Systems that Should Have Failed

Critical Infrastructure Protection in an Institutionally Fragmented Environment

Mark de Bruijne & Michel van Eeten

Faculty of Technology, Policy and Management
Delft University of Technology
The Netherlands
PO Box 5015
2600 GA Delft
The Netherlands
e-mail: m.l.c.debruijne@tbm.tudelft.nl

Draft version which was published in: Journal of Contingencies and Crisis Management,
Vol.15, No. 1, March 2007, pp. 18-29

Abstract
Recent years have witnessed major governmental initiatives regarding Critical Infrastructure Protection (CIP). At the same time, critical infrastructures have undergone massive institutional restructuring under the headings of privatization, deregulation and liberalization. Little research has gone into understanding the interactions between these two developments. In this article, we outline the consequences of institutional restructuring for the changing ways in which critical infrastructures ensure the reliability and security of their networks and services. Neither Normal Accident Theory nor High-Reliability Theory can account for reliability under these conditions. We then investigate the implications of these findings for Critical Infrastructure Protection (CIP).

Protecting critical infrastructures post-restructuring
Critical infrastructures are part of a larger set of services and products that are considered essential to the functioning of our modern economies and societies (DHS, 2003a; 2003b; NRC, 2002). An amazingly heterogeneous set of so-called ‘large technical systems’ providing services and commodities are considered to be vital. These include but are not limited to energy, information technology, telecommunications, health care, transportation, water, government and law enforcement, and banking and finance. An array of assessments argues that the collapse of services from these systems would be disastrous for whole economies and societies. At the core of every list of critical infrastructures are the large-scale technical grids of energy, water, communication and transportation. It is this subset of critical infrastructures on which we primarily focus our attention in this article. Using a somewhat tired, but still apt metaphor, they make up the arteries and veins of Western, urbanized societies (Zimmerman, 2001; Little, 2002; Moteff et al., 2003).
The specific characteristics these systems display make them, according to LaPorte (1996:67):
- tightly coupled technically, with complex organizational and management ‘imperatives’ prompted by operating requirements designed into the system, that is, unless operations are carried out in specific ways, there are no benefits and perhaps great harm can be imagined;
- prone to the operational tendencies or logic of network systems, that is, exhibit a drive to achieve maximum coverage of infrastructure and internal activity or traffic within the network;
- non-substitutionable services to the public, with few competing networks delivering the same service (the more effective the system, the more likely its monopoly);
- the objects of public anxiety about the possible wide-spread loss of capacity and interrupted service (the more effective it is, the more likely the anxiety); and …
- the source of alarm about the consequences of serious operating failures to users and outsiders, and subsequent public expressions of fear… and demands for assurances of reliable operations.6

These infrastructures might be critical to our societies, that hasn’t stopped us from using them for the great experiments of privatization, liberalization and deregulation – also known as institutional restructuring. Restructuring has changed the ways in which the reliability of networks and services is ensured, which also affects the task of Critical Infrastructure Protection (CIP). While many experts have recognized that CIP faces the challenge of dealing with infrastructures which for the most part are in private hands, few have thought through the implications of the ways in which reliability is ensured under conditions of institutional fragmentation. This article draws out these implications, based on previous empirical work in different restructured critical infrastructures. First, we discuss how restructuring has affected the ways in which reliability can be achieved in critical infrastructures. Next, we turn to how the organizations we studied in electricity and telecommunications coped with these changes. Finally, we confront these findings with two characteristics of CIP-initiatives: the reliance on anticipatory risk management and on public-private partnerships.

**Technically interconnected, institutionally fragmented**

This widespread political attention and concern about the dependence of our modern day western societies on critical infrastructures surged in the late 1990s and received new vigor in the wake of Y2K and the terrorist attacks of September 11, 2001.7

Recent studies and reports emphasize how critical infrastructures are complexly interrelated through the increasing use of information and communication technologies and are becoming more dependent on each other’s ‘always on’ availability. This means, critical infrastructures have also become increasingly vulnerable to large-scale, cascading disruptions across sectoral boundaries (e.g. Amin, 2000; 2002; Little, 2002). As a consequence of these findings, substantial research efforts are underway to analyze and assess the vulnerabilities, single points of failures and complex interdependencies in our critical infrastructures (Cf. Luijif et al., 2003a; 2003b).

Apart from the increased interactive complexity between critical infrastructures, another important characteristic of our modern day critical infrastructures has not received nearly as much attention from researchers and policy makers: the institutionally fragmented set of organizations that operate, manage and oversee these systems. Instead of providing services through large scale integrated monopolies, with centralized and hierarchical control, vital services are
provided in unbundled and competitive markets, through a patchwork of public and private ownership structures. Governments pulled away from the ownership, management and even oversight of these systems in recent decades towards a role of ‘coordinating markets and overseeing systems owned by others’ (Abbate, 1999:115). Simultaneously, ‘the number of actors and technologies in networked industries’ increased rapidly (Coutard, 1999:8), as restructuring often included the outsourcing of activities. Restructuring policies (i.e. privatization, liberalization and deregulation) led to an increased ‘splintering’ – i.e., institutional fragmentation – in the delivery, management and development of critical infrastructures (Graham & Marvin, 2001).

As a result, we face a paradoxical challenge: while our critical infrastructures have become more complex and interconnected, the management of these critical infrastructures has become increasingly institutionally fragmented. The adoption of more horizontal, market and network-based industrial structures also changed the means though which to ensure reliability. The responsibility for the reliable provision of vital services in the infrastructures changed from a primarily intra-organizational task to an inter-organizational challenge. Instead of one or comparatively few public organizations cooperating under hierarchical command and control, large networks of organizations with competing interests became involved in the management of critical infrastructures and the reliable provision of services.

A key question that arises from these developments is: How do critical infrastructure industries, consisting of networks of organizations, many with competing goals and interests, provide reliable services in the absence of conventional forms of command and control? This raises another question that, logically, precedes it: Are institutionally fragmented critical infrastructures in fact still reliable?

**Does institutional fragmentation affect the reliability of service provision?**

The exact relationship between institutional restructuring and the reliability of services and networks has so far remained largely obscured. The available empirical data on reliability – measured in terms of the frequency and length of disruptions to end-users – fail to provide an unequivocal answer. We were able, however, to draw upon extensive field research on reliability-related issues in large-scale water systems (Van Eeten and Roe, 2002; Roe and Van Eeten, 2002), electricity grids (Schulman et al., 2004; Roe et al., 2005) and telecommunication networks (Van Eeten et al., 2005; De Bruijne, 2006). Together, these field studies comprise over 130 interviews, extensive control room observations and literature reviews.

Without repeating previous discussions of our findings, we can draw out a number of implications, primarily based on our studies in electricity and telecommunications. First of all, while there are no conclusive data regarding the reliability of services and networks post-restructuring, the data that is available suggests that the network operators and service providers have managed to cope with these changes. The two focal organizations that we studied – the California Independent System Operator (ISO) and Dutch mobile telephony operator KPN Mobile – succeeded in maintaining a high reliability of service provision. The organizations displayed virtually unchanged levels of service provision before and after restructuring. The ISO’s reliability performance during California’s electricity crisis in 2000 and 2001 – one of the most turbulent periods in which any restructured critical infrastructure industry ever operated – did not lead to outage rates that differed significantly from those of the utilities
before restructuring. In the end, the lights stayed on for most of the time, notwithstanding the popular images in the media of sweeping blackouts across. The reported rolling blackouts occurred on eight days for 27 hours, compared to the 125 days on which just 1.5 percent of operating reserves remained and stage 3 emergencies were declared. The aggregate amount of load shed during California’s electricity blackouts was quite small, adding up to no more than one hour’s worth of electricity to all residential homes in the state. This performance fell within the margins of the average annual reliability performance of the investor-owned utilities before restructuring. However, other key reliability indicators (e.g. the number of high-voltage transmission line overloads and the number of violations in the ISO’s control area) did show that the system was operating closer to the edge of failure – demonstrating the massive pressure under which the system was operating. In other words, although negative effects of restructuring could be identified, the organizations involved, most notable the ISO, managed to cope with these effects and maintain acceptable levels of service and network reliability.

Similarly, the Dutch mobile telephone operator KPN Mobile displayed a steady reliability performance from 1996 to 2001, notwithstanding seven-fold increase in customers, the rapid expansion and innovation of its mobile network and the six-fold increase in the number of services it provided over this network. From 1996 to 2001, the company displayed steadily rising call completion rates (CCR) and call setup success rates (CSSR), which in the telecommunication industry are considered key proxies for the reliability of service provision. In addition to these steadily improving reliability indicators, KPN experienced ‘only’ a 50 percent increase in the number of ‘calamities’ – which they define as incidents with an impact on customers.

While significant, this number pales in comparison to the growth rate of customers, network and services. KPN Mobile achieved this performance under cut-throat competition in the market which forced them to undertake drastic cost reductions in their operations.

Considering the effects of institutional fragmentation on how these critical infrastructures were organized and operated, the abovementioned performance of both the ISO and KPN Mobile may be considered an astonishing feat. Despite operating under conditions with significantly reduced resources time and again the organizations managed to maintain a reliable provision of critical infrastructure services. These findings are all the more puzzling since the two dominant organizational theories that are used to assess the reliability, or lack thereof, of complex, large-scale technological systems would predict a negative impact on the ability of organizations to reliably manage these critical infrastructures.

The Normal Accident Theory (NAT) (Perrow, 1999a) and High-Reliability Theory (HRT) (Roberts, 1993) both expect that institutional fragmentation caused by restructuring negatively affects the ability to reliably manage these infrastructures and that reliability of service provision accordingly should have suffered. However, the case studies did not confirm the theoretically assumed negative relationship between the effects of institutional fragmentation and ability to reliably manage these infrastructures even though infrastructure operations did become more complex to manage and behaved more volatile (De Bruijne et al., 2006). Evidence did show that the infrastructures operated ‘closer to the edge’ than before restructuring. Yet what could explain the performance record of restructured critical infrastructures and the more or less continued high reliability of the provided services in the researched cases?
Coping with institutional fragmentation

Based on these findings, it could be concluded that institutional fragmentation and restructuring not only negatively affected the ability of organizations that manage critical infrastructures to provide highly reliable services, but also offered new options that enabled organizations involved in the management of these systems to maintain reliability under extremely demanding conditions. The case studies revealed a large number of hitherto unknown or unrecognized conditions that enabled these organizations to cope with the effects of institutional fragmentation (De Bruijne, 2006). Examples include the increased use of real-time, on-line experimenting; the gradual redefinition of reliability norms and criteria to fit the new conditions and the increased use of support staff and informal wheeling and dealing in real-time in control rooms. These conditions, which many at first glance would consider detrimental to the provision of reliable services, were found to contribute to the ability of the organizations to maintain a reliable provision of services.

The research found both NAT and HRT flawed in their assumptions on the main relationships between the conditions that facilitate reliability and the levels of reliability achieved. The networked environment clearly emphasized different reliability-enhancing characteristics than those identified by NAT and HRT (cf. Grabowski & Roberts, 1996; Schulman et al., 2004). The implication is that NAT and HRT, which until now have been presented as generic organizational theories of (un)reliability, need to be modified in order to be valid under conditions of networked reliability (see also Schulman et al, 2004; De Bruijne, 2006). In general terms, we have identified three shifts of emphasis in organizational processes and resource allocation.

1. From long-term planning to real-time management

Institutional fragmentation and the introduction of competition create more volatile and technologically more complex infrastructures. Many of the procedures and routines that had been designed to reliably operate the critical infrastructures do not function anymore. Infrastructure operations used to emphasize the importance of complete information, centralized planning and command and control. Institutional fragmentation caused those in control of infrastructure operations to be confronted with less than adequate information and control, leading to more surprises and reliability-threatening events. This in turn emphasizes a need for more flexible response capability to maintain reliable services. Real-time operations – typically focused in and around control centers – increases in importance, reducing the strong reliance on long-term, detailed planning that has characterized critical infrastructures (cf. De Bruijne et al., 2006; Van Eeten et al., 2006; Roe et al., 2002).

2. From design and analysis to improvisation and experience

More volatility and complexity also means more unpredictability. As Demchak (1991, p. 3) has said, the chief manifestation of complexity is surprise. Operations move more often ‘outside analysis’, beyond the well-studied situations for which technology has been designed and procedures have been tested. Under these circumstances, relying on established procedures, routines and guidelines decreases rather than ensures reliability. In real-time, control room operators increasingly have to rely on their experience and improvisational abilities to deal with surprises and volatile events. Referential knowledge, improvisation, ‘instinct’ and experience gain precedence in comparison to detailed procedures and routines. It becomes more important to train operators to know when not to follow
procedure and how to still maintain reliability.

3. From standardized and formal to real-time informal communication and coordination

The third shift moves infrastructure operations away from formal and hierarchical towards informal and ‘rich’ modes of communication and coordination. To put it differently: real-time resists formalization. Faced with surprises and threatening events, critical infrastructure operations are constrained by hierarchical, unilateral, and formal modes of communication and coordination; albeit legacies from the pre-restructuring days or those installed after restructuring to ensure competition and level playing fields. Both types severely handicap operators’ abilities to improvise and provide reliable services. Especially when faced with reliability-threatening events, informal communication and coordination mechanisms take over or augment formal mechanisms. The need for real-time communication has already been identified in the literature on coordination in networks of organizations as well. In the absence of formal communication and coordination arrangements between organizations in networks, informal coordination and communication evolve and take over (cf. Chisholm, 1989). Powell (1990:304) finds information passed through networks (of organizations) must be “thicker” than information obtained through markets and “freer” than information communicated through hierarchies.

Real-time, ‘rich’ informal communication and coordination has been identified as one of the most important sources of networked reliability: “[R]eal-time values and privileges the non-routine over the routine, the informal over the formal, and the relational over the representational” (Roe et al., 2002:9-5). In other words, the ability of system operators to engage in a rich exchange of information and informal deals enhances their knowledge of system conditions, stimulates creativity and increases their options for maintaining reliability. To be sure, under ‘normal’ operating regimes, the need for ‘rich’ and varied communication and coordination is constrained by the competitive environment in which critical infrastructures nowadays operate. However, when threats occur and move towards real-time, ‘rich’ and informal communication and coordination become increasingly important. Real-time informal infrastructure operations enable types of interventions and control that are typically unacceptable at any other time or place.

The provision of reliable services in critical infrastructures is pushed and pulled to real-time as a consequence of institutional fragmentation and the introduction of competition. When summarizing these shifts in terms of Wildavsky’s (1988) framework of risk management strategies, we see a changing focus of organizational processes and resources allocation from anticipatory long-term planning towards real-time resilience. The ability to reliably manage critical infrastructures is pushed back to the very last minute. Infrastructure operators can only count on sufficient amounts of information, resources and the necessary authority once they get to in real-time. At that point – and often only then – were experienced system operators able to find the means and resources to maintain reliability.

Institutional fragmentation and CIP

Although the abovementioned conclusions may come as no surprise to those engaged in the study of crisis management, the specific relationship between institutional fragmentation and the push to real-time reliability management provides an interesting perspective on current CIP-policies and programs. When we confront these developments, two characteristics shared by most CIP-initiatives stand out:
(1) the reliance on anticipation as the dominant risk management strategy; and
(2) the reliance on public-private partnerships as the preferred institutional arrangement to improve CIP. Neither of these characteristics fit comfortably with what we now know about networked reliability.

CIP-initiatives typically organize a process around some form of risk assessment – be it risk analysis, threat assessment, vulnerability assessment, impact assessment, interdependency assessment or a comparable methodology (for an overview, see Dunn and Wigert, 2004). Common to all methodologies is their objective to identify threats and risks beforehand, so that measures can be taken to these threats and risks – which is typically the next phase in CIP-processes. All of these methodologies are clear examples of a risk management strategy which Wildavsky calls anticipation: a risk-averse approach which seeks to anticipate risks beforehand and to allocate resources so as to build a defense against this risk, thereby prevent harm from occurring.

While any approach to CIP would need to balance anticipation and resilience, it seems clear that the conditions for effective anticipation have deteriorated. In order to make a strategy of anticipation effective, it is necessary to know the properties of the expected risk, its probability, and the existence of effective responses. Wildavsky (1988, 2006) demonstrated clearly that the knowledge requirements and the organizational capacities required to make anticipation effective are extremely demanding. Critical infrastructures obviously still need to invest in anticipation, but they have been forced to develop more resilience-based strategies.

The focus of CIP on anticipation might not be exclusive – it also includes efforts on disaster response and emergency preparedness, for example – but it is dominant enough to hamper the difficult adaptive process currently going on in critical infrastructures. Wildavsky's main objection to anticipation was that it sinks resources into specific defenses – resources which are then no longer available for resilience and dealing with unanticipated risks. CIP’s reliance on such a strategy is particularly ironic, given that the threats which have fueled the recent surge in CIP-initiatives those posed by terrorists – who make it a point to defeat anticipation. If one is looking for examples of sinking resources in highly ineffective defenses, one only needs to look at the revamped airport check-in security procedures. Massive resources are now being spent to prevent people from taking potentially explosive liquids onto the plain. This is on top of the already expanded and costly security procedures for more conventional weapons – which, as it turns out, don't work. These procedures may prevent you from taking your water bottle onto the plane, but tests demonstrate time and again that when it comes to catching bombs and guns, they are not very good (e.g., Frederickson & LaPorte, 2002; Marsico, 2006).

In contrast to this kind of approach, Wildavsky (2006) explains that resilience, "requires the accumulation of large amounts of generalizable resources, such as organizational capacity, knowledge, wealth, energy, and communication, that can be used to craft solutions to problems that the people involved did not know would occur. Thus, a strategy of resilience requires much less predictive capacity but much more growth, not only in wealth but also in knowledge.” It is not hard to see why CIP-initiatives have a difficult time organizing “generalizable resources.” If nothing else, political accountability forces investments in specific defenses and risk-avoidance, rather than in resilience and risk-tolerance. All of this means that it is not self-evident that CIP-initiatives, even when they do reach their goals, actually result in safer and more reliable critical infrastructures.
CIP and governmental involvement

The institutional restructuring of the past decade means that most parts of the critical infrastructures are in private hands. In a policy note on the Dutch CIP-effort that was sent to the parliament plan it is estimated that as much as 70-80% is owned by market parties (Ministry of the Interior and Kingdom Relations, 2005, p. 59). In the U.S., private industry involvement in critical infrastructures is assessed to hover somewhere around 85 percent. That raises the issue of what role, if any, there is for governments in CIP. Advocates of liberalization often claim that markets should be perfectly capable of determining the proper levels of reliability of services. Why not use the same argument with regard to CIP? Many would argue that the chief difference is that security is a public good and, hence, subject to market failure (Lewis, 2005).

Without governmental intervention, CIP-efforts in privatized industries will not take into account the full social costs and benefits of security. In more practical terms, policy makers have an important role in the safeguarding of reliable infrastructures for at least three reasons:

- **Unknown, but massive societal costs of (threatening) large-scale failures**
  The organizations operating, managing and overseeing critical infrastructure face fundamental uncertainties about the paths to large-scale failures and, perhaps more importantly, the potential costs of such failures. Even though nobody knows whether the effects of large-scale failures are as devastating as certain experts forecast or believe, they often are tremendous even if the worst-case scenarios are only half true. For example, what are the societal costs of a failing Internet infrastructure, a failing banking industry, a failing railway operator or a failing electricity market? A rough but conservative calculation of the societal costs of California’s electricity crisis easily reached US$60-70 billion. Furthermore, it is often argued that these costs are bound to increase as our dependence on critical infrastructure services increases. The central argument is that society is highly risk averse, and increasingly so (i.e. Beck, 1992). Central in these discussions are not the odds of things happening, but the costs of public intervention to prevent a large-scale failure versus the absolute cost of such a failure. Framed this way, it is hard, if not impossible, to resist public intervention.

- **Institutionally fragmented critical infrastructures lack an obvious fall-back mode**
  Before critical infrastructures became institutionally fragmented, vertically integrated state-owned critical infrastructures had a series of more or less controlled fall-back solutions based on reliability engineering considerations. The institutional framework around critical infrastructures allowed those who managed these systems to employ these fall-back modes when needed. Important values and resources in critical infrastructure operations could clearly be traded off. The central political authority over state-owned critical infrastructure service providers meant that fall-back solutions could be implemented. If power lines were overloaded, electricity service providers could create rolling blackouts to maintain the overall reliability of services. Whatever the fall-back mode, their costs were borne by the collective.

  As a result of institutional fragmentation in critical infrastructures, this safety net has been largely removed. Instead, critical infrastructure restructuring largely shifted the task of ensuring reliability from operations to markets. However, markets cannot easily provide substitutions for the engineering-based fall-back mechanisms employed under vertical structures and public ownership. Consequently, restructuring significantly lengthens the list of potential reliability risks and reduces the number of fall-back...
modes that allow for a reliable provision of services. As we argued in our California study, the grid is engineered to maintain a reliability standard of N-1 – which means that the grid has to be able to cope with the failure of any one of its components. There is no such equivalent for markets – no concept of market reliability – even though there the market is the main institutional mechanism to provide the coordination of critical system operations. If the market fails to function, as it did in California, there is no fallback.

**Institutionally fragmented critical infrastructures lack an obvious ‘lender of last resort’**

Closely related to the previous consideration, critical infrastructure operations demand a ‘lender of last resort’ to maintain the reliability of service provision. The universally applied strategies to deal with problems affecting critical infrastructures in the late 19th and early 20th century were nationalization and regulation. When the small-scale and usually monopolistic private critical infrastructure operators failed to comply with public and politically desired performance levels, they were nationalized and transformed into vertically integrated utilities or subjected to tight regulation. In that period, even local governments could easily take over from failing private providers of critical infrastructures (e.g., Jacobson, 2000). Until well into the 20th century, governments stepped in if, for whatever reason, integration and nationalization of infrastructures appeared necessary. As critical infrastructure services were provided through national utilities, governments functioned as a lender of last resort to ensure a sufficient financing to maintain minimum levels of reliability.

The question that becomes increasingly important is whether national governments are able to make up for this role in a world of internationally interconnected restructured critical infrastructures and globally connected critical infrastructure markets? The risks and especially the (financial) consequences of large-scale disruptions of services to society become so large that governments of nation-states will find it increasingly difficult to play a leading role in efforts to maintain or recover reliability.

But when we consider the declining influence and possible roles of national governments as ‘lender of last resort’, a second question needs answering: do people still expect governments to step in and maintain the reliability of service provision in the face of large-scale threatening events? If we look at California’s electricity crisis as a litmus test, the answer is a resounding Yes. Even though this role was not envisioned in the institutional design and even though there was no formal reason that required the state government to step, it did. The pressure was so overwhelming that it forced the state to more or less drive itself into bankruptcy.

Faced with threats of large-scale critical infrastructure disruptions, governments will be more or less pushed to become the lender of last resort. Societal pressure for both fair prices and sufficient levels of reliability pushes political representatives to act, even when formally they may lack the authority to do so. Governments would do well to acknowledge this task and prepare for it. A failure to provide reliable critical infrastructure services at reasonable prices and the threat of large-scale failures will only increase demands for political oversight and scrutiny as well as public involvement (cf. McCurdy, 2001; Heimann, 1997). The reliability of service provision we as a society expect from our critical infrastructures is ‘always-on’. Institutionally fragmented critical infrastructures increase the moral hazard problems in the absence of the institutions that shielded society from large-scale disruptions in reliability of service provision in the past.
CIP and the reliance on Public Private Partnerships
If governments do have a role in CIP, it is much less clear what that role entails exactly. As a consequence of institutional fragmentation, CIP has moved from an activity embedded in the public sphere towards one where “[g]overnments are increasingly dependent on other parties when it comes to protecting critical infrastructures” (Luiijf & Klaver, 2004:1189). In other words, in a competitive environment, divergent or even conflicting interests between organizations involved in CIP are to be expected. Ongoing processes of liberalization and globalization will only increase this feature in CIP in the near future. In this sense, the task of governments involved in CIP is roughly similar to the organizations responsible for the reliable provision of services in the previously described research: they may have, formally or informally, the overall responsibility for the reliable provision of services, but they lack the authority and resources to actually fulfill that responsibility. Central governments bodies and policy makers involved in CIP to a large extent lack the technical expertise and the means to monitor or control critical infrastructure operations. This void has been acknowledged in many CIP-policies and publications. Policy makers and researchers have responded to this dilemma by emphasizing the need for Public Private Partnerships (PPPs) (e.g. PCCIP, 1997; DHS, 2003a, Anderson and Malm, 2006).

While the need for PPPs may seem self-evident, such arrangements actually dodge the underlying question: how is such voluntary collaboration supposed to work in an institutionally fragmented environment? All parties may agree that CIP is important, but that shallow consensus quickly disappears once it becomes clear that governments want the private sector to invest in security and reliability beyond its normal business continuity requirements. Often, as Schneier notes (2003, p. 41), these public claims are articulated in moral terms that simply ignore that underlying institutional issue: “Officials appealed to the CEO’s sense of patriotism, reminding them that improving safety would help their country. That this had little real effect should surprise no one. If the CEO of a major company announced that he was going to reduce corporate earnings by 25 percent to improve security for the good of the nation, he would almost certainly be fired.” Somewhat less academic in tone, he adds: “If I were on the company’s board of directors, I would fire him. Sure the corporation has to be concerned about national security, but only to the point where its cost is not substantial.”

Similarly, researchers have argued how “most critical infrastructure stakeholders start to understand that national and international collaboration is inevitable” (Luiijf & Klaver 2004, p. 1189) without indicating why exactly this is ‘inevitable’ and why this would make private companies willing and able to bear the costs of increased CIP (see also Lewis, 2005). After arguing at length that PPPs are the preferred strategy for CIP, Anderson and Malm (2006, p. 159) end their analysis with these two sentences: “Who will foot the bill for agreed emergency preparedness measures? For PPPs to succeed... these types of issues must be resolved.” It seems to us, that should have been the start of their analysis, not the outcome.

In our research we did find some evidence for the existence of what one could call a collective responsibility for reliability, but it only existed in real-time. Private parties shy away from committing to any responsibility for the provision of reliable services beyond their organizational mandate until confronted with the threats of large-scale disruptions in real-time. To paraphrase a well-known saying: the prospect of hanging focuses the
mind wondrously. In the absence of such a prospect – that is, outside of real-time – such a collective responsibility seems to evaporate.

In sum, most CIPs find themselves at intersection of incongruous institutional roles: government argues that measures are needed to protect the public interests, but that it is the responsibility of the private infrastructure owners and operators to implement these measures. The private parties, in turn, predictably resist taking measures that go substantially beyond their business continuity requirements, arguing that these threaten the viability of their business model. That leaves the government with two basic options. The first is to provide the necessary resources itself – for example through tax incentives or subsidies. This seems highly unlikely. The costs of implementing a substantial CIP-program are prohibitive for most governments – not to mention the associated problems of rent seeking and moral hazards. The second option is regulation. Here, we come full circle. The reason governments en masse adopted PPPs is because they typically lack the expertise, the authority – also in light of internationalized markets and treaties – and the ability to adapt standards to the rate of technological change and innovation in many infrastructures. Unsurprisingly, most CIP-initiatives do neither. That leaves them restricted to working on issues that stay close to the status quo, such as raising awareness, exchanging best practices or identifying possible measures which have no commitment power with regard to private actors.

All of this looks pale in comparison to the ambitious language that characterizes most CIP-initiatives. As long as they do not overcome the challenge of institutional fragmentation, these well-intentioned initiatives bear an uncomfortable resemblance to what Schneier (2004, p. 38) calls “security theater” – measures that provide the feeling of security, rather than the reality.

Conclusion
Our findings lead us to the rather sobering conclusion that current CIP-efforts seem very vulnerable with regard to institutional fragmentation and networked forms of reliability. The preceding discussion suggests two basic implications for developing more effective CIP-initiatives. First, CIP needs to better balance anticipation and resilience. This means sinking fewer resources in specific defenses against the plethora of risks, threats and vulnerabilities that come out the formal assessment procedures that are typical of CIP-processes. Rather, it requires the development of what Wildavksy (1988) calls “generalized resources.” Typically, these means improving the knowledge base and operating experience of the relevant organizations. CIP could contribute by developing real-life simulation exercises and gathering other forms of intelligence that better mimic the surprising events that these systems need to deal with. A fringe benefit from such an approach is also that these measures are substantially less expensive than investments in specific infrastructure upgrades to avoid certain risk scenarios which may or may not occur.

The issue of cost brings us to our second implication: either CIP-initiatives accept the current incongruous institutional roles between public and private actors and they adjust the objectives and approaches accordingly – i.e., work within the status quo; or they set out to establish a different governance arrangement for CIP which overcomes this issue. So far, they’ve done neither. PPPs have been used to mask this flaw and that has left us with some elaborate examples of security theater, rather than protected infrastructures.

Networked reliability may have some sobering implications for CIP, but what about the other way around? Some would argue that networked reliability isn’t really reliability, but operating at the edge
of failure, mixed with a dose of luck. It doesn’t fit with CIP precisely because it relies too heavily on resilience, rather than anticipation. Remember, it was anticipation that made these infrastructures so reliable in the first place. What we now see at organizations like CAISO and KPN are the remnants of what were once High Reliability Organizations, before institutional fragmentation exposed them to conditions that no HRO should be exposed to, gradually degrading their reliability-seeking processes into something that looks more like crisis management than anything else.

That may be true. We are struck, however, by how these organizations are able to operate at the edge of failure, without falling off. We would further hypothesize that there are more like them – dare one say, a ‘class of organizations’ – that according to NAT and HRT should fail, but don’t. Neither NAT and HRT can explain how an organization could survive at the edge. Both abhor the edge and the conditions that drive organizations there. If our hypothesis is correct, that means that studying such organizations could point us to patterns that allow these organization to occupy such uncomfortable niches – in the same vein as the way in which the original HRO-research challenged NAT. We would expect real-time resilience to be an important part of those patterns. Wildavsky (1988) advocated resilience, but perhaps what he and other didn’t quite realize is how unsettling the experience of resilience is, both to the organizations and its observers.

References


Amin, M. (2005), ‘Powering the 21st century, we can – and must – modernize the grid’, in: IEEE power & energy magazine 3(2), pp. 96-94


Beck, U. (1992), Risk Society: Towards a New Modernity, Sage, Londen


Booz Allen & Hamilton (1997), Economic Impacts of Infrastructure Failures, Report to the President’s Commission on Critical Infrastructure Protection, Washington D.C.

Chisholm, D. (1989), Coordination without Hierarchy, Informal Structures in Multiorganizational Systems,


Technologies, University of Michigan Press, Ann Arbor, MI
Mayntz, R., and T.P. Hughes (Eds.) (1988), The Development of Large Technical Systems, Campus Verlag, Frankfurt am Main, Germany
National Research Council (NRC)(2002), Making the Nation Safer, The Role of Science and Technology in Countering Terrorism, National Academies Press, Washington D.C.
Office of Science and Technology Policy (OSTP)(1998), Cybernation: The American Infrastructure in the Information Age, National Security and International Affairs Division, Washington D.C.
Organisation for Economic Cooperation and Development (OECD)(1993), Infrastructure policies for the 1990s, Paris


Critical Infrastructure Protection (CIP) is by no means a new phenomenon. Already during the late 19th century, infrastructure industries that provided products that were considered beneficial, if not essential to the military, economic and social development could count on the warm interest of policy makers from nation states (Jacobsen & Tar, 1995; Schneider & Jäger, 2003; Millward, 2004). In fact, one could claim that the dominant CIP-policy during the late 19th and 20th centuries stimulated if not caused the nationalization of critical infrastructures. Especially in Europe, the political, military and economic benefits of expansion, standardization and integration of infrastructures stimulated that most infrastructure industries became firmly

1 Large Technical Systems (LTS) may be defined as “[s]ocioeconomic or infrastructural sectors built around a core technology” (Schneider, 1991:20), or “[c]omplex and heterogeneous systems of physical structures and complex machineries which are materially integrated, or ‘coupled’ over large spans of space and time, quite irrespective of their particular cultural, political, economic and corporate make-up, and support or sustain the functioning of very large numbers of other technical systems, whose organizations they thereby link” (Joerges, 1988:24). These systems are composed of networks of humans and technical resources with social and organizational connections and relationships. Large-scale technical systems perform tasks and support the missions and goals of more than one organization (e.g. Mayntz & Hughes, 1988; Genschel & Werle, 1993; Grabowski & Roberts, 1996; 1997).

2 In the Netherlands, 11 vital sectors and 31 vital products and services were identified during a quick-scan study (Luijff et al., 2003a). Cf. Moteff, et al. (2003).

3 See Luijff et al. (2003b), Bijlage E for an international comparison of government efforts to protect vital sectors that produce essential services and commodities for modern Western societies.

4 Exactly what the effects of large-scale, cascading failures in our critical infrastructures will be is less well known (Steetskamp & Van Wijk, 1994; Booz Allen & Hamilton, 1997). Although experts refer to the consequences of natural disasters and/or large incidents (e.g. OSTP, 1998; Schneider et al., 1998; Little, 2002; Boin et al., 2003), little effort has been made to systematically study the potential effects of large-scale cascading outages on society. However, some studies provide a glimpse into the unknown. Rusman (2000) mentions a Swiss study that estimated that banks were, on average, able to survive without electricity for up to 48 hours before being seriously threatened with bankruptcy. Likewise, the study estimated that insurance companies could cope without electricity for a maximum of 100 hours. Rusman, P. (2000), ‘War without bombs and shells’ (In Dutch: ‘Oorlog zonder bommen en granaten’), in: Vrij Nederland, February 19, 2000. A Dutch study in the mid-1990s estimated that the effects of large-scale electricity disruptions lasting for more than eight hours would have increasingly severe consequences for other infrastructures and society (Steetskamp & Van Wijk, 1994).

5 While preparing for the millennium transition, the Dutch government identified four vital national infrastructures – energy, drinking water, telecommunications and banking. Large-scale unavailability of these infrastructures would seriously disrupt society within eight hours (Coördinatiecommissie Millennium OOV, 1999:2).

6 According to Fay and Yepes (2003:2), the world’s infrastructure stock is valued at about US$15 trillion, of which 60 percent is found in industrialized countries. Infrastructure industries and the services they provide significantly influence the performance of national economies and provide a substantial amount of capital (Kessides, 2004:29; OECD, 1993). For instance, network utilities in the United Kingdom accounted for 18-30 percent of total net fixed assets in the United Kingdom between 1850 and 1960, while in 1995 the worth of utility stocks accounted for 15 percent of GDP. In the United States in 1992, the gross capital stock of telecommunications alone amounted to 11 percent of GDP (Newberry, 1999:27).

7 However, Critical Infrastructure Protection (CIP) is by no means a new phenomenon. Already during the late 19th century, infrastructure industries that provided products that were considered beneficial, if not essential to the military, economic and social development could count on the warm interest of policy makers from nation states (Jacobsen & Tar, 1995; Schneider & Jäger, 2003; Millward, 2004). In fact, one could claim that the dominant CIP-policy during the late 19th and 20th centuries stimulated if not caused the nationalization of critical infrastructures. Especially in Europe, the political, military and economic benefits of expansion, standardization and integration of infrastructures stimulated that most infrastructure industries became firmly

17
embedded within the public sector. During the Cold War, critical infrastructure protection figured prominently as an integrated element in both government and military plans and preparations (cf. Luijif, 2003a; Farrell et al., 2004). After having been more or less neglected for the better part of a decade since the end of the Cold War, widespread public (policy) attention and concern about the dependence of modern day Western societies on such basic needs as electricity and transportation only (re)emerged in the late 1990s. Instead of focusing on large-scale, armed conflicts, attention instead turned to ‘new’ uncertainties provided by the increased complexity and interdependence of our critical infrastructures (PCCIP, 1997). Protection from conventional military attack was replaced by protection from so-called ‘cyber-threats’, which sought to make use of the vulnerability of interconnected and computerized critical infrastructures. This focus gained new vigor in the wake of the Y2K problem (Perrow, 1999b). The more ‘conventional’ string of terrorist attacks since September 11, 2001 (Wallace et al., 2002) showed how CIP at the same needed to pay attention to its roots. Consequently, many Western countries have invested heavily in research to protect their critical infrastructures from external threats such as malicious attack by terrorists and cyber threats (see Luijif et al. (2003b) for an international comparison of government efforts to protect vital sectors that produce essential services and commodities for modern western societies).

8 Until they were subjected to restructuring, critical infrastructure development was characterized by the parallel and mutually reinforcing trends of technical growth, expansion and interconnection, coupled with an ongoing centralization and vertical integration in management. Although the history of different critical infrastructures varies considerably across countries and sectors, the dominant pattern shows infrastructure industries developing “[f]rom local to regional and finally large-scale, integrated, hierarchical systems” in the course of the 19th and 20th centuries (Coutard, 1999:3). There are few comprehensive scientific studies that provide an overall and integrated assessment of the effects of restructuring or the reliability performance of any infrastructure industry, let alone a comparison of different infrastructures (e.g. Héritier, 2002). A small sample of key reliability performance indicators yields contradictory findings. For example, some claim that the reliability of service provision of the electricity industry in the United States has not been significantly affected by electricity restructuring (Kendall, 2001). On the other hand, other reports have shown disturbing trends and warned that restructuring reduced the reliability of electricity infrastructures (Jansson & Michelfelder, 2004; Amin, 2003; 2005).

9 For a more in-depth discussion of this argument, see De Bruijne (2006).


12 Total societal costs of California’s electricity crisis may be roughly calculated by comparing electricity expenditures before, during and after the crisis. The costs of electricity in California’s restructured wholesale energy markets costs soared from US$7.4 billion in 1999 to US$27 billion in 2000 and US$26 billion in 2001, only to fall back to about US$11 billion in later years (De Bruijne, 2006: 130). Even allowing for increased gas prices, the crisis at least added US$30 billion to the energy costs. Not yet included in these estimates are the long-term power purchases that were undertaken by the California Power Authority and financed by the state of California. Estimates of the costs of these contracts range between US$40–43 billion, which in hindsight seem to have been priced too high as much as 50 percent representing another $20 billion (e.g. Sweeney, 2002). Then there are the direct costs related to the actual electricity disruptions, which according to Rinaldi et al. (2001:11), disrupted key industries with interruptible power contracts (e.g. Silicon Valley) and “led to billions of dollars of lost productivity” which Weare (2003:3) estimates between 0.7 and 1.5 percent. Assuming the state’s 2001 gross state product (as provided by California’s Department of Finance (see: <www.dof.ca.gov/HTML/FS_DATA/LatestEconData/Data/Miscellaneous/Bbrank.xls>, August, 25, 2005) to be affected by the crisis. This would add another US$9.5-20.4 billion. This brings the total sum of societal costs from California’s electricity crisis somewhere between US$ 60-70 billion.

13 Like some of Perrow’s high-risk systems, critical infrastructures are amongst the high-risk systems that are not easily abandoned (1999a). Consequently, redesigning these systems seems the only option left.