Cyberspace and Infrastructure

William D. O'Neil
w.d.oneil@pobox.com
Author's final draft
23 Jan 2008

This chapter addresses the related and overlapping but non-identical subjects of (1) protecting infrastructures against cyber attack and (2) protecting cyber infrastructure against all forms of attack.[1]

The history of infrastructure attack – a sketch

Infrastructure attack is a story as old as war. Time out of mind attackers sought to cut off their target's water supply and transportation, often with decisive results.

The rise of modern infrastructure systems, starting in the 19th century, quickly brought heightened concerns about vulnerability. As one widely-read futurist and social critic put it in 1929:

> [S]omething on the order of one hundred key men, opening its veins of water, power, gas, sewage disposal, milk supply, [and] communication, could bring the life of a great city to an end—almost as neatly as though its every crevice had been soaked with poison gas. Even in rural areas with the growing use of electric power, the telephone, gasoline, and imported foodstuffs, the factor of dependence on an unknown technology is very great. … The machine has presented us with a central nervous system, protected with no spinal vertebrae, lying almost naked for the cutting. If, for one reason or another, the severance is made, we face a terrifying, perhaps mortal crisis.[2] …
>
> Day by day the complexity, and hence the potential danger, accelerates; materials and structures ceaselessly and silently deterio-

---

1 A somewhat different version of this appears as "Cyberspace and Infrastructure," Chapter 5 of *Cyberpower and National Security*, ed. Franklin D. Kramer, Stuart H. Starr, and Larry Wentz (Washington: National Defense University Press and Potomac Books, 2008).

2 Stuart Chase, *Men and Machines* (New York: Macmillan, 1929), pp. 288-9.

> rate. One may look for some very ugly happenings in the next ten years.[3]

In the United States, in particular, early air power enthusiasts drawing on these currents of thought became convinced that infrastructure attack held the key to victory in modern war. Infrastructures – especially the electric grid – were seen as relatively easy to take down and so critical that their slightest disruption would severely disrupt war-making potential and economic activity generally. Electrical generation and distribution was identified as being particularly vulnerable and crucial.

When war came, however, Air Force planners decided that electric power was not as critical as previously thought and turned their attention to other target complexes. Later analysis suggested that this was probably an error and that attacking power plants would have been quite productive. German electric production was curtailed when attacks on the rail infrastructure cut back coal shipments severely, but this came quite late – in part because concerted attack on transportation was not decided upon until late in the war. Japanese electric production (largely hydroelectric) was even less affected by bombing.[4]

Of the 1.5 million tons of air ordnance delivered by U.S. forces against German and German-held targets in 1943-45, 41% fell on transportation targets (largely rail) and a further 6% on oil, chemical, and synthetic rubber targets.[5] In the war against Japan, naval submarine and surface forces as well as air forces devoted much of their weight of attack to enemy transportation, particularly at sea.[6] In both cases it was concluded that attacks on transportation infrastructure had severely affected enemy war effort and that even greater effort against transportation would have been very worthwhile. (Attacks against Japan's oil infrastructure were judged to

---

3 Ibid., p. 297.

4 Mark Clodfelter, "Pinpointing Devastation: American Air Campaign Planning before Pearl Harbor," *The Journal of Military History* 58 (Jan 1994): 75-101; United States Strategic Bombing Survey (USSBS), *Over-All Report* (*European War*) (Washington: Government Printing Office (GPO), 1945); Over-All Economic Effects Division, USSBS, "The Effects of Strategic Bombing on the German War Economy" (Washington: GPO, 31 Oct 1945); Idem, "The Effects of Strategic Bombing on Japan's War Economy" (Washington: GPO, Dec 1946).

5 USSBS, "Statistical Appendix to Over-All Report (European War)" (Washington, Feb 1947).

6 USSBS, "The War Against Japanese Transportation, 1941-1945" (Washington: Transportation Division, May 1947).

have been "almost superfluous" because the war on transportation had already largely idled the refineries for want of feedstocks.[7])

Wars since 1945 have continued to feature attacks on transportation and often on oil as well. The 1991 Gulf War included major campaigns against both and added systematic attacks on Iraq's communications infrastructure.

Since World War II U.S. bombing campaigns generally have made electric power infrastructure a major target. The best documented case is the Gulf War, in which 88% of Iraq's electric grid capacity was knocked out, most in the first few days of the war.[8]

Guerilla and terrorist forces as well have targeted infrastructure in many conflicts. In the internal conflicts that followed the invasion of Iraq in 2003 there have been repeated attacks on infrastructures, especially those for electric power and oil.[9]

So far as has been publicly revealed there have as yet been no military campaigns against the infrastructure of cyberspace, nor any military cyber attacks on other infrastructures. But there have been a great many attacks by "hackers" whose identities and motives often are shadowy, and some believe that some of these have been state sponsored.

We will look later at some of the lessons which can be drawn from this history regarding defending infrastructures. But first it will help to probe a little into the nature of infrastructures themselves.

Networks

It is a common observation that infrastructures often depend on networks. We speak of the *road network*, the *rail network*, the *telephone network*, the *electricity network*, and more recently the *Internet*. The network consists of the points of supply or origin, the routes of transportation or movement, and the points of destination or consumption. Each infrastructure is physically distinct but as the terminology suggests they share something important at the level of abstract structure.

---

7 USSBS Over-All Economic Effects Division, "The Effects of Strategic Bombing on Japan's War Economy," pp. 46-7.

8 Thomas A. Keaney and Elliot A. Cohen, *Gulf War Air Power Survey: Summary Report* (Washington, 1993).

9 James Glanz, "Iraq Insurgents Starve Capital of Electricity," *New York Times* (19 Dec 2006).

The theory of networks in fact is important to a great many fields of science and technology. It has been studied intensively and a few of its relevant findings are very briefly outlined in Appendix A to this chapter.

As is shown in the appendix, scientists studying large, irregular networks (which most infrastructures are) have identified two broad classes of importance. In one, most nodes have roughly the same number of links and nodes with a great many more links are very rare. Networks with this sort of egalitarian structure are often called *exponential networks* (because highly-connected nodes are exponentially unlikely) or, more descriptively, *uniform-random* networks.

(It seems odd to speak of "random" networks in connection with infrastructures, where the design choices are not made by rolling dice or drawing from a hat. But infrastructures do tend to develop through a sequence of choices reflecting a variety of changing considerations, and this gives them a certain statistical or random-like character.)

In the other major class, nodes tend to be preferentially connected to nodes which already have a great many connections – the rich get richer – but still with a random (or random-like) element. These are called *power-law networks* for technical mathematical reasons which reflect their relative abundance of highly-connected nodes. Even more commonly they are called *scale-free networks*, a reference to their lack of a dominant typical or average scale in the sense of number of connections per node. A more descriptive term might be *hub-and-spoke random network*.

These two kinds of nets are illustrated in Figure 1. It is obvious immediately that the diagram on the right, **b**, showing a scale-free power-law network, has a number of highly-connected hub nodes as well as a great many nodes with only a single link connection. The uniform-random like exponential network on the left, **a**, has a much more uniform pattern of connection, contrasting sharply with **b**.

If an accident or attack were to disable a node picked at random from the exponential network at the left of the figure, it usually would disconnect only a handful of other nodes that happen to connect uniquely to the one disabled. In the scale-free network a random disablement will do even less damage in most cases since so few nodes have any other node which connects only through them. But the worst case is worse in **b** than in **a** – taking out only a few of **b**'s highly connected hubs does a lot of damage. This is an important distinction in terms of infrastructure protection.

Cyber networks

So far we've addressed only the *topology* of the networks, the logic of how nodes and links connect without regard to the physical nature and spatial location of these elements. We need to look into these factors as well in order to understand the issues involved in protecting networked infrastructure systems.

Infrastructures are not only networks but networks upon networks. This is outlined with respect to cyberspace by Ed Skoudis in Chapter 7 and by Marjory Blumenthal and David Clark in Chapter 8. **Table 1** sketches the levels involved in the cyber network in order to emphasize that at bottom cyber content rests on a structure of physical elements which have physical properties and locations.[10] Even though its topology is not identical with that of the network layers it is built upon, cyberspace itself ultimately has a geography as well as a topology, both affecting its vulnerability and survivability. The same is so, *mutatis mutandis*, of other infrastructure networks.

The topologies of the Internet and World Wide Web are the subject of particularly intense study, for a variety of reasons – including their complexity, availability for study, and practical importance.[11] No simple model can fully capture the complexity of these structures, but it is well established that in broad terms they are both power-law or scale-free networks to a close approximation.

Scale-free networks arise typically through growth in which new nodes link preferentially to old nodes that are already highly linked, thus forming very highly connected hubs. It is easy to see how this happens in the Web, where it costs no more to link a Web page to a richly connected hub Web site such as http://www.google.com/ or http://en.wikipedia.org/ than to an isolated site run by a specialized organization such as http://www.ndu.edu/ctnsp/ or an individual such as http://www.analysis.williamdoneil.com/.

Internet nodes consist of computers (or devices which incorporate computers). In the simplest (and very common) case this is a single isolated computer in a home or small business establishment. The cheapest connection one could make would involve running a cable or wireless link to

---

[10] A closer look would reveal many more layers within those shown. Analyses of the Internet Protocol stack, for instance, often are carried to seven or more layers.

[11] Dmitri Krioukov, *et al.*, "The Workshop on Internet Topology (WIT) Report," *Computer Communication Review* 37, no.1 (2007): 69-73.

the computer next door. In most cases, however, this would not accomplish very much since usually only a small portion of our information needs can be supplied by our immediate neighbors. Of course I might be able to piggyback on the information channels available to my neighbor, but this would cut into the bandwidth available to him and so would be unattractive from his standpoint. So even though it costs more we generally buy our service from an Internet service provider (ISP) who offers a connection to his hub or server bank (a group of high-speed computers, usually housed in a single warehouse-like building) via some miles of telephone wire, coaxial cable, fiber-optic cable, wireless cellular radio link, or satellite radio link. Higher bandwidth connections which provide greater information capacity cost more but most users find the expense worthwhile and users in areas where they are not available frequently complain bitterly.

An ISP whose server bank services thousands of high-speed connections over a an area of many acres or square miles faces similar choices. Connections to nearby ISPs would be relatively inexpensive in terms of the cost of the cable but would not meet his needs for a rich flow of data to meet his customers' demands. Thus the ISP finds it worthwhile to pay for a very high bandwidth connection to a major hub with a massive server bank that handles a great deal of Internet traffic and thus is able to tap its riches. It is clear that processes like these, repeated at all levels, drive the Internet toward a hub-and-spoke scale-free architecture resembling that of Figure 1**b**.

We noted earlier that scale-free networks are robust in the face of random or undirected failures, which fall most heavily on the large numbers of nodes with only a few connections. The experience of the Internet has borne this out. Nodes often fail or are shut down for a variety of reasons but this has scarcely any discernable effect on overall network performance. Even more massive failures, such as those caused by widespread power outages or the 9-11 attacks, have been quite localized in the effects. (Of course such incidents can generate a surge in traffic which may slow response in itself, but this is not a vulnerability of the Internet *per se*.)[12]

In contrast to the extensive experience with random outages there has been little with outages that preferentially target the Internet's major hubs. Nevertheless, there is every reason to think that what is true in theory would hold equally in practice – that successful attacks on many of the biggest hubs would have severe and pervasive effects, with a great many Internet

[12] Computer Science and Telecommunications Board, *The Internet Under Crisis Conditions: Learning from September 11* (Washington: National Academies Press, 2003).

nodes isolated or able to communicate with only a small number of other nodes. Thus protection of major Internet hubs is a cornerstone of rational policy for cyberspace infrastructure defense.

From Figure 1**b** it would seem that link outages are much less of a concern in a scale-free network like the Internet than are outages of key nodes. Severing links could not do as much damage to network connectivity as disabling an equal number of critical nodes. A closer look at the physical layers underlying the Internet, however, shows that this may be too sanguine a view in practice. Links which are logically and topologically separate may in fact be carried over the same physical communications infrastructure through multiplexing. (Indeed, links of entirely separate networks, having no logical interfaces at all, may be multiplexed onto one fiber-optic strand.) Even if they use physically separate communications lines, it is very possible that the lines for multiple links may share the same conduit or otherwise be vulnerable to the same damage agents. Thus a single attack might take out a great many links at one time – thousands or even tens of thousands. In critical cases this could cut multiple nodes off from the network. The places where this can occur likewise must be protected to assure cyberspace infrastructure integrity.

This is a particular concern for nodes located in physically isolated sites, as many critical to national security are. Where economy or convenience are the dominant considerations, such sites often may be served by only one or two pathways for all communications links.

The electrical network

Loss of electricity does not ordinarily take down a major Internet hub, at least not at once, since most have emergency backup sources that can carry them for hours or days. In a larger sense, however, the Internet clearly is critically dependent on electric supply, as is virtually our entire society. **Figure 2** provides a schematic view of the central role of the electricity supply network, which we usually know as the *electric grid*.

Figure 3 shows an electric grid, that of California. It is immediately clear that its topology is not very much like that of the power-law or scale-free network shown in Figure 1**b**. While there are core areas representing major centers of population and industry, there are no hubs with large numbers of nodes connecting directly to them and most nodes have more than one link. Indeed, quantitative studies show that electric grids lie closer in topology to the uniform-random, exponential network depicted in Figure 1**a**. Although the electric grid, like the Internet, grows and changes as nodes

and links are added, modified, or sometimes deleted, its economic and technological forces are quite different and result in a different kind of pattern. These forces are changing in important ways today and will result in a different grid eventually, as will be discussed below. We must look first at historical forces to understand today's grid, with some notes along the way regarding changes.

Even though it comes in different forms – alternating or direct current at any of a number of voltage levels – electricity is all of a kind.[13] With suitable conversion of form, any electrical energy will serve any electrical load.[14] It is a bulk commodity lacking in the specificity that distinguishes information. Electricity is most economically generated in bulk, resulting in a grid dominated by a relatively small number of large central station plants usually located at or near their energy sources. (This may well change as the costs of carbon emissions and other environmental damage are figured into the cost of generation; some of the generating technologies with low environmental impact, such as wind turbines and solar-electric systems, may favor smaller-scale operation.)

Electricity is also most economically transported in bulk, at high levels of energy and voltage. (The advent of new transmission technology such as high-temperature superconductors (HTS) may reduce the advantage of high-capacity transmission but not void it entirely.)

Neither bulk generation nor bulk transmission in themselves dictate a uniform-random electric network. A key factor is that most electrical transmission is in alternating current (AC) form at high voltages and most electrical use is AC at lower voltages. A relatively simple passive device, the transformer, allows high-voltage AC (HVAC) to be tapped at lower voltage with scarcely any loss of energy. Thus major corridors are served by a few high-capacity HVAC lines along which are strung distribution stations feeding local bulk users as well as local retail distribution networks. The corridors themselves are determined by economic geography – where the customers are. Of course there is some feedback since customers may find it economical to locate in corridors served by major transmission facilities. **Figure 4** illustrates the relationships among the major components of the electric grid.

---

[13] For an overview of the electricity grid see Jack Casazza and Frank Delea, *Understanding Electric Power Systems: An Overview of the Technology and the Marketplace* (Hoboken, New Jersey: IEEE Press and Wiley-Interscience, 2003).

[14] In electrical terminology any equipment or system which draws electric power is a *load*.

Except for the very largest of customers who use electric power on truly massive scales it is more economical to draw power from a nearby distribution station than to run lines directly to a distant central station. Because the distribution station draws its power from a major HVAC line, it can supply large quantities, and since all power is the same it makes no difference where it comes from as long as it is sufficient in quantity. This is why when we look at a portion of the electric grid such as shown in **Figure 3**, we see a network that more closely resembles the uniform-random pattern of Figure 1**a** than the scale-free hub-and-spoke layout of Figure 1**b**. Thus the electric grid is a fundamentally different kind of network from the Internet.

The earliest commercial electric utilities used direct current (DC). AC won out as the standard in large part because it is so easy to tap HVAC transmission lines with transformers to produce lower voltage for distribution and final use – and so difficult and costly to step down from HVDC. (Transformers do not work for DC and there is no simple DC equivalent.) DC continued in use for specialized local applications (such as shipboard electrical systems) for some decades but these applications too gradually died out. For moving very large flows of energy over long distances, however, HVDC lines can be more economical than HVAC. This has led to the use of HVDC *intertie* lines connect distant "islands" of intense electric use across wide stretches with little use, foregoing distribution stations. (To distribute current from HVDC lines it is necessary first to convert it to AC.) Looking at the relatively isolated California electric grid in **Figure 3** we see several HVDC intertie lines which connect to distant areas. Today the North American electric grid (encompassing the Continental United States, Canada, and a small portion of Northwestern Mexico) is divided into four large regions which connect almost entirely via HVDC links. We will see later how this greatly reduces the risks of a continent-wide grid failure.

When there are two or more possible routes from generator to load electricity will flow over all of them, with the greater amount following the paths with lower resistance. If one path is cut off the flow automatically redirects itself over the remaining links. When the flow in a transmission network is near the limits of its capacity to handle power flow without breakdown the failure of one link can throw more load on remaining links than they can carry. This leads to a cascade of failures as links either break down (due to overheating) or are shut down to save them from damage.

On an AC network the current everywhere must alternate at the same frequency in what is called *synchronous* operation. Any failure of this frequency synchronization produces unbalanced forces that can literally tear

equipment apart. Synchronization failures too can cascade as generation or transmission equipment drops off line to avoid catastrophic failures.

The loading on the grid varies from moment to moment and the organizations responsible for its operation have only limited tools for managing it. Users can add loads by throwing a switch while generators and transmission equipment can go off line for a variety of reasons. Grid operators may have the ability to shed some loads (customers who have bought "interruptible power" at reduced rates) but this is very limited. In an emergency a block of customers in a particular area may be blacked out to shed load, but many systems are not set up to allow this to be done quickly and in any event utilities are naturally reluctant to do this except as a last resort. An overstressed link or node may have to be shut down, which increases the load on other components. If local overloading drags the frequency of a generator down then it and the area it serves must be immediately disconnected from the grid. On a wide scale this can cut the grid up into isolated islands, many or all of which may fail under local load imbalances.

Where will a failure cascade stop? Could one engulf the entire North American electrical grid, pitching the whole continent into the dark? Two factors make this unlikely. First, like the wave raised by a rock thrown into a pool the disturbance following a major fault in the grid weakens as it spreads. Beyond that, the HVDC intertie lines that link the four major synchronous regions in North America serve to isolate them from frequency disturbances in other regions. Regardless of what may happen in any one region the others should be able to adjust and continue normal operation without major disruption.[15]

Readers of a certain age may recall that prior to the mid 1960s widespread grid failures were unknown. In earlier periods electricity was very largely local. With a few exceptions (mostly involving large hydropower systems) most electric power was generated, distributed, and delivered within a compact area served by a local power utility that enjoyed a regulated monopoly. This cellular structure meant that there was little opportunity for failures to spread beyond a single utility. Moreover,

---

[15] There is a partial exception to this in that one of the synchronous regions, that in Northeastern Canada operated by Hydro-Québec, is a major power exporter from its large hydroelectric generating facilities. The power is almost all exported via an HVDC line which prevents frequency disturbances from spreading, but a sudden major voltage disturbance in this region or the Northeastern U.S. region it sells power to would put the other region under stress.

regulators held utilities responsible for reliability of service and could apply effective sanctions when reliability fell short.

Although this regime had led to steadily decreasing electrical prices for decades as the utilities incorporated new technology, economists and political leaders argued that the monopolistic structure, even with regulation, was economically inefficient and led to added cost. At the same time new technologies appeared to offer the potential to economically generate and transmit electricity on scales that transcended the bounds of individual utility companies, however large. As a result, from the mid 1970s the federal government moved to "deregulate" electricity – in fact to change the regulatory basis so as to encourage competition in generation and transmission. States have followed suit, although not in a uniform way, leading to further fragmentation of ownership and control over generation and distribution equipment and operation.[16]

Deregulation in general has not been followed by further significant decreases in the costs of electricity, although it is argued by proponents that it resulted in avoidance of large increases as well as other benefits. It has, however, helped to open the way to other problems which were not fully anticipated and which are still being worked out.

Because every part of the grid influences every other part, it has been difficult to construct a deregulation regime which would allow the truly independent operation necessary for fully effective competition. In 2000 and 2001 Enron Corporation and other power producers and speculators exploited the physical properties of California's electricity grid in combination with its "deregulated" operating rules to manipulate prices to their great advantage, at the same time causing or exacerbating electricity shortages in the State. Analyses of this event show clearly how difficult the problems of ensuring the smooth running of the physically tightly coupled but economically fragmented electric market system are.[17] The same limitations that permit participants to impose costs on others without inherent limits (beyond those interposed by the remaining regulators) equally allow some serious technical problems to develop and spread

---

16 Richard F. Hirsh, *Power Loss: The Origins of Deregulation and Restructuring in the American Electric Utility System* (Cambridge: MIT Press, 1999.)

17 Frank A. Wolak, "Diagnosing the California Electricity Crisis," *The Electricity Journal* 16, No. 7 (Aug 2003): 11-37.

without any individual participating firm or organization seeing it in its direct interest to take corrective action.[18]

The overall commitment to deregulation nevertheless remains strong on the part of legislators and public officials. But even if restoration of the earlier regime of regulated local vertical power monopolies were politically and economically feasible it is not clear how it could work physically today. Many regions have come to depend on power generated in distant places and transmitted over long distances. Heavy long-distance power flows have become a fact of life and any attempt to again divide the grid into relatively small, self-sufficient cells operated by separate local firms would involve not only major investment costs but serious environmental concerns. But without some such structure there is no simple way to assign responsibility for maintaining adequate and reliable power service.

The physics of electricity simply does not allow a fully laissez faire, every-man-for-himself operating regime. Just as on the highway there must be some consistent set of operating rules which everyone is constrained to obey if the system is to operate stably and safely. Despite warnings this is a realization which has been somewhat slow in emerging, perhaps in part because authorities were thinking in terms of analogies with networks which were not as tightly coupled as the electricity grid and thus less in need of tightly disciplined operation.

Later we will discuss what has been done to address the policy issues raised by these facts and what more may need to be done.

Lessons from a blackout

An illustration of how tightly coupled the grid is and what this implies for its operation and protection is provided by major outages. The most recent of these in North America occurred on 14 Aug 2003 and eventually covered large areas of the United States and Canada, affecting electric service to approximately 50 million Americans and Canadians for an extended period.[19] The extent of the blackout is graphically illustrated by **Figure 5**.

---

[18] Secretary of Energy Advisory Board, "Maintaining Reliability in a Competitive U.S. Electricity Industry: Final Report of the Task Force on Electric System Reliability" (Washington: Department of Energy, 29 Sep 1998).

[19] U.S.-Canada Power System Outage Task Force, "Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations" (Washington and Ottawa: U.S. Department of Energy and Natural Resources Canada, April 2004) is the comprehensive official report. It is very illuminating about the mechanisms of failure. Also useful is North American Electric Reliability Council, "Technical Analysis of the

The process which led to blackout started near Cleveland after 3 p.m. The day was hot and air conditioning loads were heavy but not near the heaviest which the system had handled before. Investigation revealed a number of hardware and software failures together with faulty operational procedures on the part both of the local utility operator and the organization responsible for ensuring the reliability of the grid in the region. Most of these did not directly contribute to the blackout, but it is clear that many of them could have led to major failures under slightly different circumstances. Many aspects of the operation were in violation of accepted (but then voluntary) industry standards. Even if all equipment and software had been functioning properly and fully in compliance with existing standards, however, the tools available to the operators for system awareness would have been critically limited.

The immediate cause of the blackout was a series of instances of high-voltage transmission lines contacting trees in their rights of way which had been allowed to grow too tall. Autonomous safety systems sensed the resulting ground faults and automatically disconnected or "tripped" the lines to prevent more serious damage and fires. Operator response was very poor, in part because critical warning and analysis systems had failed. One by one the faults accumulated until the point was reached at which human intervention could no longer be effective. Over a period of eight seconds, from 4:10:37 p.m. to 4:10:45, automatic safety relays all over the Northeast shut down lines and generators that had violated pre-set acceptable operating limits, severing grid links and blacking out areas throughout the region in a process shown in **Figure 6**.[20]

Not withstanding the various hardware and software failures, if the operators had been well trained and effective in applying the existing procedures – despite the limitations of the procedures – the huge blackout would never have occurred in the way that it did. At worst, a very limited

---

August 14, 2003, Blackout: What Happened, Why, and What Did We Learn?" (Princeton: NERC, 13 Jul 2004). For related news account see Richard Pérez-Peña, "Utility Could Have Halted '03 Blackout, Panel Says," *New York Times* (6 Apr 2004): A:16. "Blackout 101," a series of tutorial presentations developed by experts to inform Congress, is at http://www.ieee.org/portal/site/pes/menuitem.2b4756efb9a16c58fb2275875bac26c8/index.jsp?&pName=pes_level1&path=pes/subpages/meetings-folder/other_meetings&file=Blackout_101.xml&xsl=generic.xsl, or http://tinyurl.com/yur6o4.

[20] In many cases the lines and generators were not actually immediately threatened, but only appeared so as an artifact of the large power surges triggered by the cascade. If the safety relays had been better able to discriminate between real and apparent threats the outage could have been much less widespread.

area with a few tens of thousands of customers might have been affected for a few hours. Indeed, part of the problem was excessive and inappropriate operator reliance on limited and fallible warning and diagnosis systems.

Attacks against the electrical infrastructure

Coming just 23 months after the 11 Sep 2001 attacks it was natural to wonder whether the Northeast blackout could have been caused or worsened by terrorist attacks. Indeed, claims of responsibility purportedly from Al-Qaeda appeared within a few days of the outage. Moreover the "Blaster" Internet worm had first been seen on 11 Aug, leading to speculation that it might have been involved. Investigation showed that the alleged Al-Qaeda attack had not occurred and that Blaster was not involved, but also revealed significant potential vulnerabilities.[21]

As previously mentioned, software failures had played a role in the process which set the stage for the Aug 2003 Northeast blackout by denying operators information tools they were accustomed to. These failures were accidental and/or intrinsic to the system design, but as we shall see it is conceivable that comparable failures could have resulted from cyberspace attacks. In general the operators of the grid rely on a variety of cyberspace services to gather operating information, communicate with other control personnel, and issue instructions. Of particular concern are SCADA (supervisory control and data acquisition) and energy management systems (EMS). These gather data on the operational parameters of equipment throughout a particular segment of the electric grid, report it to a central location and present it to operators, and change set-points for equipment controllers in response to operator decisions about system configuration and operation.

We will address the policy and standards efforts being undertaken to meet these problems later in this chapter.

The (secure) grid of the future?

Just as there is a good deal of thought about the Internet of the future so have engineers and policy-makers devoted attention and development efforts to defining the future of the electric power grid. The visions differ in detail but generally involve a "smart grid" able to adapt to failures in real

---

[21] U.S.-Canada Power System Outage Task Force, "Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations," (Washington and Ottawa: U.S. Department of Energy and Natural Resources Canada, April 2004), pp. 131-7.

time with limited if any degradation.[22] In a sense this represents a return to the cellular structure of the pre-deregulation grid, but with smaller cells that are regulated by software rather than governmental agencies and with provision to take advantage of distant power sources. One key is distributed local power sources and perhaps power storage systems. Fuel cells in particular appear to offer promise of small-scale but highly-efficient generating units that would serve these purposes.[23]

Growing concern about global climate change may affect these visions in various ways and directions. One possibility is a renewed emphasis on very large nuclear central station generating plants. This offers zero emissions of greenhouse gases – even better than the reduced emissions of fuel cells – but would represent a step toward greater concentration of power generation.[24] Solar power is another zero-emissions option and could integrate more naturally into a cellular structure, although efficient and economical means to store the energy from solar systems for release in the hours of darkness must be found if they are to become a major electrical source.[25] Effective use of wind power at large scale depends on solutions to the challenges posed by its unsteady flow.[26] Various other advanced technologies for energy production are more speculative at this time.[27] None of the alternative sources so far conceived can obviate or substantially modify the need for a more reliable and robust grid for electrical transmission and distribution, and in most cases they would bring additional complexity due to their limited ability to provide steady and

---

22 S. Massoud Amin and Philip Schewe, "Preventing Blackouts," *Scientific American* 296, no. 5 (May 2007): 60-7; S. Massoud Amin and Bruce F. Wollenberg, "Toward a Smart Grid," *IEEE Power and Energy Magazine* 3, no. 5 (Sep-Oct 2005): 34-8; and Clark W. Gellings and Kurt E.Yeager, "Transforming the Electric Infrastructure," *Physics Today* 57, No. 12 (Dec 2004): 45-51. See also "Ideas Generated for Transforming the Electric Infrastructure," *Physics Today* 58, No. 5 (May 2005): 13-15.

23 Supramaniam Srinivasan, et al., "Fuel Cells: Reaching the Era of Clean and Efficient Power Generation in the Twenty-First Century," *Annual Reviews of Energy and the Environment* 24 (1999): 281-328.

24 John M. Deutch and Ernest J. Moniz, "The Nuclear Option," *Scientific American* 295, No. 3 (Sep 2006): 76-83; James A. Lake, Ralph G. Bennett, and John F. Kotek, "Next-Generation Nuclear Power," *Scientific American* 286, No. 1 (Jan 2002): 72-81.

25 George W. Crabtree and Nathan S. Lewis, "Solar Energy Conversion," *Physics Today* 60, No. 3 (Mar 2007): 37-42; Ken Zweibel, James Mason and Vasilis Fthenakis, "A Solar Grand Plan," *Scientific American* 298, No. 1 (Jan 2008): 64-73; Daniel M. Kammen, "The Rise of Renewable Energy," *Scientific American* 295, No. 3 (Sep 2006): 84-93.

26 Kammen, Op. Cit.; and Karl Stahlkopf, "Taking Wind Mainstream," *IEEE Spectrum* (Jun 2006).

27 W. Wyatt Gibbs, "Plan B for Energy," *Scientific American* 295, No. 3 (Sep 2006): 102-14.

continuous power or to rapidly vary their output in response to load fluctuations.

Virtually all proposed schemes for improved electricity delivery depend on networked smart control – which is to say that they depend more on cyberspace. For the most part proposals to date have devoted little attention to security against cyber attack. Clearly, this must change before much further work is done along these lines in order to ensure that efforts to improve the reliability and efficiency of power distribution do not increase vulnerability to attack.

Pipeline networks

The electrical infrastructure is unique both in its degree of coupling and its central role, but other infrastructures present parallel concerns even if at lower overall risk levels. This is particularly true of two other major energy sector infrastructures, oil and natural gas.[28] Both are also networked infrastructures, with about 170,000 miles of oil pipelines and 1.4 million miles of natural gas pipelines.[29] More than three quarters of U.S. crude oil supplies are carried by pipeline and about 60% of refined products, while virtually all natural gas flows by pipeline.[30]

Notwithstanding the very obvious dangers inherent in pipes filled with flammable and potentially explosive fluids, the overall safety record of U.S. pipeline systems is good. And despite their obvious vulnerability to sabotage there have been few attacks or attempts on U.S. pipelines, at least to date. So far all known threats have been of physical attack, not by cyber means.

Both oil and gas pipelines make use of SCADA and operational management systems, although not at the same level as the electrical infrastructure. The issues of cyber security in these infrastructures are generally similar to those affecting electricity.

Infrastructure threats

The operators of infrastructure systems of all types routinely face a spectrum of threats, whether from natural causes (e.g., lightning,

---

28 The Nation's 4 million miles of public roads, 100,000 miles of Class I rail lines, and 26,000 miles of waterways also represent networked infrastructures, but are not dealt with here because of their lower vulnerability to cyber attack.

29 Department of Transportation, "National Transportation Statistics, 2007," Table 1-10, http://www.bts.gov/publications/national_transportation_statistics/.

30 Paul W. Parfomak, "Pipeline Safety and Security: Federal Programs," CRS Report for Congress RL33347 (Washington: Congressional Research Service, 11 Jul 2007) pp. 1-2.

earthquakes, hurricanes), intrinsic faults (e.g., stuck valves or circuit breakers, failing electronics, or unstable software), or criminal action (e.g., by vandals, thieves, extortionists, or hackers). Motivated by a community sense of responsibility, regulatory and legal requirements, and economic self-interest, they take action to avert these threats, minimize the damage they do, and recover rapidly from damage which does occur. Many national security threats resemble more intense and deliberate versions of these normal infrastructure threats, emphasizing the need to integrate all aspects of infrastructure protection.

Our special focus here, of course, is on cyber threats. There have been a number of attacks on infrastructure cyber systems but not coordinated large-scale attacks. Damage to date has been limited. However, the CIA has warned of the threat of cyberattack against electrical utilities especially, saying that "cyber attackers have hacked into the computer systems of utility companies outside the United States and made demands, in at least one case causing a power outage that affected multiple cities." "We do not know who executed these attacks or why, but all involved intrusions through the Internet," the CIA analyst reported, adding, "We suspect, but cannot confirm, that some of the attackers had the benefit of inside knowledge."[31] These are not the sole examples of attacks, however, and some have been domestic. Generally they have received no publicity in an effort to avoid giving the attackers useful feedback.[32]

In some cases it has been clear that cyber-extortionists were behind the attacks, but in most instances the identity and motivation of the attackers is unclear. Following the 11 September 2001 terrorist attacks it was widely predicted that al Qaeda would follow up with massive cyber attacks on infrastructure targets but these have not materialized and the likelihood of large-scale cyber-infrastructure attacks by terrorists is disputed.[33]

Even if terrorists never find the means or motivation to do so there is little doubt that a conventional state enemy determined to mount a military attack on the U.S. could and very well might launch massive coordinated

---

[31] Ellen Nakashima and Steven Mufson, "Hackers Have Attacked Foreign Utilities, CIA Analyst Says," *Washington Post* (January 19, 2008): A4.

[32] Andy Greenberg, "America's Hackable Backbone," *Forbes.com* (22 Aug 2007) http://www.forbes.com/2007/08/22/scada-hackers-infrastructure-tech-security-cx_ag_0822hack.html.

[33] John Rollins and Clay Wilson, "Terrorist Capabilities for Cyberattack: Overview and Policy Issues," CRS Report for Congress RL33123 (Washington: Congressional Research Service, 22 Jan 2007).

cyber attacks on infrastructure as a part of an overall strategy of infrastructure attack. Thus we need to ask how much damage can be done by cyber means and what may be done to limit it.

A crucial question is the extent to which systems can be accessed via the Internet. Wide-open access is rare but many systems may have some Internet portals through which an attacker might be able to reach critical functions. In most cases there are active efforts to close these or at least provide highly secure protection, and to the extent these efforts succeed it will be impossible to attack systems from the Internet. This does not rule out attacks via individuals with inside access, however

The consequences of failures are always foremost in the minds of the engineers who design infrastructure systems and components. Well aware that complex systems generally and software particularly are prone to failure they design to limit the consequences. Where possible critical control functions with the potential for severe equipment damage are lodged with simple, entirely autonomous, self-contained systems such as governors and overload trips. Thus an EMS may be able to overload a transmission line with current but cannot prevent the circuit breakers from tripping and opening the circuit to forestall damage.

This strategy is not universally applicable, however. In an aircraft with a fly-by-wire control system, for instance, it is inherently possible that a run-away of the system could fly the airplane into the ground or exceed its limiting flight loads. When such vulnerabilities are unavoidable, engineers seek to go to extraordinary lengths to assure the reliability and integrity of the critical control system – as they do in the case of aircraft controls. Yet experience suggests that there will be cases where inadequate care is taken, leaving open vulnerabilities that cyber attackers might be able to exploit with devastating results. System engineering disciplines do exist to minimize the chances of this but they require high levels of expense and intrusive oversight and thus are unlikely to be uniformly applied. Moreover, there is a large overhang of legacy systems designed before the risks of cyber attack were recognized.

It can be very tempting for suppliers and customers to incorporate widely-available subsystems and software modules when constructing a system. Indeed there are many strong proponents of *commercial off the shelf* (COTS) or *open source* systems for virtually all uses. Clearly, such approaches can save substantial amounts of time and money, but they can also greatly increase vulnerability to common modes of attack, perhaps catastrophically so when they are applied to critical applications such as

EMS or SCADA systems. Thus their use should be subject to policies that assure management of risks and adequate weighing of risks and costs.[34]

There is also danger in policies of commonality – the use of the same systems or the same family of systems across an organization or industry. No doubt this can save cost, but it opens the potential for exploitation of a single vulnerability to affect a wide range of operations. Thus here too care should be taken to weigh savings against risk.

Concerns about their security emerged in the 1990s and have been heightened by recurrent hacker attacks. Even though these attacks have so far not gone beyond the nuisance stage, at least domestically, it is clear that when the grid is operating under stress a successful denial of service attack on a SCADA system or EMS could at least in theory lead to a situation comparable to that which occurred by accident and inattention in Aug 2003, in which operators lacked important information and/or could not exercise effective control. Thus lack of appropriate operator action could again set the stage for a massive failure cascade, this time as a result of cyberspace attack.

Worse still could be capture of a SCADA system or EMS by attackers who could use it to exert control over generating and transmission equipment. An attacker who had sufficient information and understanding of the affected portion of the grid and full enough access could use his control to increase stress on the system by configuring it inappropriately. If the attacker could simultaneously block or spoof system reporting and operator control functions he could render the manual safeguards ineffective or even counterproductive. This would be true whether the capture was effected by physical intrusion or by remote means via cyberspace.

There are some limits, however, to how much an attacker could accomplish simply by capturing control of an EMS and/or SCADA system. SCADA and energy management systems inherently lack the capability to override self-contained automatic safety relays associated with generators, transmission lines, and transformers. Design engineers, aware of the possibilities of SCADA and EMS failures for a variety of reasons, avoid giving them more control authority than is strictly necessary for their intended functions. Unless an attacker knew of and were able to exploit a very serious design flaw, capture of a SCADA system or EMS in itself generally would not in itself allow him to inflict major long term damage on the electrical system

---

[34] Howard F. Lipson, Nancy R. Mead, and Andrew P. Moore, "Can We Ever Build Survivable Systems from COTS Components?" CMU/SEI-2001-TN-030 (Pittsburgh: Carnegie Mellon University, Software Engineering Institute, Dec 2001).

per se, and would be unlikely to permit him to initiate a major failure cascade in the absence of heavy external system loading.[35] If combined with effective physical attacks on critical equipment, however, capture of the control systems could allow an attacker to greatly magnify the damage he could do. Even without coordinated physical attacks, however, a cyber attacker with sufficient access and knowledge could trigger a widespread blackout comparable, within the affected area, to the Northeast Blackout of August 2003. As the experience of that blackout shows, it can take several days to restore full service and there can be large economic losses, as well as some risk to life and property.

SCADA systems and EMS are cyberspace systems by the definition used in this book, but their susceptibility to attack by cyberspace means varies. Industry standards call for them to be isolated from contact with the public Internet and other sources of possible outside cyber entry.[36] (They are also specified to be protected from physical intrusion and from surreptitious insider takeover.) Many instances have been found, however, in which the systems have failed fully to meet these standards. In some cases deficiencies have been revealed through hacker attacks, but most have been discovered in the course of testing programs.

One potential threat that is often overlooked is that of Trojan horses introduced in the process of developing or maintaining the software. The concealed fault deliberately planted in a device or software program is a familiar fictional device,[37] but little has been done to forestall such threats, perhaps in part because there has been no publicity about actual cases. Although more complex to mount than a virus or denial-of-service attack, a surreptitious "fifth column" attack of this sort could potentially be more damaging and more difficult to diagnose and correct. The danger is greatest in the case of open source systems, where there is little control over who

---

35 It may be objected that attackers do routinely find and exploit important design flaws in the software on Internet-connected computers. As will be discussed below, EMS and SCADA systems are much less available for examination and probing, however. Moreover, the relative simplicity of SCADA systems in particular allows less opportunity for serious hidden flaws. Thus devastating attacks on these systems are much less likely. This is borne out by experience, as attacks on EMS and SCADA systems by hackers have thus far been much less common and generally less serious than those directed at Internet computers and servers.

36 Control Systems Security and Test Center, "A Comparison of Electrical Sector Cyber Security Standards and Guidelines," INEEL/EXT-04-02428, Revision 0 (Idaho National Engineering and Environmental Laboratory, 28 Oct 2004).

37 See, e.g., Richard A. Clarke, *Breakpoint* (New York: Putnam, 2007) for a recent example or Will O'Neil, *The Libyan Kill* (New York: W. W. Norton, 1980) for an earlier one.

may have modified the code. But COTS and even purpose-built systems may incorporate key modules supplied by obscure low-tier subcontractors with little oversight.

While an attacker who finds and exploits a key cyber vulnerability may be able to do severe and lasting damage to a particular system, many systems will be competently and conscientiously designed and operated and will not offer such opportunities. If an attacker targets a system for which he cannot identify a catastrophic cyber vulnerability then he will have to employ physical attack to do major damage to it.

In general it is necessary not only to do physical damage but to do enough of it to saturate the capacity for near-term restoration. Electrical utility companies, for instance, generally are very well prepared to quickly restore considerable numbers of downed transmission lines, since natural causes such as ice storms or hurricanes can do damage of this sort. If the attacker's goals involve putting the system out of operation for more than a few days then he will do better to attack other elements. The capacity to quickly replace large transformers is very limited, for instance, and that to replace major generator facilities is more so. An exception is a case where the attacker is able to interfere with repair and restoration activities and/or mount repeated attacks with little cost, as in Iraq.[38] In any event, the importance of physical security for key facilities is clear. In recognition of the threat posed by attacks on transformers and generators, efforts are being made by utilities and coordinating groups to improve capabilities for quickly restoring them, an example which should be widely followed.

Systems engineering and dependability

Engineers in many fields have long sought to make their systems perform their intended functions dependably in the face of a wide spectrum of threats. Over the course of this effort they have developed a body of practice, usually referred to as *systems engineering*, which encompasses specification, analysis, design, and testing practices to ensure that a system will meet definite standards of dependable operation. Although not always fully effective, experience has shown that thorough application of systems engineering practice greatly improves dependability.

Application of systems engineering has been notably weak in most areas of software development. The techniques for effective systems

[38] The experience with infrastructure attack in Iraq since the 2003 invasion is discussed in Appendix B to this chapter.

engineering for software are well understood and documented,[39] but the structure of the industry has not supported their application in most commercial software. This makes commercial software cheaper but undependable, as almost every reader of this report will know from personal experience. But most customers find it easier to assess and evaluate price than dependability.

Securing infrastructures against cyber attack is impossible without dependable software. Thus any program for infrastructure protection must mandate good software systems engineering in order to be effective.

Policy and organization

Existing top-level policy on infrastructure protection

Concerns regarding protection of infrastructure are of long standing but it was in the second Clinton Administration that the first steps toward a comprehensive policy were taken. A presidential commission was convened in 1996 and reported in 1997, emphasizing government-industry cooperative efforts.[40] On 22 May 1998 then-President Bill Clinton signed ***Presidential Decision Directive/NSC-63 (PDD 63)***, "Critical Infrastructure Protection." Although now superseded, PDD 63 was the root of most infrastructure protection policy.

As this is written, the principal current policy directives regarding infrastructure protection include the following:

- ***Executive Order on Critical Infrastructure Protection,*** signed by President George W. Bush on 16 October 2001. The primary focus of this EO is "continuous efforts to secure information systems for critical infrastructure, including emergency preparedness communications, and the physical assets that support such systems." It sets forth that, "The Secretary of Defense and the Director of Central Intelligence (DCI) shall have responsibility to oversee, develop, and ensure implementation of policies, principles, standards, and guidelines for the security of information systems that support the operations under their respective control. In consultation with the Assistant to the President for National Security Affairs and the affected

---

[39] Daniel Jackson, Martyn Thomas, and Lynette I. Millett, editors, *Software for Dependable Systems: Sufficient Evidence?* (Washington: National Academies Press, 2007).

[40] *Critical Foundations: Protecting America's Infrastructures: The Report of the President's Commission on Critical Infrastructure Protection*, Oct 1997, http://www.fas.org/sgp/library/pccip.pdf.

departments and agencies, the Secretary of Defense and the DCI shall develop policies, principles, standards, and guidelines for the security of national security information systems that support the operations of other executive branch departments and agencies with national security information." However the policy and oversight structure set up by this EO has been considerably modified since its promulgation.

- ***The Homeland Security Act of 2002***, signed into law by President George W. Bush on 25 November 2002, established the DHS (Department of Homeland Security) and assigned it lead responsibility for preventing terrorist attacks in the United States, reducing national vulnerability to terrorist attacks, and minimizing the damage and assisting in recovery from attacks that do occur. It gives the DHS broad responsibilities for protection of critical infrastructure in the United States both against terrorism and natural disaster. DHS, however, was not given responsibilities for protecting critical infrastructure from intrinsic or natural faults such as those involved in the Northeast Blackout of 14 Aug 2003, or from non-terrorist attacks.
- ***The National Strategy for the Physical Protection of Critical Infrastructures and Key Assets***, approved by President George W. Bush in February 2003. The focus is very specifically on protection against terrorist attack, rather than protection generally. It lays out a cooperative effort shared between various levels of government and the private sector without for the most part defining definite responsibilities.
- **The National Strategy to Secure Cyberspace**, approved by President George W. Bush in February 2003. By contrast to The National Strategy for the Physical Protection of Critical Infrastructures and Key Assets (NSPPCIKA) this does not focus so exclusively on terrorist threats, mentioning criminal threats and threats of military attacks as well. In overall structure and approach, however, this is comparable to the NSPPCIKA.

  The directive's stated purpose is "to engage and empower Americans to secure the portions of cyberspace that they own, operate, control, or with which they interact." "In general," it states, "the private sector is best equipped and structured to respond to an evolving cyber threat. There are specific instances, however, where federal government response is most appropriate and justified. ... Externally, a government role in cybersecurity is warranted in cases where high

transaction costs or legal barriers lead to significant coordination problems; cases in which governments operate in the absence of private sector forces; resolution of incentive problems that lead to under provisioning of critical shared resources; and raising awareness." The role of the DHS is most strongly emphasized and clearly detailed.

- ***Homeland Security Presidential Directive/HSPD-7***, Subject: "Critical Infrastructure Identification, Prioritization, and Protection," signed by President George W. Bush on 17 December 2003. "[E]stablishes a national policy for Federal departments and agencies to identify and prioritize United States critical infrastructure and key resources and to protect them from terrorist attacks." The DoD is assigned specific responsibility for protecting infrastructure relating to the defense industrial base. Most other national-level infrastructures, including many critical to DoD operations, are placed under the responsibility of the DHS so far as terrorist attacks are concerned.
- ***National Infrastructure Protection Plan 2006***, agreed among multiple agency heads (including then-Secretary of Defense Donald H. Rumsfeld) in June 2006. The NIPP defines its goals in terms of "enhancing protection of the Nation's CI/KR to prevent, deter, neutralize, or mitigate the effects of deliberate efforts by terrorists to destroy, incapacitate, or exploit them; and to strengthen national preparedness, timely response, and rapid recovery in the event of an attack, natural disaster, or other emergency." In military terms, the NIPP is analogous to a strategic plan, whereas the other directives more closely resemble broad statements of policy. Sector-specific plans are in the process of development and approval.

  The NIPP integrates terrorist and natural-disaster threats to the infrastructure and gives passing attention to criminal threats. Neither it nor the other directives address threats of warlike attack or intrinsic failure. It clearly would be best and most efficient to deal with them in an integrated and comprehensive way.

Aside from the now-superseded Clinton-Administration PDD 63, the focus of all these directives is very strongly on terrorist threats. There is some limited treatment of criminal and military threats, but virtually none of damage that might result from intrinsic, accidental, or natural causes.

This of course is not to say that the government has no policies with respect to other threats. For the most part, however, these are scattered among many laws, regulations, and directives relating to the responsibilities and functions of specific departments, agencies, and organizations. For that reason we will turn next to an organizational perspective.

Organizational responsibilities

Since the duties of the DHS are the subject of the policy documents just described our discussion will concentrate on other organizations.

The Department of Defense role

Regardless of other considerations, each federal department and agency must see to protecting its own infrastructures against all threats, in cooperation and coordination with other agencies as applicable. In addition, the DoD must defend the nation's infrastructures against military attack and participate with allies in defending their infrastructures against military attack. And it may be called upon to aid in protecting and restoring the nation's infrastructure or the infrastructures of allied or friendly nations against natural disasters. A comprehensive DoD policy regarding infrastructure defense and protection must deal with all these needs.

As a practical matter, however, the DoD's concerns cannot stop there. It is in the nature of infrastructure that it is pervasive, highly networked, and frequently largely invisible. Few clear boundaries can be drawn. All its great efforts toward self-sufficiency notwithstanding, the DoD is dependent on many infrastructures not under its control. Even though their defense against certain threats is not within its defined responsibilities, the DoD cannot afford to wash its hands of them.

Finally, the distinctions among threat sources – military, terrorist, criminal, natural, and intrinsic – are often not operationally meaningful in the sense that it can be difficult or impossible to discern the actual source of a threat, at least not in time to affect operations. For instance it may not be feasible or prudent to await definitive information about whether a specific problem is the result of military attack before taking defensive action.

Thus the DoD policy-maker confronts an uncomfortable conundrum regarding infrastructure protection. It is impossible for the DoD at once to stick solely to its own business and at the same time discharge its responsibilities. There will inevitably be very substantial and important ambiguities and overlaps of responsibility and spheres of action, and consequent potential for costly conflict with other agencies and entities.

There can be no bureaucratic "good fences" to make "good neighbors" with other agencies and organizations in infrastructure protection. Unless close and cooperative give-and-take relationships can be developed in advance, counter-productive friction is very likely to hamper needed efforts.

None of these issues or considerations are new to DoD, which has confronted them in various guises for many years. But the changing nature of the threats as well as new organizational responses in other areas of the government have led to significant changes. These are reflected in three key policy and doctrine documents:

- ***Department of Defense Strategy for Homeland Security and Civil Support***, approved by Deputy Secretary of Defense Gordon England in June 2005, is a broad statement of policy and approach.
- ***DoD Directive 3020.40 of 19 August 2005, Subject: Defense Critical Infrastructure Program (DCIP)***, approved by DEPSECDEF England, specifically defines DoD policy with respect to protection of defense-related critical infrastructure and assigns responsibilities.
- ***Homeland Defense, Joint Publication 3-27*** of 12 July 2007 is a joint doctrine published by the Joint Staff. It is especially lengthy, reflecting the complexity of the issues involved, and provides commanders at all levels with authoritative guidance covering a wide range of situations and contingencies.

Only experience will tell whether DoD's policy and doctrine will prove adequate and adequately implemented but what has been produced is encouraging.

Other federal agencies

Five major federal agencies share responsibilities relating to energy infrastructure: the Department of Homeland Security (DHS), Department of Energy (DoE), Department of Transportation (DoT), Federal Energy Regulatory Commission (FERC), and Nuclear Regulatory Commission (NRC). DoE and FERC are both involved in protection of all energy (electrical, oil, and natural gas) infrastructures against natural and intrinsic threats as well as sharing terrorism-protection responsibilities with DHS. NRC plays a comparable role in respect of nuclear energy infrastructure. DoT has responsibility for pipeline safety, exercised by its Office of Pipeline Safety, and coordinates with DHS regarding pipeline security.[41] The DoE has several national laboratories (outgrowths of the development of nuclear weapons)

---

41 Paul W. Parfomak, "Pipeline Safety and Security: Federal Programs."

and its Idaho and Sandia National Laboratories are active in energy infrastructure security research and development.

The Federal Communications Commission (FCC) has responsibility for all federal communications regulation. In the past it has commissioned a recurring series of Network Reliability and Interoperability Councils (NRICs), comprising a representatives of a broad spectrum of communications industry entities as well as concerned government organizations, chartered "to provide recommendations to the FCC and to the communications industry that, if implemented, shall under all reasonably foreseeable circumstances assure optimal reliability and interoperability of wireless, wireline, satellite, cable, and public data networks."[42] The Commission's Public Safety and Homeland Security Bureau works with DHS on security and protection issues.

National Communications System (NCS)

The NCS is an outgrowth of a Cold War initiative with origins in the 1960s, intended to assure critical Executive Branch communications under any circumstances, including nuclear attack. An interagency group long run by DoD, it is today lodged in the DHS.

Rather than build dedicated government-owned communications infrastructure for the purpose the NCS has stressed close cooperation with the telecommunications industry to assure the necessary reliability. The closely-related National Security Telecommunications Advisory Committee (NSTAC), provides industry-based advice to the Executive Branch on communications security issues.

North American Electric Reliability Organization (ERO) and North American Electric Reliability Corporation (NERC)

The first widespread power outage in North America was the Northeast blackout of Nov 1965, which affected large areas of Ontario, New York, New Jersey, and New England. One result was the formation of regional reliability councils and a top-level National Electric Reliability Council (NERC) to coordinate them. The regional councils and NERC operated under federal authority but were funded and staffed from the utility industry; their standards were consensual and compliance voluntary and self-policed. Experience showed that this was not adequate and that the same causes cropped up again and again in power failures. Industry leaders urged that NERC (which by then had become the North American Electric Reliability

---

42 http://www.nric.org/. The latest NRIC concluded its work in 2005.

Council, with coverage of Canada and a small portion of Mexico whose grid is linked to that of California) needed to be given teeth so it could formulate and enforce mandatory standards.[43] Finally, after the 14 August 2003 Northeast blackout, necessary legislation was passed.

Under the Energy Policy Act of 2005, FERC was given responsibility and authority for the reliability of the bulk electric power delivery system throughout the United States. It was authorized to designate an independent Electric Reliability Organization (ERO), which it was hoped would also be recognized by Canada and Mexico, to set and enforce standards throughout the North American electric grid.[44] After reorganizing itself to comply with the independence requirements and submitting a concrete proposal, NERC (by now North American Electric Reliability *Corporation*, subsidiary to the *Council*) was certified by FERC as the U.S. ERO in Jul 2006. It is empowered to establish and enforce reliability standards subject to FERC approval, with penalties for infraction. Among these are standards for security, including cyber security. Inevitably the standards reflect a balance among security and other considerations, notably cost. They have been reviewed by concerned government and industry organizations and are widely but not universally believed to be adequate. The standards were given FERC's imprimatur in Jan 2008, to take effect in March.[45]

The standards are process and objective oriented and broad enough to cover a range of situations. The ERO issues implementing instructions and assess compliance, recommending action to the FERC as necessary to correct problems that cannot be resolved administratively.

State agencies

Although they have embraced deregulation in various ways and degrees, the State governments retain their inherent powers to regulate infrastructures operating in their territory. Most States have one or more independent agencies devoted to these functions. In addition, of course, State and local law enforcement agencies play major roles in protecting infrastructure systems.

---

43 Secretary of Energy Advisory Board, Op. Cit.

44 However only a small portion Mexico's electrical infrastructure, in the extreme Northwestern part of the country, is currently integrated into the North American grid.

45 *Mandatory Reliability Standards for Critical Infrastructure Protection*, FERC Docket No. RM06-22-000; Order No. 706, issued 18 Jan 2008.

Information Sharing and Analysis Centers (ISACs)

In 1998 PDD 63 called for the establishment of an Information Sharing and Analysis Center as part of its apparatus for critical infrastructure protection (ISAC). This evolved into a series of 11 organizations: Communications ISAC, Electricity Sector ISAC (ES-ISAC), Emergency Management and Response ISAC, Financial Services ISAC, Highway ISAC, Information Technology ISAC (IT-ISAC), Multi-State ISAC, Public Transit ISAC, Surface Transportation ISAC, Supply Chain ISAC, and Water ISAC. Loosely coordinated by an overall Council, the ISACs serve as conduits for government-industry and industry-industry communication about operational threats and protective measures.

The Communications ISAC is the National Coordinating Center for Telecommunications (NCC), an arm of the NCS. The ES-ISAC is the NERC.

Critical Infrastructure Partnership Advisory Council (CIPAC), Government Coordinating Councils (GCC), and Sector Coordinating Councils (SCC)

The CIPAC was established Mar 2006 by DHS as a forum for confidential interchange among all parties concerned with critical infrastructure protection. It is divided into 15 sectors: Chemical, Commercial Facilities, Communications, Dams, Defense Industrial Base, Electricity, Emergency Services, Financial Services, Food and Agriculture, Information Technology, Nuclear, Oil and Natural Gas, Postal and Shipping, Transportation, and Water. For each sector there is a committee with representation from each major concerned governmental and industrial organization. The governmental organizations constitute the GCC for the sector and the industrial organizations comprise the SCC. For example the membership of the IT committee includes the National Association of State Chief Information Officers; the U.S. Departments of Commerce, Defense, Homeland Security, Justice, State, and Treasury; as well as Arxan; Business Software Alliance; Bearing Point; Bell Security Solutions Inc.; Center for Internet Security; Cisco Systems, Inc.; Computer and Communications Industry Association; Computer Associates International; Computer Sciences Corporation; Cyber Security Industry Alliance; Computing Technology Industry Association; EWA Information & Infrastructure Technologies, Inc.; Electronic Industries Alliance; Entrust, Inc.; General Atomics; General Dynamics; Hatha Systems; IBM Corporation; Information Systems Security Association (ISSA); Information Technology Association of America; Intel Corporation; Information Technology Information Sharing & Analysis Center

(IT-ISAC); International Systems Security Engineering Association (ISSEA); Internet Security Alliance; Internet Security Systems, Inc.; International Security Trust and Privacy Alliance (ISTPA); Juniper Networks; KPMG LLP; Lockheed Martin; McAfee, Inc.; Microsoft Corporation; NTT America; R & H Security Consulting LLC; Seagate Technology; System 1, Inc.; Symantec Corporation; TestPros, Inc.; U.S. Internet Service Provider Association; Unisys Corporation; and VeriSign.

Partnership for Critical Infrastructure Security (PCIS)

Representatives from the CIPAC SCCs comprise the PCIS, a cross-sector coordinating organization established in Dec 1999 under the auspices of the Department of Commerce and now taken up under the CIPAC.

Operating companies and organizations

U.S. infrastructures are operated by thousands of commercial and other organizations. Virtually every one of them takes measures to ensure reliability and security of operations.

Policy issues

Fragmentation of policy and organization

In the years since PDD 63 was issued in 1998 a great deal has been accomplished to make the nation's critical infrastructure systems more secure and robust and to improve their protection against cyber as well as physical attack, with emphasis on defense against terrorists. At the same time law enforcement agencies have greatly stepped up their activities in the area of cyber crime. Yet potential threats also have burgeoned over this period. It is a race in which to stand still is to fall seriously and swiftly behind.

Throughout this period cyber attacks have mounted steadily. Many have been directed at the infrastructure of cyberspace itself, and a smaller but still substantial number against other infrastructures via their SCADA and management systems. In the great majority of cases it has been impossible to determine the identity of the attackers or the motivations of their attacks. Vandalism, criminal gain, terrorism, intelligence gathering, or even covert military attack are all possibilities and usually there has been no way to tell.

These cyber attacks have been costly. Yet in terms of deaths, economic losses, and sheer misery and inconvenience their effects have been much

less than those stemming from other sources of infrastructure damage. Many more Americans have been much more seriously affected by loss of electrical, communications, transportation, natural gas, and oil service resulting from stressful weather, geological disaster, accident, and/or intrinsic faults in design or construction. As a natural and logical result our society invests more attention and capital in averting and containing these more common and costly problems.

In practice, it often is not clear whether damage was initiated by human or natural attack. We have seen in connection with the 14 Aug 2003 Northeast Blackout that it took months of investigation to determine that the cyber infrastructure failures that had an important bearing on the extent of the damage had not been caused by hostile attack. In fact, very similar damage might have been produced by a combination of physical attacks on transmission lines and cyber attacks on EMS and SCADA systems.

At the physical and engineering level there is thus a very large area of overlap in the measures needed to guard infrastructures against damage from whatever cause. To ignore this underlying unity in framing policy is to fight against nature and cannot fail to generate needless conflicts, gaps, or duplications of effort.

Yet our survey of the welter of policies and governing organizations reveals little evidence of unity in dealing with infrastructure protection.

A basis for unified policy

The most fundamental axiom of U.S. policy in every field is that to the greatest extent possible responsibilities should be assigned to individuals or small, unified groups and that responsibility and authority should always be closely aligned. That is the basis for our free enterprise system, for restricting the powers of government as narrowly as possible, and for assigning governmental powers to the lowest and most local level possible.

We are also very wary, as a society, of the hazards of mixed motives and conflicting interests. We know all too well how difficult it is to serve two different masters or pursue two divergent interests.

These principles have informed America's decisions about infrastructures. Unlike many countries, we have never made the telecommunications, rail, or petroleum infrastructures into government departments. The Internet, created at federal government initiative, was divested as soon as it seemed feasible. Governmental control and operation

of other infrastructures is quite limited and largely confined to local authorities.[46]

Yet we have seen how the private enterprise structure in electrical power distribution has itself contributed to conflicting interests and mismatches of responsibility and authority, leading to massive artificial shortages and huge blackouts. The companies involved found that they could increase profit potential by withholding electricity (in the case of the California energy crisis of 2000-2001) or neglecting safeguards (as in the case of the 14 August 2003 blackout). They understood that these actions were undesirable from the standpoint of American society as a whole but it is our society, after all, which mandates such powerful incentives to individual and company profit. This was foretold, but no effective preventive measures were established. If we wish different outcomes we must either restructure the marketplace to assure that profit motives align with society's needs or else impose effective regulation to prevent companies from finding profit in damaging or dangerous actions.

Market solutions

Aligning profit motives with needs for infrastructure protection against attack would be most desirable, providing maximum delegation of power and responsibility while minimizing conflict of motive and interest. The most direct approach to this is to make companies bear the costs that result from successful attacks on their facilities and services. This in principle would motivate them to do what is needed to avoid or mitigate damaging attacks, including banding together as necessary to take collective action. This would be the pure free-market solution and the one that is arguably best aligned with American principles and values.

Scarcely anyone in our society questions the efficacy of the free market in providing well-defined products and services to meet customer demands at the lowest price. But in a case such as this experience and theory combine to raise a series of issues to be considered, including:

- Because the incentives would take the form of threat of need to repay money already collected from its customers, they would depend on the credibility of some external enforcement mechanism, inevitably governmental in character. The government would not only have to be prepared to act in a rigidly punitive fashion but to convince the

[46] The Tennessee Valley Authority is the most prominent exception.

companies that it was so. It has always been difficult for democratic government to do this.

- The companies most concerned would be limited-liability corporations and their inherent limitations of liability would imply a cutoff in threat response. That is, any threat severe enough to threaten the viability of the company should evoke the same level of protection regardless of whether its effect on society was catastrophic or only serious. Thus society might be relatively under-protected against the gravest threats.
- Corporations are run by agents, executives whose own incentives may be more or less misaligned with those of the corporation's owners, its shareholders. Alignment of executive and owner interests is essential to any free-market solution to the infrastructure protection problem. But this alignment is difficult to achieve when income and cost are separated in time and the magnitude of cost is uncertain. This is a case where cost may come a long way down the road and where its amount (and even incidence) are wildly uncertain. In such circumstances it is extremely tempting for executives to focus very strongly on present-day income and neglect the highly uncertain future costs of infrastructure attack.
- Many of the most effective potential responses to threats of attack involve foreign intelligence collection or the exercise of police powers or military force. While some might perhaps welcome broad delegation of such powers to individuals or corporations, to do so would raise issues regarding the nature of our nation and government far transcending the bounds of this discussion.

Regulatory solutions

Unless some way can be found to avoid the problems of a market-based approach, security for our infrastructures will have to depend either on direct government control or on regulation. Direct government control of course would fly directly in the face of the fundamental principles of delegation of responsibility and control and of alignment of motivations discussed earlier, and we will not address it further.

In principle the most desirable way to regulate infrastructure security might very well be by *private orderings*, in which industry participants spontaneously and out of self interest and/or their sense of social responsibility evolved structures which society could rely upon. This could

minimize the costs of regulation.[47] Needless to say, Internet governance is a prime example of private ordering in action. But there seem to have been no serious efforts to date to develop proposals for private orderings for infrastructure security. For the moment it must remain an open question and a challenge to policy analysts. In the absence of serious suggestions regarding private orderings we turn to *public orderings.*

In broad principle all regulation operates by manipulation of the incentives of income and costs. There is a great difference in practice, however, between regulation which threatens to cost executives their freedom and that which merely promises to modify the firm's profit and loss calculus. Here we will distinguish between *incentive regulation* and *directive regulation.*

A well-known example of incentive regulation is *pollution credits* (also discussed under the rubrics of *emissions trading* or *cap and trade*). The regulator creates a certain number of credits, each conferring the right to emit a defined quantity of pollutants – for instance, 10 million tons of carbon dioxide per year. The credits are allocated to firms by administrative fiat or auction and thereafter firms are free either to keep a credit and emit that quantity of pollutant or to sell it to another firm. The price of a credit acts as an incentive to the firm to invest in pollution reduction so it can sell the credit. The net effect ideally is to concentrate investment in pollution reduction in areas where the greatest cuts can be achieved at least cost to society as a whole.

There are pitfalls for regulators in such schemes and they do not always work well. As outlined earlier the incentive regulation regime employed to regulate electricity and natural gas distribution in California in the early 2000s offered opportunities for gaming which were exploited by Enron and other suppliers to gain billions of dollars in extra profits. There is wide (although by no means universal) agreement, however, that where they can be appropriately designed and well implemented, incentives provide the most efficient means of regulation.

But there are many areas of regulation where incentive regimes have not been found feasible or attractive, at least not so far. For instance, issuance of credits permitting a firm to cause a certain number of deaths or maimings would not be widely accepted by the public as a substitute for

---

[47] Steven L. Schwarcz, "Private Ordering," *Northwestern University Law Review* 97, No. 1 (Jan 2002): 319-49.

affirmative direct regulation of safety measures, regardless of the theoretical merits of such a scheme.

Regulation affecting vulnerability to infrastructure attack seems open to similar objections against dependence on incentives. Beyond this, however, attacks themselves are infrequent and variable enough in nature and intensity to raise severe problems in measuring vulnerability. It is a very different situation from that of carbon dioxide emissions or even workplace accidents, where it is possible to gather relatively immediate and direct data on the impact of any control measures.[48]

Thus it appears that in many areas effective protection of infrastructures against attack can best – and perhaps only – be assured through directive regulation of infrastructure firms. To the greatest possible extent, this regulation should take the form of performance-oriented requirements which leave to the individual firm the choice of means by which the necessary performance is to be achieved. It is generally found in safety-related regulation, however, that there are areas in which the regulators have effectively no choice but to mandate the use or avoidance of specified procedures and equipment. The dangers of this sort of regulation are clear and significant, since regulators are given power to impose increased costs without having to answer to the firm's owners, whose only recourse is through administrative, legal, or political appeals. The regulators present the same agency problems as management while being possibly even less accountable to owners. The only mitigation is that, unlike managers, regulators are not able to profit personally by actions which may damage the firm.

In any event, good practice in process- and equipment-oriented regulation always dictates that firms should be given the opportunity to propose alternatives based on an evidentiary case that what they propose will produce results at least as satisfactory as those mandated in the regulation. It is also important that regulators be very attentive to industry arguments regarding changes in technology or circumstance and the resulting needs for regulatory revision.

One of the important variant of public ordering is *public ordering with private enforcement*, in which the rules are publicly determined but are enforced in part or in whole by private appeal to the courts or administrative

---

[48] It is in some way similar to the problem of regulating hedge funds and like entities, which can have very infrequent but extremely costly failures. See Dean P. Foster and H. Peyton Young, "The Hedge Fund Game," Brookings Institution Center on Social and Economic Dynamics Working Paper No. 53, November 14, 2007.

tribunals. In principle this offers opportunities to reduce costs and reduce the opportunities for costly bureaucratic meddling. A variant of this is now being employed in regulating electrical system reliability, and it merits attention.

We saw earlier that the quest to better secure the electrical power infrastructure against natural and intrinsic threats, as well as against many forms of attack, has led recently to the establishment of a formal Electric Reliability Organization (ERO). The ERO takes the form of a private non-profit corporation with close ties to the electric power industry, but endowed with regulatory powers under the supervision and control of a federal agency, FERC. Earlier experience with NERC (which now runs the ERO) had demonstrated that admonition and appeals to industry-wide and national interest were not adequate, but it is expected that the ERO will not have to regulate with a heavy hand and will not be a source of significant needless cost for the industry and its customers. On the other hand there is reasonable confidence that it will be able to bring important improvements in electric grid reliability (bearing in mind factors, such as the narrow margins between capacity and maximum demand, that lie beyond its control).

This system of regulation is only now starting to operate and we cannot be certain how well it will fulfill expectations. Even if it operates exactly as hoped it will not eliminate blackouts, for that is not possible with an electric grid anything like the one we have. What it should do is both to reduce their frequency and greatly reduce their severity. If it is successful in establishing and enforcing appropriate standards, based in present knowledge, then a blackout like that of 14 August 2003 should never happen again.

It is not possible to be quite as definite about the potential of the ERO to protect the grid from attack, simply because our experience of attack is not as comprehensive as that of natural and intrinsic casualties. But analysis and experience does indicate that consistent enforcement of ERO standards should make the risk of damage from an attack significantly lower.

Overall, it appears that the ERO model offers promise as a mechanism for regulation to improve the survivability and operability of many kinds of infrastructures in the face of attacks as well as natural and intrinsic threats.

Cyberspace infrastructure

Many U.S. infrastructures are shared to some degree with our neighbors in Canada and Mexico, and actions to protect them need to be coordinated closely with those taken by the governments of these nations. The cyberspace infrastructure, however, is unique in the extent of its international connections and dependencies.

This raises unique problems of governance and regulation. In electrical power we can and do operate with standards that differ considerably from those used in distant countries – even in such basic matters as AC distribution frequency (60 Hz here but 50 Hz in many other places). In cyberspace, however, international coordination issues are much more complex, as outlined by Harold Kwalwasser in Chapter 9. Many aspects of cyberspace infrastructure protection policy must therefore be coordinated with foreign and international bodies.

How much is enough?

The title of this section of course refers to what is always the most fundamental of questions in defense planning: what level of protection is needed? There is no absolute answer to such a question, but it needs to be addressed explicitly and systematically.

We are frequently warned of threats to cyberspace infrastructure and to infrastructures generally. But how are we to weigh these threats against others, and to assess how much of our attention and resources we should devote to countering them? Our very limited experience of such threats makes the problem much more difficult.

Examples drawn from other risk fields help to illustrate the issues:

- In 2001 about 5,000 people died because more than a quarter of car and truck occupants failed to wear seatbelts.[49] In the same year, accidental drownings killed 3,300.[50] Both tolls exceeded that caused by terrorist attacks, but public concern about terrorist attacks is far higher than that over seatbelt use or water hazards, and far more resources are being devoted to combating terrorism.
- The average annual toll from asteroid impacts is estimated to be lower than that from machinery accidents, but the two averages are arrived at in very different ways. Machinery accidents occur frequently and

---

49 Based on data in "National Transportation Statistics."

50 Centers for Disease Control and Prevention, "Web-based Injury Statistics Query and Reporting System (WISQARS)," http://www.cdc.gov/ncipc/wisqars/.

> relatively regularly, each involving a small number of deaths, while fatal asteroid impacts come at intervals of 1,000 years (for relatively small incidents) to 100 million years or more (for catastrophic ones) and could involve huge numbers of fatalities – possibly even extinction of our species. Until rather recently evidence regarding asteroid threats was generally discounted, but over the past two decades has become a matter of some public concern.[51]

Surveys and experimental studies in how we evaluate and respond to perceived risks confirm, as these examples suggest, that people are not rigorously "logical" about such matters. The perceived "dreadfulness" of a threat has a lot to do with response to it, as does the form in which information regarding its probability of occurrence is received.[52] We are prone to be less concerned, relative to objective quantitative risk level, about common and familiar risks such as heart disease or motor vehicle accident than shadowy and little-understood menaces such as cyber attack.

A further complicating factor is the concern decision-makers often feel regarding public reactions to attacks or failures. Fear of mass panic in the face of danger is one prevalent concern, as expressed with special vividness by the famous General William "Billy" Mitchell:

> What would we do if the United States were attacked and New York menaced? ... A deafening roar—another and another.... There is another blast—and the rush to the streets begins.... The streets are tightly filled before a third of the office workers have poured out. Tardy ones claw and clutch and scramble, clambering on top of those who have fallen. Before long there is a yelling, bloody, fighting mass of humanity.[53]

Less immediately, decision-makers fear weakening of public support for necessary measures in the face of sacrifices. For instance virtually every wartime president since Abraham Lincoln has worried that the people would be unwilling to accept casualties as the price of victory.

Social scientists find, however, that support for a conflict is not a question so much of a particular level of casualties as of belief in the cause

---

51 Clark R. Chapman, "The Hazard of Near-Earth Asteroid Impacts on Earth," *Earth and Planetary Science Letters* 222, No. 1 (May 2004): 1-15.

52 M. Granger Morgan, "Risk Analysis and Management," *Scientific American* 269, No. 1 (Jul 1993): 32-41; Paul Slovic, et al., "Risk As Analysis and Risk As Feelings," *Risk Analysis* 24, No. 2 (Apr 2004): 311-22.

53 William Mitchell, "When the Air Raiders come," *Collier's* (1 May 1926): 8-9 and 35, p. 8.

for which it is fought and the probability of success.[54] Similarly, it is found that mass panic is very rare even in circumstances which might seem to provide ample justification.[55] It is well established that for most people emotional factors play the dominant role in determining overall response to issues like demand for defense, but it is insufficiently appreciated that this by no means implies that the public is "irrational" about such subjects.[56] Our emotional apparatus evolved as it has because it aided survival in very threatening environments, and it continues to serve us in this role.[57] Decision-makers often resort to measures intended to manipulate the public's emotional responses in order to gain support. But while manipulation can seem to be effective in the short term, over the longer term it often evokes a backlash in policy matters just as it does in personal relationships.

Ultimately the question *how much is enough* of infrastructure protection can be answered only by the public through the political process. Policy-makers hoping for a sound answer will do well to provide the public with clear, credible information.

Policy recommendations

The foregoing examination of infrastructure protection issues has revealed a lack of broad and systematic policy. The following recommendations to remedy this are presented for consideration at the highest levels of government.

1. **Unify policy direction.** It is unrealistic to expect that all of the aspects of policy relating to infrastructure protection can or should be united under a single governmental department or agency, but it is essential that a positive mechanism be put in place to assure effective inter-agency coordination. Because unitary action is required to protect each infrastructure against natural disasters, accidental and intrinsic failures,

---

54 Christopher Gelpi, Peter D. Feaver, and Jason Reifler, "Success Matters: Casualty Sensitivity and the War in Iraq," *International Security* 30, No. 3 (Winter 2005/06): 47-86, which provides a guide to earlier literature. See also Louis J. Klarevas, Christopher Gelpi and Jason Reifler, "Casualties, Polls, and the Iraq War," *International Security* 31, No. 2 (Fall 2006): 186–198 for a critique and response. For a survey of data on support for wars between 1942 and 1993 see Eric V. Larson, *Casualties and Consensus: The Historical Role of Casualties in Domestic Support for U.S. Military Operations*, MR-726-RC (Santa Monica, Calif.: RAND, 1996), pp. 105-20.

55 Lee Clarke, "Panic: Myth or Reality?" *Contexts* 1, No. 3 (Fall 2002): 21-6.

56 Paul Slovic, et al., Op. Cit.

57 Antonio R. Damasio, *The Feeling of What Happens: Body and Emotion in the Making of Consciousness* (New York: Harcourt, Brace & Company, 1999).

and threats from terrorist, military, and criminal attack it is necessary that the inter-agency mechanism encompass them all. The precise organization of this mechanism requires further study by Congress and the Executive.

2. **Specialize policy direction.** While there should be unity in overall direction the various infrastructures should be subjected to policy direction tailored to their specific nature and needs. Thus for each infrastructure there should be a subordinate inter-agency process involving those agencies with specialized knowledge and responsibility.
3. **Strengthen and unify regulation.** While directive regulation of infrastructure firms at the process level has important pitfalls, there is no evident substitute for it with regard to protection of infrastructures. Absence of effective regulation leaves firms exposed to commercial pressures that work against protection and tends to prompt a "race to the bottom." For each infrastructure there should be a single, well-informed regulator with the knowledge and incentives to strike the right balance between risk and economic benefit. The ERO represents a promising approach to this which should be studied more carefully as a potential model for other infrastructures.
4. **Define State and local roles.** State primacy in policy and regulation for infrastructures has been undercut by the trend toward larger, interstate networks but State and local government agencies nevertheless retain a very important role. The federal inter-agency mechanism for infrastructure protection policy and related regulatory apparatus must be linked closely with the relevant State agencies. How this is to be accomplished will need to be worked out directly with the States.
5. **Define international interfaces.** Cyberspace infrastructure networks depend on international connections, but in this they differ in degree rather than kind from other infrastructures. In virtually all cases it is necessary to secure coordinated international action in order secure infrastructures most effectively. Again, the ERO appears to offer a promising model, with the United States playing a positive leadership role by offering a structure with mutual benefit and demonstrating readiness to modify positions to meet the legitimate interests of others.
6. **Mandate effective systems engineering for infrastructure-related software.** Undependable software is one of the greatest vulnerabilities of infrastructure systems. The cost-driven trend to wide use of undependable COTS and open-source software is exacerbating the risks.

Software dependability will not achieve the necessary standards unless effective systems engineering is mandated for infrastructure systems.

7. **Don't take *no* for an answer.** There will be some in the infrastructure industries who will express strong resistance to any directive regulation, regardless of justification. Their objections are understandable but must not be accepted. It is instructive to look at the forty-year struggle to avert massive electrical blackouts without directive regulation – a struggle culminating in the 14 August 2003 Northeast Blackout whose magnitude was multiplied by widespread failure to comply with existing voluntary standards. Any decision-maker who is tempted to give into industry pressures against regulation should consider carefully what he is going to say when to have done so is found to have opened the way for a successful and damaging attack.
8. **Establish and realize clear priorities.** There is no clear limit to potential threats against infrastructures, but there are limits to resources which can be used for protection. An attempt to protect everything against all possible threats will result in inadequate protection for the most crucial targets against the most important threats. Priorities for the allocation of financial and management resources are essential in order to provide effective protection.
9. **Inform the public clearly and accurately.** Many of the decisions which will have to be reached regarding protection of infrastructures will be technical and should be made by those with appropriate expertise. It is a serious error, however, to imagine that the key decisions in this area can be held within any closed group. The integrity of infrastructures affects everyone in our society and the public will demand that its views be heeded at critical junctures. A systematic ongoing effort to make full and objective information available is the best guarantee of informed and considered public input. It also is the best way to ensure that the public will feel confidence in those who direct infrastructure protection efforts and will pay appropriate attention to their advice and recommendations.
10. **Conduct a continuing program of research.** Many important questions remain unsettled and more will arise as threats, technology, and economic conditions change. The policy and regulation institutions must have the authority, resources, and responsibility to sponsor and guide broadly conceived programs of research to serve their information needs. Knowledge can be expensive but its absence can be much more so.

Appendix A

Network Theory

Viewed abstractly, infrastructures largely follow a common pattern despite differences in specific characteristics. Generally they involve a product which starts at certain points or nodes of origin and is transmitted or moved along certain linking routes to certain nodes of destination. Along the way it may pass through or be forwarded via multiple nodes of origin, and nodes may function both for origin and destination. The whole set of nodes of origin and destination, together with the linking routes, comprises a *network.*[58]

Not only infrastructures but a very wide variety of man-made and natural physical, biological, and social systems can be analyzed in network terms. Network theory is an active research field and a number of important discoveries have been made in recent years.[59] We will touch on a few points that are important for infrastructure protection.

While infrastructure networks are not truly random they are very complex and irregular; as a result many of the applicable tools of network theory are statistical in nature. The most fundamental statistic describing a complex network is p_k, the proportion of the network's nodes having exactly k links connecting to other nodes.[60] A node having k links is said to have

---

[58] The terms *link*, *edge*, and *line* are used interchangeably in discussing networks. Similarly, *node*, *vertex*, and *point* all refer to the same thing. Networks are also referred to as *graphs.*

[59] Albert-László Barabási, *Linked: How Everything is Connected to Everything Else and What it Means for Business, Science, and Everyday Life* (New York: Plume Book, 2003) is an excellent non-technical overview by a leader in modern network theory. For a more technical summary, Réka Albert and Albert-László Barabási, "Statistical Mechanics of Complex Networks," *Reviews of Modern Physics* 74 (Jan 2002): 47-97. Duncan J. Watts, *Six Degrees: The Science of a Connected Age* (New York: W. W. Norton & Co., 2003) is another sound non-technical overview by a prominent scientist but oriented more toward social networks rather than infrastructures.

[60] Naturally, k has to be a whole number – a node cannot have 2½ or 4.38 links.

degree k and the table or set of values of p_k for all values of k is called the *degree distribution* of the network.

If a network's nodes are connected entirely by chance the result is a random but statistically uniform network. If there are N nodes and K links all told then most nodes will have about $k_{Av} = 2N/K$ connecting links.[61] A purely random network such as this is "egalitarian" – nodes with very high numbers of links will be extremely rare. More specifically, for k much larger than k_{Av}, the proportion of nodes with k links will decline *exponentially* (i.e., as e^{-k}, where e is a mathematical constant, an irrational number approximately equal to 2.71828) as k grows. Thus networks of this kind can be called *exponential networks*.

While all purely uniform random networks are exponential, a network does not have to have arisen randomly to be exponential – it simply has to look generally like a random network. At one time it was thought that any large irregular network must be exponential in this sense, and until recently infrastructure networks were sometimes analyzed in this way. But in the 1990s closer examination of some infrastructures began to erode this assumption. It was discovered that many infrastructure networks (and a great many other kind of physical, biological and even social networks) follow an equally simple but very different distribution, $p_k \approx k^{-\gamma}$, where γ (Greek letter gamma) is a constant which varies from network to network but in a great many cases lies between 2.1 and 2.5. This distribution is called a *power law* and so networks which follow it can be called *power law networks*. For large values of k the exponential e^{-k} is a great deal smaller than $k^{-\gamma}$, meaning that highly connected nodes (nodes with large values of k) will be far more common in a power law network than in an exponential network.

A power law distribution has no hump or peak near the average value, k_{Av}. Instead, the proportion of nodes with exactly k connections in a power law network is maximum for $k = 1$ and declines monotonically thereafter. Since there is no one dominant value of k, or *scale* that characterizes the network such networks often are called *scale-free.*

These two kinds of nets are illustrated in Figure 1. It is obvious immediately that the diagram on the right, **b**, showing a scale-free power-law network, has a number of highly-connected nodes as well as a great many nodes with only a single link connection. The random-like exponential

---

61 The factor of two in the formula for the average number of links per node accounts for the fact that each link connects to two nodes.

network on the left, **a**, has a much more uniform pattern of connection, contrasting sharply with **b**. The scale-free network, **b**, gives a visual impression of interlinked wheels and the highly-connected nodes in such networks often are referred to as *hubs*.

If an accident or attack were to disable a node picked at random from the exponential network at the left of the figure, it usually would disconnect only a handful of other nodes that happen to connect uniquely to the one disabled. In the scale-free network a random disablement will do even less damage in most cases since so few nodes have any other node which connects only through them. But the worst case is worse in **b** than in **a** – taking out only a few of **b**'s highly connected hubs does a lot of damage.

Appendix B

Infrastructure Attacks in Iraq

The drawn-out conflict in Iraq following the Mar 2003 American invasion has provided a laboratory for infrastructure attack, with the insurgents targeting oil and electricity in particular. There have been repeated physical attacks on oil production facilities and especially pipelines.[62] Electric grid attacks have been aimed at high voltage transmission lines. In both cases many personnel have been killed. There are no reports of cyber attacks, but neither the oil nor electrical system have much in the way of SCADA or operational management systems that could make appealing cyber targets.

The identity of the attackers is shadowy and their strategy, goals and incentives are unclear. Their motives may be profit as much as politics. They have not destroyed the oil or electrical systems but it is not clear whether that is their intention. They certainly have imposed major problems and costs. The lack of adequate and reliable electrical power has been a factor in disaffecting the population and undercutting support for U.S. objectives. Loss of oil revenues has significantly weakened the Iraqi government and forced to U.S. to subsidize it. Substantial forces have had to be devoted to protection of infrastructure. And the limitations of domestic petroleum product production and of electrical production have forced large-scale trucking of fuels from Iran, Kuwait, and Turkey, adding substantial effort to protect the convoys.

The problems of fuels logistics have been greatly exacerbated by an ill-considered American decision at an early stage to boost Iraqi electrical generating capacity with combustion turbines – turbogenerators driven the exhaust from aircraft-type jet engines. They were ill-suited to Iraqi needs and conditions and required fuel which has to be trucked in because it is

---

62 Iraq Pipeline Watch, http://www.iags.org/iraqpipelinewatch.htm.

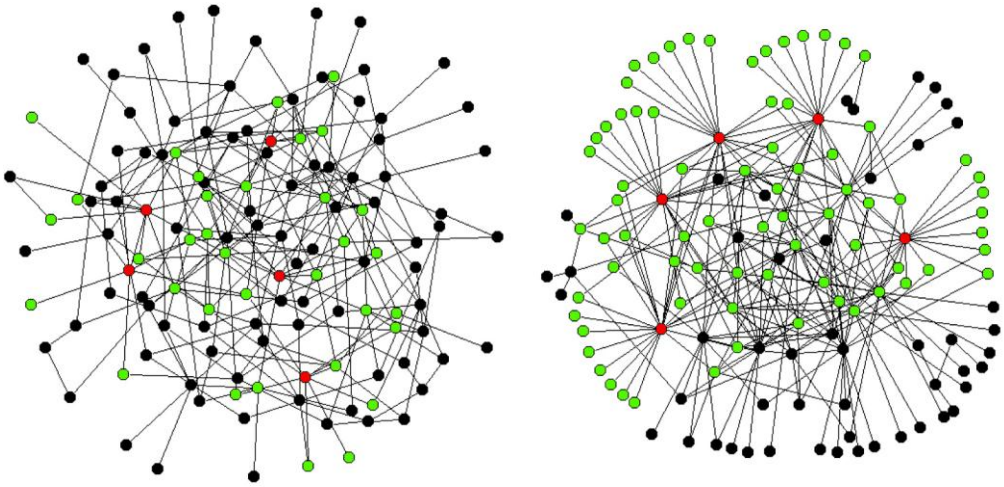
not available in Iraq.[63] More careful examination of real needs on a total-system basis would have paid significant dividends. This is a lesson which should be learned for the future.

63 Glenn Zorpette, "Re-engineering Iraq," *IEEE Spectrum* 43, No. 2 (Feb 2006) 22-35.

Glossary

Term	Meaning
AC	Alternating current
CI	Critical infrastructure
CIP	Critical infrastructure protection
CIPAC	Critical Infrastructure Partnership Advisory Council
COTS	Commercial off the shelf
DC	Direct current
DHS	Department of Homeland Security
EMS	Energy management system
EO	Executive Order
ERO	Electric Reliability Organization
FERC	Federal Energy Regulatory Commission
GCC	Government Coordinating Council
HSPD-7	Homeland Security Presidential Directive/HSPD-7, "Critical Infrastructure Identification, Prioritization, and Protection," 17 Dec 2003
HTS	High-temperature superconductor
HVAC	High-voltage AC
HVDC	High-voltage DC
Hz	Herz (frequency unit)
ISP	Internet service provider
KR	Key resources
NERC	North American Electric Reliability Corporation (formerly North American Electric Reliability

	Council, formerly National Electric Reliability Council)
NIPP	National Infrastructure Protection Plan
PCIS	Partnership for Critical Infrastructure Security
NSPPCIKA	The National Strategy for the Physical Protection of Critical Infrastructures and Key Assets, Feb 2003
PDD 63	Presidential Decision Directive/NSC-63, "Critical Infrastructure Protection," 22 May 1998
SCADA	Supervisory control and data acquisition
SCC	Sector Coordinating Council



a. Random-like, exponential network

b. Power law, scale-free network

Figure 1. These two diagrams show networks with identical total numbers of nodes (N = 130) and links (K = 215) but with different structures corresponding to their different degree distributions. In each diagram the five most-connected nodes are colored red and the nodes that connect directly to them are green. In the power-law network, ***b****, the most-connected node is connected to 17% of all the nodes in the network while in the exponential network,* ***a****, no node connects to more than 7% of the network. But if the rich are richer in connections in* ***b****, the poor are more numerous – only 12% of the nodes in* ***a*** *are singly-linked while more than 40% of the nodes in* ***b*** *have only one link. The great majority of nodes in a have degree close to the average number of links for the network, 2×215÷130 = 3.3 links per node.*

We need to be observant in interpreting diagrams such as these, which represent only the abstract topological relationships among the elements of the network and not any actual physical structure. Any rearrangement of nodes and links in the diagram would still represent the same network.

Source: University of Notre Dame Physics Department.

Level	Description	Examples
Cyber	Intellectual content	data, commands, knowledge, ideas, mental models
Logical net	Services employing physical signals to carry logical messages	"plain old telephone service" (POTS), broadcast radio and TV services, cable TV service, public Internet, private IP-based networks carried on common-carrier infrastructure, private-infrastructure IP-based networks, SCADA networks, etc.
Hard net	Infrastructures formed from base elements that carry electrical or electromagnetic signals	common-carrier telecommunications networks, tactical radio systems, dedicated wireline systems, community cable systems, cell phone systems, etc.
Base	Physical elements that underlie telecommunications services	cable headworks, optical fiber, coaxial cable, radio transmitters and receivers, radio transmission paths, communications satellites, Internet routers, modems, etc.

Table 1. *Simplified schematic overview of the levels involved in cyberspace networks.*

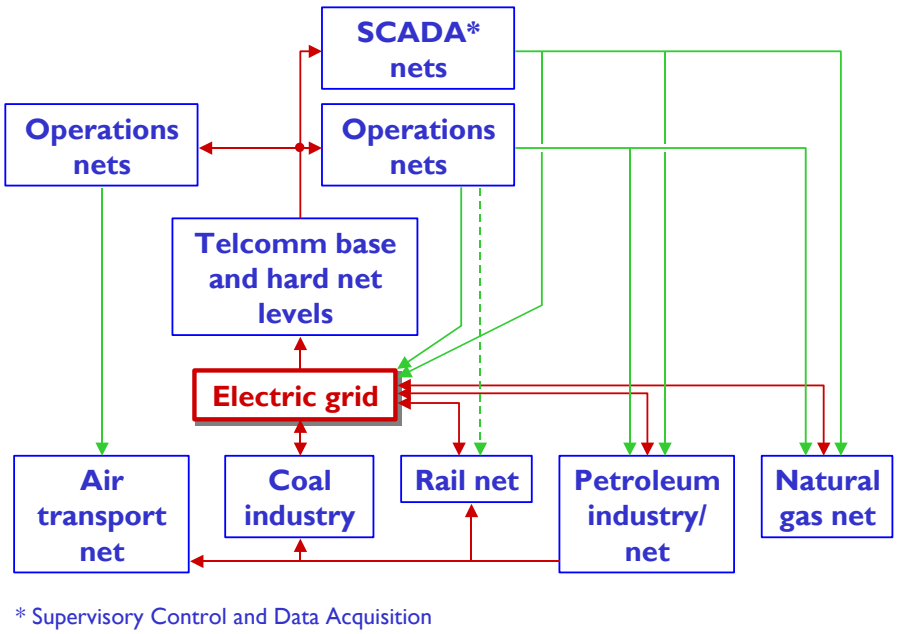


Figure 2. *Top-level interdependencies among selected infrastructures, emphasizing the central role of the electric grid. Few infrastructure systems can function for long without electric power. (Red links denote physical inputs, green, control and data inputs.)*

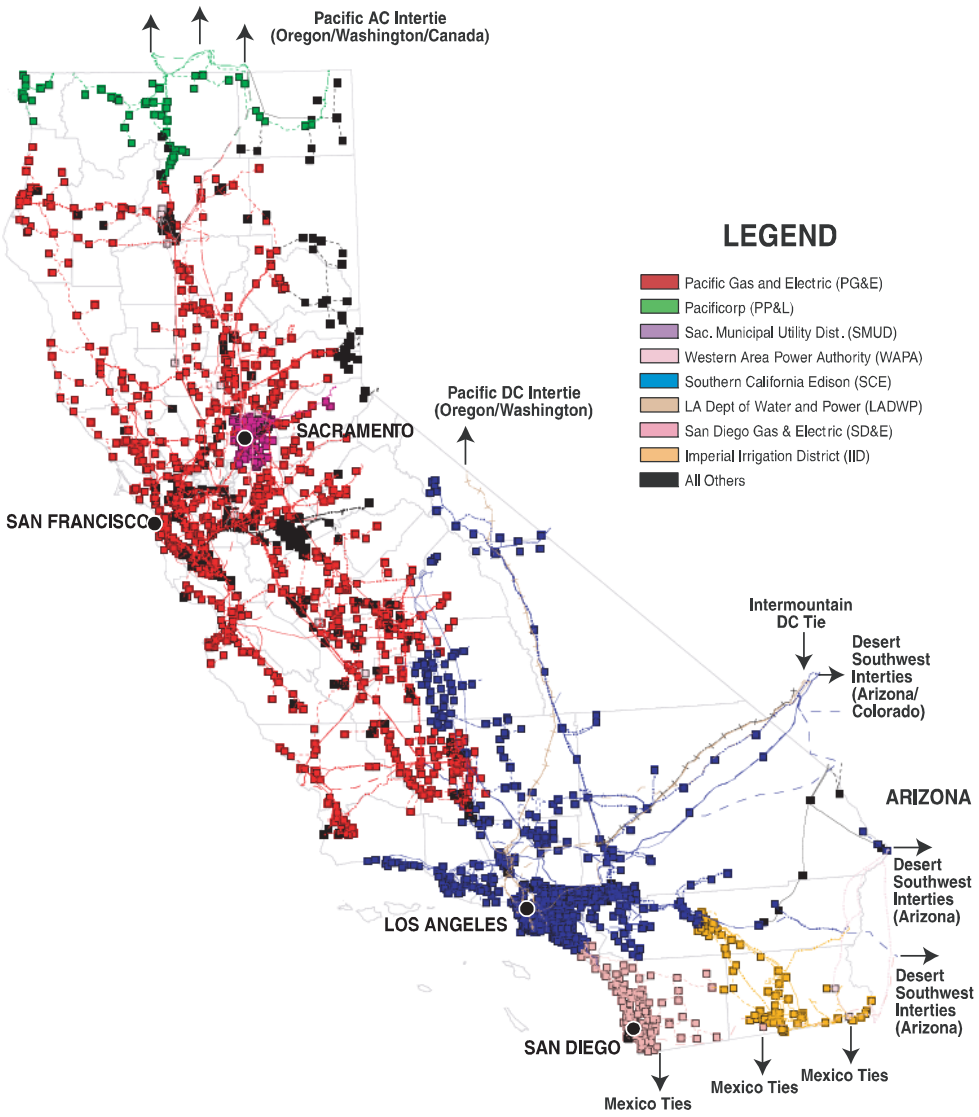


***Figure 3**. California's high-voltage bulk electric transmission grid. The nodes (boxes) are generation and distribution stations. The low-voltage local distribution networks are not shown, but in most cases distribution stations feed local networks.*

Source: California Energy Commission

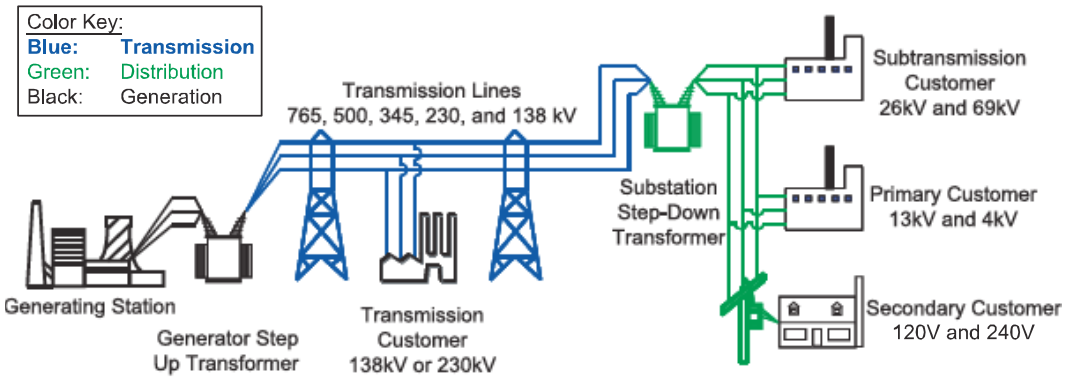


Figure 4. *Major components of the electric grid.*

Source: U.S.-Canada Power System Outage Task Force, "Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations," (Washington and Ottawa: U.S. Department of Energy and Natural Resources Canada, April 2004), p. 5.

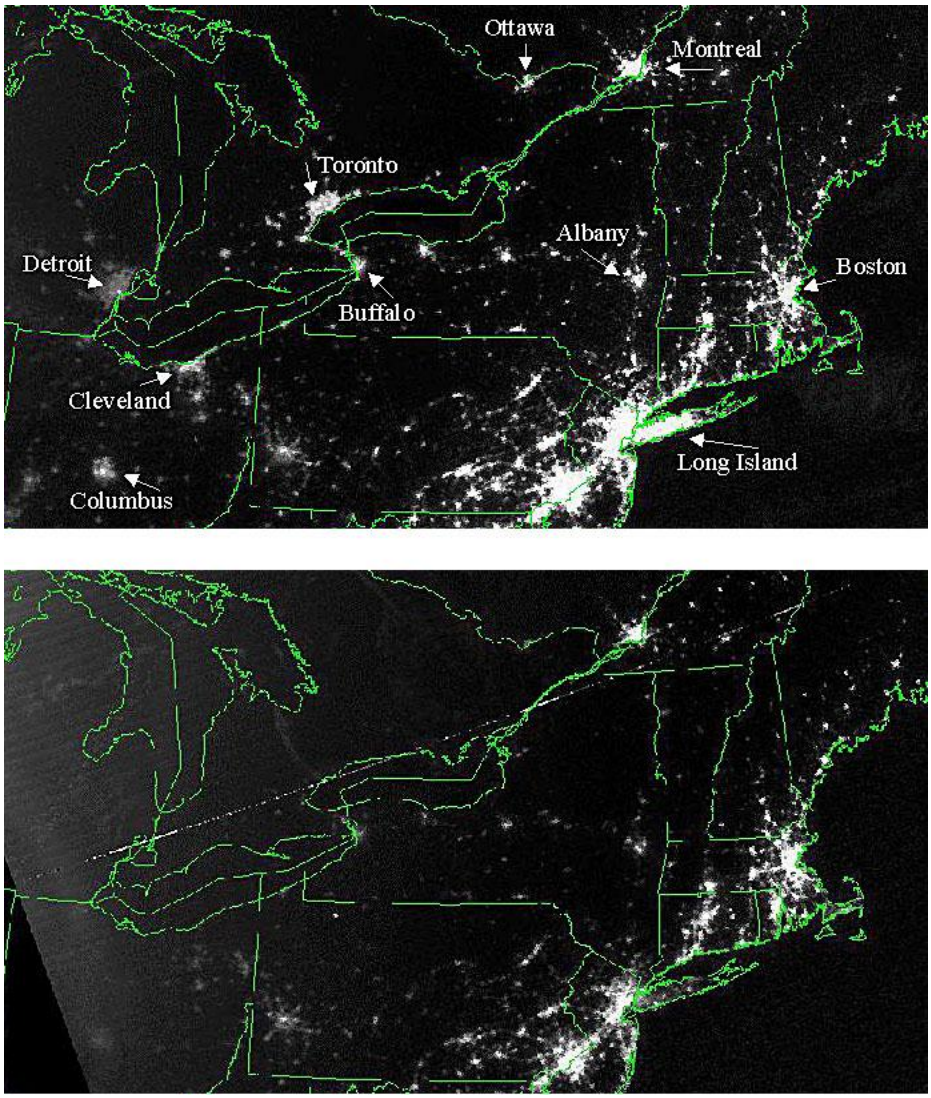


Figure 5. *The 14 Aug 2003 Northeast blackout. The upper image shows the lights in the region on the evening of 13 Aug; the lower was taken on the evening of 14 Aug, more than five hours after the grid failure. In some areas power had already been restored by then.*

Source: Defense Meteorological Satellite Program data processed by Air Force Weather Agency.

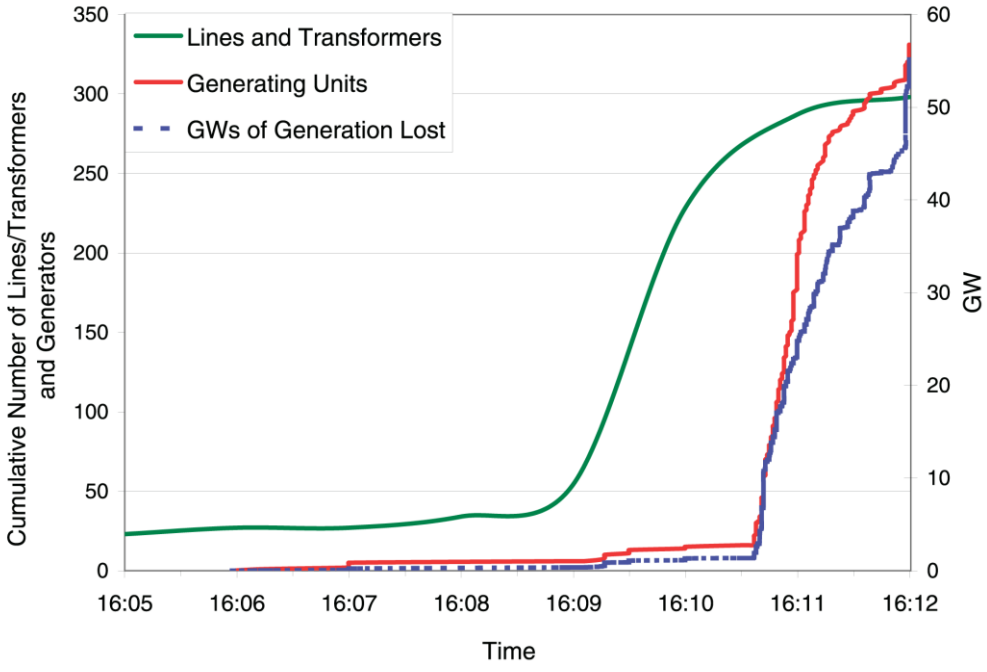


Figure 6. *Cascade of outages in the Northeast blackout of 14 Aug 2003. (Times are Eastern Daylight.)*

Source: U.S.-Canada Power System Outage Task Force, "Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations," (Washington and Ottawa: U.S. Department of Energy and Natural Resources Canada, April 2004), p. 74.