Innovation in Networked Infrastructures

Coping with Complexity

Margot P.C. Weijnen & Ivo Bouwmans, TU Delft

Based on a keynote lecture at the IEEE International Conference on Systems, Man and Cybernetics
10 October 2004, The Hague, The Netherlands

Introduction

Infrastructures are the systems that provide energy and water, that remove waste water and wastes, that facilitate the movement of people and goods, and that enable us to communicate and exchange information without being troubled by distance. Infrastructure systems are designed to satisfy specific social needs, but they shape social change at a much broader and more complex level. Electric power supply has radically changed our households, and telecommunication services and the Internet have changed mobility patterns. Like the railway and airway transport infrastructures, telecom and information infrastructures have greatly accelerated the internationalisation of companies and markets. They have also created the platform on which new infrastructures for financial transactions, health care and education could emerge.

Infrastructures are so deeply embedded in all economic and social activity, that they are often taken for granted, and they are used without specific reflection. When everything runs smoothly, we are unaware of the complexity of infrastructures and infrastructure-related vulnerability. We only realize this vulnerability when faced with service interruptions. A number of large-scale power blackouts (London, Italy, California, New York) and successful virus attacks on the Internet have brought to light how critical the role is that infrastructures play in our economies. The terrorist attacks on the World Trade Center and the Madrid railway stations have, in a different way, contributed to an emergent feeling of vulnerability.

Causes of infrastructure malfunctioning

There are many causes for infrastructure malfunctioning. Rather than providing an exhaustive overview, the following examples serve to illustrate the diversity of causes.

Natural disasters are an abundant source of service interruptions. The November storm in France in 1997 tore down major parts of the electricity transmission network, leaving many remote areas without electricity until well past Christmas. Snow and ice are a notorious cause of accidents in transport infrastructures, as well as a major cause of disruption in North American power distribution systems, due to breakage of the overground distribution wires. Thunder storms, strong winds and fog are a well known cause of substantial delays, and even accidents, in air traffic. Earthquakes and floods destroy the physical infrastructure. It is obvious that infrastructures cannot be completely protected against the forces of nature because of the excessive costs this would entail.

Human error is another source of infrastructure failure. Fatal errors can be made in working the points of railway infrastructure; human errors on the part of air traffic managers can also have catastrophic consequences; accidents may be caused by engine drivers, car drivers or aircraft pilots who overlook or disregard traffic instructions. Maintenance works are a major cause of interruptions in all infrastructures, not only for the maintenance interval itself, but also as a cause of subsequent failure (start-up and shut-down are not 'business as usual'); many underground pipelines, wires and cables are prone to damage by excavation works, if their exact location is not well administrated. Unfortunately this is all too often the case, as show by the occasional interruptions in electrical power, natural gas and drinking water distribution and in the provision of telecommunication and information services.

Sabotage of power plants and transmission lines, oil production wells and pipelines, drinking water sources and treatment plants, transatlantic telecommunication cables etc. may affect large numbers of end users in a vast geographical area. Recent terrorist acts have brought about a strong public and political awareness of the vulnerability of vital infrastructures.

The effects of wear and tear of specific components in an infrastructure are not to be neglected as a source of service interruptions. Depending on the component affected, a local failure may cascade down to other parts of the network, so that eventually many users are hit by service interruption. Preventive maintenance is obviously a relevant factor in reducing the risk of technical failure.

Probably the most frequent cause of service interruption or quality loss is simply capacity overload. Examples are traffic jams on the road system and power blackouts during peak demand caused by the use of air conditioners in summer, when the capacity of power stations is reduced by high ambient temperatures. If these conditions are forecasted, capacity rationing (service brownouts) could be considered. Unexpected peak demand or capacity loss will result in massive congestion and service blackouts.

As already mentioned, this overview is not exhaustive. Moreover, the causes of malfunctioning are often interrelated, as in the case of human errors following a natural disaster. It is clear, however, that the causes are diverse. This makes it unlikely that failures can be prevented completely, but we can at least design for robustness and resilience.

Costs of infrastructure malfunctioning

All these examples of infrastructure malfunctioning entail massive social costs. Accurate cost figures are hard to get. Cost estimates on one and the same incident often diverge widely, depending on cost definitions and strategic interests. Nevertheless, some cost figures may show the order of magnitude.

The costs of 'normal' road congestion in the EU are estimated at around 2% of GDP (ECMT, 1998). Road traffic accidents add another 1.5-2% of GDP in developed countries (Jacobs et al., 2000). The costs of the terrorist attacks of 11 September 2001 were estimated to be over one billion euros for the USA airlines alone (DG TREN, 2003).

In the telecom sector the primary cause of failure of services—breaks in cables—annually cause damages in the order of $500 million in the USA (Grover, 2004). Apart from this type of disruptions, there are other causes (software errors, human errors, overload) that raise this amount.

The costs of the major power failure in the USA and Canada on 14 August 2003, which affected 50 million people, were estimated at around 6 billion US$ (Elcon, 2004). The costs of the Italian power failure on 28 September 2003, which affected some 57 million people, can be assumed to be of the same order of magnitude. EPRI estimated the annual cost of electricity interruptions all over the USA to be 119 to 188 billion US$ (Primen, 2001).

Granted that there are many uncertainties in these figures, it can be concluded that even the more conservative estimates of the direct costs of interruptions in services are substantial, in all developed countries. As our economies are increasingly specializing and are more exposed to global competition, our dependence on these vital structures is steadily increasing. A growing sense of vulnerability is prompting us to review the mechanisms through which we have, until recently, ensured the reliability of infrastructure.

Infrastructure complexity

The notion of 'infrastructure' generally refers only to the physical network that connects the suppliers and end users of an infrastructure-bound service. However, in our view, an infrastructure system includes—besides the transport and distribution network—the carriers, conversion and storage facilities as well as the governance, management and control systems that are needed to make the system meet its functional specifications and its social objectives.

Any infrastructure is in itself a highly complex networked system, that includes both a physical and a social network (see figure 1). The behaviour of an infrastructure cannot be understood merely by looking at the structure and dynamic behaviour of either the physical or the social network. Both are interconnected and interwoven in many ways. The social and economic value of the infrastructure is only determined by the service it provides to its end users. The quality and reliability of that service, in turn, are determined by the performance of the integrated system, i.e. the integrated physical and social network. Both networks are complex in themselves – and their interaction, at the level of the integrated socio-technical system, presents another domain of complexity.

Innovation in infrastructures – the layered model

As already illustrated, there is an undeniable sense of urgency from an economic perspective to place the reliability of infrastructure-bound services high on the political agenda. Adding the complex system perspective, there is also abundant reason to place the reliability of infrastructures high on the agenda for research and innovation.

Many innovations are underway that may strengthen the functioning of vital infrastructures. Some of these innovations are technological innovations at the level of the physical assets (see figure 2), whether at the level of the links and nodes in the network, or at the level of their components. Relevant innovations are also concerned with the introduction of new control engineering concepts, often triggered by changes in the organization and management structure. The selection of innovations in this layer, whether technological or organizational, is directly influenced by the public governance structure. The highest level in this conceptual model of a layered infrastructure represents the layer of service provision to the end user. Especially at this upper layer, which also has the highest level of value added, a major part of the current innovation effort is directed.

Many infrastructures, in many countries, are now in the midst of turbulent technological and institutional change. The outcome of these change processes is hard to predict. The systems are so complex that it is often unknown how a local innovation will work out for the performance of the system as a whole, on the short term or on the long term. In order to steer the processes of change and innovation towards social benefit, a better understanding of the structure and dynamic behaviour of infrastructures is needed. In the following sections, we will try to unravel the complexity of infrastructures.

Complexity in the physical network

The physical networks of infrastructures are huge systems composed of a multitude of nodes and links. The effect of the number of nodes seems quite obvious: the more nodes, the more complex the system will seem. These nodes are not passive, but they interact with and adapt themselves to their surroundings. Their reaction to external changes is often non-linear, which can result in unpredictable behaviour of the system as a whole. Deterministic chaos is one of the forms of this unpredictable behaviour. Chaos theory shows that even if it were possible to accurately describe the changes in the nodes as they are influenced by their environment, the prediction of the state of the system could gradually diverge from its actual state, due to the exponential growth of tiny errors in the measurement of the system's initial state. Chaos theory can, however, give insight in the general behaviour of the system and the system states that can be expected.

Studies on complex systems often use the concept of *agents* for those elements that interact. In general, an agent is a model for any entity in reality that acts according to a specific set of rules. Its actions depend on input from the outside world. Agents may model people, but they can also model technical systems or components.

An elementary type of agent is the cellular automaton. It can only have a limited number of states and it changes state according to its present state and that of its neighbours. A computer is a large collection of cellular automata. In principle, all systems can be modelled by cellular automata given the right rules (Wolfram, 2002). It is evident, however, that in the case of an infrastructure system it would be a laborious—if not impossible—task to find out all the rules needed.

Agent Based Modelling theory uses agents that act and interact according to relatively simple rules to get a better insight into system behaviour (see Netlogo, 2005, for examples). The emergent behaviour, which is the behaviour of the system seen as a whole, follows from the behaviour of the agents at the lowest level.

This emergent behaviour can be much more difficult to describe than the actions of the individual agents at the lowest level. For example, the switching between states of a simple cellular automaton can approach complete randomness, which is the system behaviour that is the most difficult to predict.

On the other hand, emergent behaviour may be simpler to describe than the combined behaviour of the individual agents, as in the case of ants that have rules to follow pheromone trails set by other ants. The system behaviour is that all ants eventually follow the shortest route from the ant nest to the food supply.

An analogy can illustrate the concept of emergent behaviour. The laws of physics describe the behaviour of the elementary particles that form the atoms. The same laws dictate how the atoms combine to form molecules. However, it would be extremely difficult to describe any but the very simplest of chemical reactions on the basis of these laws. It is no wonder, therefore, that the periodic table was drawn up, based on observations of the way in which the chemical elements react. This is emergent behaviour of the laws of physics at the level of the elements.

One step further, biology (macromolecular biochemistry) uses entities such as proteins and enzymes to describe the behaviour of complex molecules. This, in turn, is emergent behaviour of the laws of chemistry at the level of molecules.

This analogy illustrates that emergent behaviour can be modelled without necessarily modelling the behaviour at lower levels of the system or at the level of its components.

In many cases, in both physical and social networks, emergent behaviour can be described by new rules, new laws, at higher levels of aggregation, disregarding the actions at the lowest level of the agents. This coincides with the notion that the value of physical infrastructure is only determined by the physical system's performance at the top level. The end users do not care what switches or cables are used, as long as they can make a phone call, watch the television, and have a comfortably heated home.

Part of the problem with understanding the behaviour of infrastructures is that most of the physical systems were not designed as an integrated system, but gradually evolved over time. A typical example is the way the electricity infrastructure evolved in the Netherlands (Hesselmans, 1995). Like most infrastructures, it originates in local networks, established through private initiative. City networks were established around the beginning of the 20th century. Interconnection of local networks and network expansion to rural areas were forged through government intervention. Provincial networks thus emerged in the first half of the 20th century; the national grid was established in the second half. Over time, the density of inter- and end-user connections increased. Transport and processing functions in the infrastructure were intensified (increasing throughput) to serve a steadily increasing number of users and a steadily increasing demand per user.

National grids were interconnected across national borders to improve the reliability of service. At the moment, most national grids in Europe are interconnected.

Infrastructure networks generally do not grow randomly (Barabási, 2002). New nodes that are added to the existing network are usualy linked to specifically selected nodes. In the gas network, for instance, new urban areas will be connected to existing main pipes. In the world wide web, new pages link to pages that already have a large number of links to them.

The advantage of such a "preferenced", or "associative", or "scale-free" network is that the network becomes very robust against failures of single nodes, or even multiple failures of random nodes. The disadvantage is that such a network is extremely vulnerable to targeted attacks. Eliminating just a few of the main hubs in the network may cause large parts of the network to become disconnected.

Complexity in the social network

In the previous sections we have not yet explored the social network that is part of the infrastructure. Each actor in the social network can be described as an agent who acts according to a set of rules determined by legislation and regulation, moral and cultural codes, etc. However, in addition to the explicit rules, each actor will have its own strategy. In the case of, e.g., a driver on the highway, behavioural psychology may shed light on the determinants of individual driver behaviour. However, if the actor is a company competing in an infrastructure-related market, business economics and strategy will influence the actor's behaviour. An incumbent actor may drop prices in order to try to keep new entrants from the market, whereas a new entrant may lobby for tighter restrictions on the incumbents' strategic behaviour. In their behaviour, actors also demonstrate learning. When faced with a similar situation for a second time, their behaviour will be influenced by the lessons drawn from the first time.

So, in infrastructures, it is not only the physical network that poses a modelling problem because of its complexity. The social network may be an even more serious modelling problem, given the often seemingly irrational behaviour of consumers and the hidden agendas of companies. Dynamic actor network analysis, game theory and simulation gaming are some of the methods employed in research to acquire a better understanding of the behaviour of the multi-actor networks in the world of infrastructures.

It will be evident that it is a difficult task to acquire an understanding of network behaviour on the basis of individual agent properties. Therefore there is a need for alternative methods that help us understand the dynamic behaviour of complex infrastructure networks.

Complexity in the socio-technical network

Until about a decade ago, most infrastructures were run as public monopolies, dominated by an engineering culture, with an almost exclusive focus on the technical assets. The infrastructure value chain was fully integrated vertically (see figure 3, left). This integration was justified by the fact that some of the technological facilities, most notably the transport and distribution network, have (or used to have) the character of a natural monopoly. In this monopoly culture, innovation was dominated by a technology push approach. By lack of competition, there was no obvious need for service innovation. Whether directly or indirectly, the government controlled the realization of the infrastructure and the universal provision of the public utility service by means of central planning and allocation of funds.

This rather transparent situation has changed dramatically. Many infrastructures, in many countries, are in the midst of a transition from a vertically integrated monopoly structure to an unbundled value chain, with competitive markets being introduced in those segments of the value chain that do not exhibit a natural monopoly character. This transition does not directly impact the technical network, but mainly the social network (see figure 3, right). As a new playing field is defined, new actors enter the playing field, often in new roles. In the European electricity markets, traders and brokers have become active. New market places have emerged; all over Europe, power exchanges and power pools have been set up. National regulators have been created to ensure non-discriminatory access to the transmission and distribution networks. Many of the technical assets have been privatised.

It will be evident that the social network has become much more complex with the larger number and variety of actors. It is also evident that the options to steer the development of the technical infrastructure have changed in nature. Whereas the old situation allowed the government to directly interfere in the planning of the physical system, investment signals (whether for the purpose of innovation or capacity expansion) are now established through market forces. The system has become a socio-technical network.

In the power system in The Netherlands, for example, there used to be only the physical system to worry about. The government interfered in the planning of generation capacity. The planning was actualy supply-driven, with a tendency towards over-investment to ensure reliability of service, including long-term reliability of service. The roles of transmission and system operator were combined in one organization, run by the collective of power utilities. In the present situation it is no longer the physical system, but the economic system that dominates the scene. Competition has been established in the power generation market as well as in the service provision market. Only the networks are treated as a natural monopoly, so that a precarious interface must be constructed between the competitive markets on the one side, and the regulated network on the other. On the short term, this system appears to produce an acceptable price/quality ratio. For the longer term, however, there appears to be a tendency towards under-investment in generating capacity, as no private investor is prepared to invest in unprofitable overcapacity for the public good. Research has shown that new mechanisms, such as a capacity market, must be brought in place to alleviate this risk (De Vries, 2004). Other concerns are emerging with respect to network quality on the longer term. As the California power crisis has shown, excessive social cost may be incurred by a faulty market design.

As the performance of the infrastructure is determined by the interplay of the technical and social networks—together forming the socio-technical network—, it can be concluded that ensuring high service reliability has become much more challenging for the following reasons.

Rather than being a product of individual organizations, highly reliable services are more and more the outcome of networks of organizations, many with competing goals and interests. This creates new challenges for effective market and network regulation and new needs for communication and information sharing. The California crisis could have been much worse than it was if the operators of the various subsystems had communicated less intensively (EPRI/PIER, 2002).

The overall demand for infrastructure services is increasing. At the same time, society demands ever higher reliability of service as we grow more dependent on infrastructure-bound services. Given the market tendency towards capacity scarcity, this poses new demands on infrastructure capacity management. Innovations are needed to make better use of available capacity. Where central coordination mechanisms are lacking, making way for distributed control strategies, infrastructures should be equipped with self-organizing and self-healing properties so as to deal intelligently with disturbances and recover more effectively from incidents (Van Mieghem, 2005).

By lack of possibilities to directly intervene in the development of the technical system, the only option is to ensure that the collective actions of players are steered towards the public interests through adequate market design, adequate network regulation (in cases where the network has retained its monopoly character) and additional legislation and regulation for safety, health, environment, etc. Innovative market designs are needed. For example, an energy-only market may not ensure sufficient investment in new power generation capacity, unless an additional capacity market mechanism is introduced.

In many cases, we have to make do with the existing infrastructures. Most of the established systems, capital intensive as they are, and embedded in the economic and spatial structure, are slow in responding to changing economic conditions and changing user demands. In the design of new infrastructures, it is still unclear how the many uncertainties that the system will face during its projected lifetime should be dealt with properly. Opportunities for greenfield infrastructure design are relatively scarce. However, if they arise, the challenge is to provide the technical flexibility and the budget flexibility that ensure the adaptivity of the design to changing requirements: changing user demands, changing public values and changes in technology. Some of the methods employed for this purpose are scenario analysis and real options theory. Life-cycle-costing approaches help to explicitly trade-off the capital cost of intrinsic reliability and maintenance costs in the conceptual design of the infrastructure system.

A specific challenge is to stimulate private actors to invest in infrastructure development.

Interdependent infrastructures: a new domain of complexity

Infrastructures are more and more interconnected, not only across national borders and continents, but also across infrastructure sectors. All infrastructures depend on telecommunication and information services; all infrastructures depend on energy services, etc. The Amsterdam power exchange, for example, was the first power exchange to be entirely conducted through the Internet. These links between infrastructures, electronically or otherwise, are a new research area that generates many questions on the vulnerability of vital infrastructures. How do we allocate the risk management responsibilities in interconnected infrastructure systems? How do we prevent failures to cascade from one infrastructure to another?

Interlinkage and interdependency between infrastructures is also created by so-called convergence phenomena. Bauer et al. (Bauer et al., 2003) distinguish four basic types of convergence:

- physical convergence (multifunctional infrastructures)
- organizational convergence (multi-utility companies)
- market convergence (in substitutes or in complements)
- spatial convergence (clustering in corridors)

The liberalization of infrastructure markets has greatly contributed to a new momentum in infrastructure innovation. The proliferation of new telecommunication and information infrastructures gives testimony of that momentum. In the world of energy infrastructures, many innovations towards distributed utility supply are underway, which may ultimately result in new network configurations, such as local networks. As a natural consequence of the market setting, the technology push mechanism has made way for market pull. The most remarkable result is the massive effort in service innovation. New metering systems, distance-meter-reading, time-of-use pricing, consolidated billing, personalized service packages, service-on-demand are examples of such innovations. The tendency of the market towards service quality differentiation, where possible, and even personalized services, entails an unprecedented dependency of infrastructure service providers on the information infrastructure. The end of this revolution is not in sight yet, as many innovative information and communication infrastructures will be entering the market in the decades to come.

Conclusion

Innovation in infrastructures occurs at all scale levels (component – node/link – network) and in all network types (physical, social, socio-technical and inter-infra). Understanding the behaviour of infrastructures as a result of these innovations can be pursued bottom-up, starting with the individual agents, or top-down, modelling the emergent behaviour of the infrastructure as a whole. In our opinion, both approaches are needed. We have no illusions about one unified model that will capture both the physical and the social complexity of infrastructures. A variety of models and disciplines will be needed to retain a realistically rich picture of these complex adaptive systems. It is also evident, that this requires a collaborative and integrative effort.

This challenge calls for a knowledge infrastructure that is equipped to cross infrastructure borders, to cross disciplinary borders, and to cross national and continental borders. In the Netherlands, the national Next Generation Infrastructures programme (www.nginfra.nl) has started to build that knowledge infrastructure, supported by a consortium of institutes, ministries and companies. The sub-programme 'Understanding Complex Networks' specifically targets questions relating to infrastructure complexity:

- What kind of models are used or should be developed to represent and model infrastructures?
- How do the constituent parts of networks interact and cooperate to make the network function?
- How can the network development be modelled so that it can be managed effectively, given the restrictions posed by the liberalized markets?
- How do interdependencies, interconnectedness and convergence affect the network development and its functioning?

The variety of innovations generated in a competitive market and the situation of distributed decision making by a multitude of actors in unbundled infrastructure markets increases the complexity of infrastructure systems. Given this complexity, steering infrastructure development so as to ensure a high reliability and quality of infrastructure bound services, now and in the future, is an ambitious goal. The disruptive effect of infrastructure malfunctioning on society and the economy, however, necessitates a better control of the development of our vital infrastructures through technological innovation, organizational change and sophisticated design of market structure and regulation.

Acknowledgement

The authors are indebted to Dr. Michel van Eeten and Prof.mr.dr. Ernst ten Heuvelhof for their valuable contributions on the social costs of infrastructure malfunctioning and the typology of infrastructure value chain unbundling, respectively. The Next Generation Infrastructures Foundation is acknowledged for the funding of this research.

References

Barabási, A.-L. (2002); Linked; Penguin, ISBN 0-452-28439-2

Bauer J.M., M.P.C. Weijnen, A.L. Turk, P.M. Herder (2003); Delineating the scope of convergence in infrastructures: new frontiers for competition, In: Thissen W.A.H., P.M. Herder (eds.), Critical Infrastructures - State of the Art in Research and Application, Kluwer Academic Publishers, Boston/Dordrecht/London, USA, pp. 209-232.

De Vries, L.J. (2004); Securing the public interest in electricity generation markets, The myths of the invisible hand and the copper plate; thesis, Delft University of Technology; available via www.tbm.tudelft.nl/webstaf/laurensv

DG TREN (2003), Analysis of the European Air Transport Industry 2001, Final Report – Summary, Contract Nr. B2-7040B-S07.17962; available online: http://europa.eu.int/comm/transport/air/rules/doc/2001_summary_report_en.pdf

Elcon (2004), The Economic Impacts of the August 2003 Blackout, Washington D.C., available online: www.elcon.org/Documents/EconomicImpactsOfAugust2003Blackout.pdf

EPRI/PIER (2002); California's Electricity Restructuring; Technical Report prepared for EPRI & PIER by Mills College & Delft University of Technology; available online: http://www.epri.com/attachments/287226_1007388.pdf

Grover, W. (2004); Fiber Cable Failure Impacts, Survivability Principles, and Measures of Survivability, available online: www.informit.com/articles/printerfriendly.asp?p=169456

ECMT (1998); European Conference of Ministers of Transport, The Spread of Congestion in Europe, Conclusions of Round Table 110, Paris, 12-13 March, http://www1.oecd.org/cem/online/conclus/rt110e.pdf

Hesselmans, A.N. (1995); De ware ingenieur; thesis, Delft University of Technology; Stichting Histosearch, Utrecht; ISBN 90-72105-04-4

Jacobs, G., A. Aeron-Thomas, A. Astrop, (2000); Estimating global road fatalities, TRL Report 445, Crowthorne, available online: www.transport-links.org/transport_links/filearea/publications/1_329_TRL445.pdf

Netlogo (2005); http://ccl.northwestern.edu/netlogo/

Primen (2001); The Cost of Power Disturbances to Industrial and Digital Economy Companies; Primen. TR-1006274 (available through EPRI)

Ten Heuvelhof, E.F., M. Kuit & H.D. Stout (2004); Innovations in Infrastructures: New solutions to increase the reliability of vital infrastructures; in: Conference Papers, TTE Conference 2004, pp.6-17; available online: www.tteconference2004.nl/pdf/TTE_A4papers.pdf

Van Mieghem, P. (2005); Robustness of Large Networks; to be published in: IEEE SMC 2005 Conference on Systems, Man and Cybernetics; Hawaii, USA, October 2005

Wolfram, S. (2002); A New Kind of Science; Wolfram Media (2002), ISBN 1-57955-008-8; available online: www.wolframscience.com

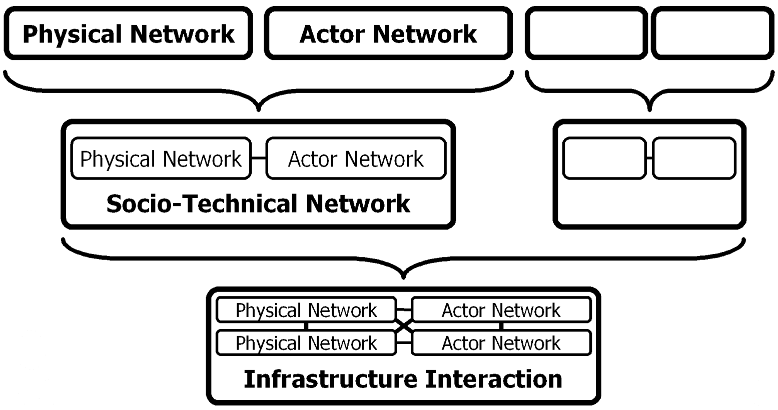

Figure 1. Domains of complexity in infrastructures.

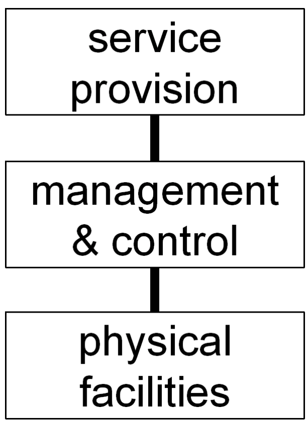

Figure 2. Layer model of infrastructure service provision.

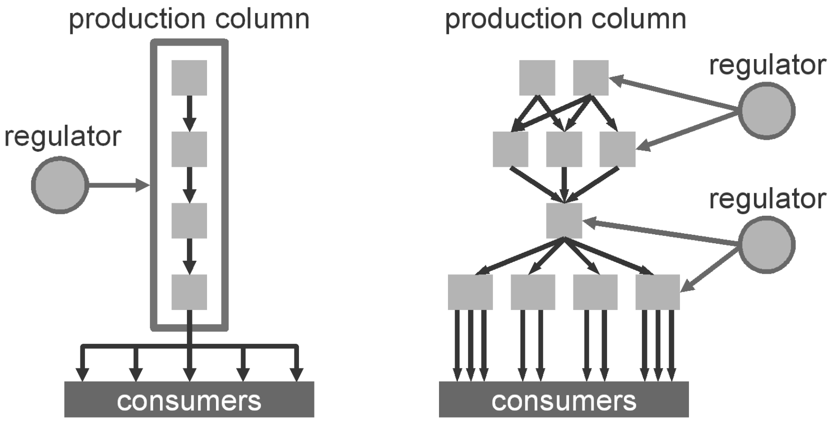

Figure 3. Model of infrastructure market before and after liberalization: vertically integrated monopoly (left) and unbundled infrastructure value chain (right), with competitive markets in non-monopoly segments. (after Ten Heuvelhof et al., 2004)