number 2 *Volume 2, No. 2: May 2009*

Article 4

Fragility: The Next Wave in Critical Infrastructure Protection

Allan McDougall
Transportation Security Community & Strategic Leadership in Federal Government Security

pp. 91-98

Recommended Citation
McDougall, Allan. "Fragility: The Next Wave in Critical Infrastructure Protection." *Journal of Strategic Security* 2, no. 2 (2010) : 91-98.
DOI: http://dx.doi.org/10.5038/1944-0472.2.2.4
Available at: https://scholarcommons.usf.edu/jss/vol2/iss2/4

Fragility: The Next Wave in Critical Infrastructure Protection

Allan McDougall

In North America today, we are about to embark on a significant effort to repair, or even upgrade, many aspects of our infrastructure. Many of these efforts are linked to economic recovery packages. Others are based on sheer need. The challenge for decision makers and planners involves ensuring that scarce economic resources are put to their best use. Understanding the concept of fragility plays a pivotal part in reaching that understanding.

Fragility, like many other systems—particularly Information Technology (IT) systems—works on the concept of subjects and objects. *Subjects* are those entities that seek to exploit the services (or capacity) offered by the object. *Objects*, on the other hand, are those entities that deliver some good or service to the overall system. Of course, something may act as the object in one pairing and the subject in another pairing—they are not exclusive in nature. For example, the driver of a car may be considered the subject in the relationship between the driver and the car, while the car is considered the object. The car may become the subject when looking at the relationship between the car and a bridge, insofar as the car (subject) is now exploiting a service that the bridge (object) provides, namely, getting the car from Point A to Point B.

Subjects and objects can be measured using a consistent framework. The subject is measured in terms of the demand that it puts on the overall system, and these measurements are contextual. If the need for more space is the core issue, then the measurement system for the subject will likely seek to quantify how much space is required as its core criterion. If the issue rotates around the number of transactions per unit of time, the subject will likely be measured in terms of how long it takes to process a single transaction. Objects, on the other hand, are measured based upon the capacity that they deliver into the system. Object measurements will generally focus on the performance of the object and how it relates to the demands placed on it by all the subjects that seek to exploit its services.

The chance that the object will fail in terms of its relationship to the subject can be viewed in terms of three perspectives. The first, referred to as *designed fragility* can trace its roots to reliability engineering. The

engineer designs something so that it can be assured to work, given certain stringent conditions, a certain percentage of the time. Naturally, the more grave the impact associated with failure (both in terms of consequence and potential liability), the greater the assurance will need to be—and the lower the level of *designed fragility*. One might even express this in terms of fragility being the difference between all possible outcomes and those that the engineer can assure will be positive outcomes (F = 100 – R).

The reliability of the infrastructure can change as operating conditions change. The engineer has assured the reliability of the infrastructure based on certain operating conditions. Where the infrastructure is operating outside of those conditions, the engineer makes no such promises. Operating conditions may include factors such as temperature, humidity, chemical exposure, age, and so on. For those that have looked at Safety Management Systems, this concept will be reasonably familiar as the gradual operation of equipment outside of acceptable parameters is generally accepted as increasing the risk of failure and, consequently, hazard to the operator and those nearby. This *natural fragility* reflects the conditions present in the real world as opposed to the engineering environment.

The conditions that impact natural fragility can be episodic in nature. Seasons change as do daily conditions. Hence, *cyclical fragility* describes natural fragility and its behavior over periods of time. Not all natural fragility will operate in cycles; sometimes the fragility is more linear in nature. Natural fragility is defined in terms of two elements. First, there is the change in natural fragility that happens along a curve over time. At particular times on the curve, certain conditions may be more prevalent and, as a result, the overall fragility of the system may either suffer or improve. The second element is the wear and tear on the infrastructure as it is subjected to repeated strains. Imagine a cycle of freezing and thawing water. As the water freezes, it expands, putting pressure on things around it, like the sides of a container. As the temperature rises, however, the ice melts, leaving an empty spot that can be filled with a larger amount of water.

The following rules extend from the relationship of subjects and objects. The following rules might be called the *local fragility rules*:

- The *design fragility* of an object is the difference between the total population of outcomes less those that assured through the engineering associated with the system (F = 100 – R).
- The *natural fragility* of an object can be described as either the lowest number of desirable outcomes or where

https://scholarcommons.usf.edu/jss/vol2/iss2/4
DOI: http://dx.doi.org/10.5038/1944-0472.2.2.4

$Fragility_{(design)} \times Factor_{(environmental\ impact)}$. This can also be described in terms of $Fragility_{(natural)} = Fragility_{(design)} \times Factor_{(loss\ of\ effectiveness\ due\ to\ environment)}$.

- The *cyclical fragility* of an object can be described in terms of the curve defined by the maximum natural fragility over a period of time where the conditions associated with the natural fragility repeat themselves.

From a physical infrastructure perspective, these rules have two significant impacts. The first is that it is not enough for the asset protection specialist to calculate impact simply based upon the engineering specifications of the target (the object). In short, a more comprehensive, intelligent assessment of the infrastructure will need to be made to account for fragility. The core challenge here will be identifying the knowledge sets that apply to the infrastructure and then building the capacity of the assessor or, in some cases, the assessment team. The second element is that this requires much more awareness regarding the impacts of decisions and how those decisions will affect the infrastructure. If we change the conditions that surround the object, we have to understand how that will change the object and whether or not this will have an impact on how the subject and the object relate. This will be a core challenge for planners as it would necessitate maintaining running inventories of their infrastructure points and understanding how their decisions would affect those on categorical, if not individual, levels.

Understanding the relationship between these concepts is vital to the understanding of how fragility works at the local level. The local level, sometimes referred to as the tactical level, is the foundation of the strategic infrastructure system at the regional and even national level. This consideration is often neglected when one becomes preoccupied with the protection of the local facility. What needs to be understood and accounted for is how the local facility or infrastructure contributes to the overall performance of the system.

Recall that the concept of *capacity* and *demand* was touched upon earlier in the description of the subject (demanding services) and the object (delivering capacity). In a system operating at full capacity, these two elements exist in a delicate balance that cannot be disrupted without causing some level of disruption ($D = C$). Where the capacity of the system exceeds the demands placed on it and depending upon the configuration of the network, the redundancy of the system allows the system to respond to some kinds of disruption by simply rerouting to new routes and locations where there is surplus capacity available ($D < C$). On the other hand, where there is more demand than capacity ($C > D$), a situation exists where not all subjects' demands can be met.

At that point, the subject must either reroute itself (to find a new source of capacity), remove itself from the system, or become idle within the system. We have all seen this situation arise during traffic jams.

The balance of demand and capacity will determine what state subjects can remain at within the overall system. When the objects are adequately meeting demands, the subjects will continue to carry on through the system or continue to remain active until they have reached or achieved their ultimate goals. A subject can be described as being in an *active state* when the subject is continuing to attempt to exploit capacity. Subjects caught in a situation where there is no capacity to be exploited will find themselves entering a neutral or passive state. In this context, the subject is waiting for some surplus capacity so that it can be exploited. What needs to be clear, however, is that the passive subject is occupying capacity (as opposed to just space) within the system.

Let us apply these principles and concepts to a more concrete example—the surface network for the City of Ottawa, Canada. In this context, let us assume that the surface system in any given area can handle 1000 cars per minute on the highway and 100 cars per minute in the downtown core. Where there is one driver per car, then one might assume that the two aspects of the system can handle 1000 drivers and 100 drivers respectively. When there are 500 cars on the highway, there are 500 cars worth of capacity to be exploited. The system can continue to function and the cars remain active, occasionally changing lanes to exploit areas that appear to have more capacity. When there are 1000 cars operating in the system, the system is operating at capacity. Any new vehicle attempting to enter the system may cause another vehicle already in the system to slow, or even stop, meaning that the system quickly stalls behind the blockage. If, as the result of the introduction of this additional vehicle, an accident occurs, then many more subjects in the system become inactive, essentially entering a passive state. This passive state occupies more and more capacity (as defined in terms of space), until the subjects have taken up all the available capacity and the system begins to stall. Where this becomes even more challenging is when the subjects occupy an object, move into a passive state and then force the object to enter a passive state. This begins to approach the challenges associated with gridlock that occurs when vehicles simply have no alternatives or capacity to exploit, fill up the overall grid, and then become the disruption themselves.

Now consider a system that has the capacity to handle 4000 persons traveling down a particular route. Let's assume a case where carpooling means that four persons occupy one vehicle. This means that the overall

https://scholarcommons.usf.edu/jss/vol2/iss2/4
DOI: http://dx.doi.org/10.5038/1944-0472.2.2.4

ratio of persons to cars has increased to four to one. Mass transit, such as busses, increases this ratio again, but on two fronts. The bus can handle forty persons. On the other hand, a bus only requires the space of approximately four vehicles. First, the ratio of passengers (subjects) to the bus (object) has increased to forty to one. This does not reflect the true value, however, because the bus takes up more than a car's allocated space—it takes up approximately four times that amount. That means that the space allocated would normally hold four cars and, therefore, sixteen persons. The bottom line conclusion is that the mass transit system has improved the system's capacity by approximately six cars (forty minus sixteen split with four persons per vehicle). In short, the bus has shifted the ratio of subjects to objects, creating a condition by which many can exploit a single entity in the overall system.

As a result of this analysis, we can apply certain rules when looking at the interaction between subjects and objects. These are the following:

- *Chaining Sequence Rule*—Given Subject A: Object B and Subject B: Object C, then where Object B and Subject B are the same entity, one can infer Subject A: Object C.
- *Efficiency Rule*—The efficiency of a system can be improved by increasing the Subject: Object.

At the regional level, the relationship between the subject and the object has a logical limit. This limit is based on the maximum efficiency of the subject and nature of the object. Consider the bus example. In this example, we have increased the efficiency of the subject from one to four (carpooling) to ten (the bus—forty in the space of four). We have not changed the nature of the object—it can still only handle a certain number of transactions per unit of time (4000 cars per route). In this case, the subject (bus) has become more efficient, but the ratio between the subjects (cars) to the object has not improved.

This situation leads to a condition where the *capacity* of the system gradually becomes fragile. This is because the system cannot respond effectively to the loss of the efficiency within the system. When looking at our example in Ottawa, we have to consider an aspect of *cyclic fragility* that occurred as part of the labor negotiation cycle. In December 2008, the object failed when the union went on strike, essentially dropping the value of the subject from forty persons/four cars or ten persons per car down to four units in a single car. In essence, approximately twenty-four person-trips worth of demand were suddenly forced back into the system, the equivalent of six cars per bus lost.

The result of this impact depends upon two functions. The first function is how the system can adjust its rate of performance in response to the new demand. In this case, the surface road system is not something that one can add capacity (infrastructure) quickly—it takes time to build roads. With respect to the amount of infrastructure available, the transportation system follows physically fixed routes that are necessary to the operation of the mode of conveyance. The question is not, however, about whether or not new routes can be created, but whether or not they can be adequately controlled. Additional flexibility could also be inputted to the transportation system (to an extent) in the context of aviation and marine industries, thereby altering the nature of the system. In essence, the first function describes the ability of the network to create and add capacity within the system so that the system can rebalance itself, which is a pure resiliency function.

Where the first function cannot be achieved, the second function must come into play. The second function is the attempt to locate and reroute the demand that is not being met into other avenues that offer a surplus of capacity. This may involve alternate routes, the use of side streets, and a host of other means—the important part is that the infrastructure has untapped capacity and can direct disrupted demand onto that capacity. This premise is also not new; it is the foundation of Intelligent Transportation Systems that attempt to route around traffic jams, etc.

Where the system cannot achieve these two goals, however, the next layer of fragility comes into play. This fragility is based upon the fragmentation and potential dissolution of networks. Consider that when a node or a conduit is completely filled (demand meets or exceeds capacity), then it cannot deliver any more service. These are essentially pockets within the system and, depending on what capacity they offer to the system, one will find that the impact begins to cascade upstream (where the system becomes clogged) and downstream (where the expected resources and so on fail to arrive). This is common within the airline industry, particularly during bad weather, and one only has to look at a major hub during that bad weather to see the breadth and depth of the impact.

This impact is again based upon the capacity at the disrupted points and the connections between the various nodes. When the nodes are affected, the conduits between those nodes are all affected, following the same principles as a single point of failure from the Business Continuity domain. When only one of many conduits between nodes is affected, the system may be able to adjust accordingly (such as we would see where aircraft and ships are routed to new airways or shipping lanes in response to bad weather). What should be clear to the reader, however, is that there is a level of

https://scholarcommons.usf.edu/jss/vol2/iss2/4
DOI: http://dx.doi.org/10.5038/1944-0472.2.2.4

dependence and independence in how this impact operates. If the node is configured with set points and does not have the ability to reprioritize or adjust its own configuration (such as paid gates at an airport); then the individual lines or conduits of disruption function independently of each other. This results in a condition where the sum of disruptions must be taken into account. On the other hand, where a node has the ability to adjust and prioritize accordingly, then the calculation of disruption moves much more in line with those associated with dependent events.

Consider this example; if each service point held by the node is firmly and irrevocably allocated to one family of lines, then only those service points need to be disrupted for the whole family of lines to be disrupted. If, however, there are clauses and similar mechanisms built in that allow for the organization to move a disrupted line from one family into another group of service points (likely with a cost), then the final failure of the event is not determined by whether or not all of one administrative group of service points are disrupted. At the tactical level, the organization needs to arrange its service points and its contracts to prevent a single incident from affecting all service lines to allow for flexibility in its operational context.

Fragility, at this point, indicates the potential for fragmentation and we must, therefore, take into account the risk of disruption looking at both the infrastructure side of the equation and also the administrative side of the equation. Where all elements are vulnerable to a certain kind of attack, for instance, and there are no other options available, then the system is fragile. At the regional level, if there is only one option available that can meet all services, this local or tactical fragility can quickly affect the regional fragility—meaning that the regional system is vulnerable to a single attack at a certain point.

Finally, by looking at how the capacity lost as a result of that disruption affects the overall system—movement to and through—we can calculate the disruptions due to fragmentation and dissolution. Fragmentation occurs at key points that segregate or connect the various nodes of a network. These might be referred to as the hubs in the transportation system. Depending upon whether or not the physical, procedural, technical, and psychological measures are in place to connect behind those nodes (such as a couple of airports with the correct runways, landing systems, communications systems, trained personnel, etc), the system will cut away from the network. This leads to fragmentation.

Fragmentation and its associated impacts eventually lead to a situation where demand in the system cannot locate any reasonable route by which

it can accomplish its goals. When this happens, the system gradually fills and then collapses under its own weight. Essentially, it is overwhelmed by a shift in the demand to capacity ratio. The end result is a condition where the pressure in the system has to be relieved to such a point that it can restart its operations and generate the capacity necessary to meet demand.

Fragility, in a networked transportation system operates across all three levels: local, regional, and national or strategic. The strategic and regional levels are founded upon the capacity and vulnerabilities inherent at the local levels and then exacerbated through regional and national disconnects, the lack of redundancy, and similar factors. At the local level, an understanding of fragility must be combined with the need to conduct appropriate impact assessments in addition to understanding the vulnerabilities associated with each input that allows for work to proceed so that the potential capacity to be delivered can actually be communicated or delivered into the system. Failing to understand the concept of fragility at the local level can lead to a misinterpretation of impacts by failing to understand how the infrastructure delivers capacity within in the broader regional or national context.

Author Biography

Mr. McDougall has held positions as the Senior Inspector for Ports and Marine Facilities at Transport Canada, National Coordinator Security Policy and Projects at the Department of Fisheries and Oceans, and Compliance Auditor for Fleet Security within the Canadian Coast Guard. In each of these roles, he has provided guidance and advice with respect to Physical Security and Infrastructure Assurance ranging across the local, regional, and strategic levels of government and into the private sector. He has a BMAS from the Royal Military College and a BA from the University of Western Ontario. He also has certifications in Critical Infrastructure Protection (PCIP), Antiterrorism (CMAS), and Information Systems Security (CISSP). He currently contributes on a variety of training projects within the Transportation Security community (including as a recognized trainer under the International Maritime Organization's Train the Trainer program and Transport Canada's marine security programs) and has contributed on a number of specialized courses, including the Strategic Leadership in Federal Government Security course offered to emerging Departmental Security Officers.

https://scholarcommons.usf.edu/jss/vol2/iss2/4
DOI: http://dx.doi.org/10.5038/1944-0472.2.2.4