CRITICAL NETWORK INFRASTRUCTURES:

ALIGNING INSTITUTIONS AND TECHNOLOGY

Professor John Groenewegen

Technology University Delft (The Netherlands)

Professor Rolf Kunneke

Technology University Delft (The Netherlands)

Professor Claude Ménard

Centre d'Economie de la Sorbonne,
University of Paris (Panthéon-Sorbonne)

This version: MAY 2014

CRITICAL NETWORK INFRASTRUCTURES:

ALIGNING INSTITUTIONS AND TECHNOLOGY

INTRODUCTION

This book is about the alignment/misalignment between technological coordination and institutional coordination in network infrastructures, with a view at how these two dimensions interact and adapt, but also at how discrepancies or changes within as well as across these dimensions can challenge the coherence of a network and plague its performance.

1. In the dark

On 16:00 EDT on August 14, 2003, an electrical blackout considered as the largest in North American history paralyzed the city of New York and a wide corridor all the way up to Detroit (USA) and Toronto (Canada). All in all, about 50 million inhabitants were deprived from electricity and thousands of businesses shot down, with a total loss estimated to range between USD 4 billion and USD 10 billion.

The causes of the outage are now relatively well known, thanks among others to the report from the *U.S.-Canada Power System Outage Task Force* created after the event.[1] According to this report, the Cleveland-Akron area was highly vulnerable to voltage instability problems, largely because a private operator "had not conducted the long-term and operational planning studies needed to understand those vulnerabilities and their operational implications" (Task Force report, 2005: 23). It was so because the company did not fulfill the standards and practices codified by the

[1] *Final Report on the August 14th Blackout in the United States and Canada: Causes and Recommendations*. U.S. Secretary of Energy, April 2004.

electricity industry through the ***North American Electric Reliability Corporation*** (NERC)[2], a pattern that seemed to have been shared by other parties to the network. Indeed, according to the Task Force report, several operators in the Midwest consistently under-forecasted load levels the days before the blackout, with the institutional mechanisms of control (in this case delegated to the NERC) not perceiving the problem early enough and/or not responding adequately to the problem. So the cause of the outage would be 'the inadequate understanding' of how the system worked.

Following the public outcry the outage caused with much given coverage, the American Congress promptly reacted in adjusting the regulation and adopted in 2005 the *Energy Policy Act*, which delegates to NERC the responsibilities for drawing, implementing, and enforcing reliable standards on all the U.S. bulk-power system. However, this rapid adjustment may not have fully took into account the deep changes in the technology for the production and transportation of electricity and in the institutions following the important deregulation and liberalization of the sector since the creation of NERC (in 1968). Also, the new legislation might well have underestimated the impact of these changes as a source of misalignment between institutions and technology that weakened the system by challenging its coherence.

Notwithstanding its specificity, the complex combination of technological as well as institutional flaws that created the major disruption ever in the distribution of energy in North America also points out the much more general problem of the potential misalignment between these two dimensions, which is relevant to all network infrastructures and beyond.

2. What is this book about?

[2] NERC is a nonprofit organization created by the electrical industry in 1968 to coordinate norms and standards in distribution.

Indeed, the so-called 'New York blackout' (or, more appropriately, the 'Northeast blackout') provides an excellent illustration of the problem that we explore in this book, which is about the alignment (or misalignment) between technological requirements and the 'rules of the game' that shape network infrastructures and determine the successes or failures of their organization. Blackouts in electric systems, repeated train accidents, disruptions and delays in underground transportation, faults in communication systems in emergency services, provide examples of failures in the coordination between the interdependent technological requirements and institutional conditions. On a more positive side, the spectacular development of information and communication technologies and of internet illustrates the emergence of technological innovations that already reshape network infrastructures with highly demanding adaptation in regulatory institutions.

Our analysis of these interactions between institutions and technology and of the resulting successes and failures in network infrastructures is based on the concept of **criticality**. 'Critical' in this book is understood as pinpointing factors that can provide indispensable support but also obstruct or even derail the alignment between technology and institutions. On the technological side, we refer to four critical technical functions that are needed to make a network technology operational: capacity allocation in the network, mechanisms for controlling the system, interoperability among the components of the system, and interconnection between its segments. On the institutional side, we focus our attention on the critical rules implemented to allocate and monitor rights among parties involved in specific network infrastructures. The alignment or misalignment between these two dimensions, the technological one and the institutional one, depends on how transactions are organized, how their governance is implemented, and how adequately these transactions are embedded in the interaction between the technological

architecture and the rules of the game institutionally established. These different components, defined at different levels, also determine the resulting degree of coherence and performance of the system. These interactions between the two dimensions and the content of our book as developed in the coming chapters are summarized in figure 1.

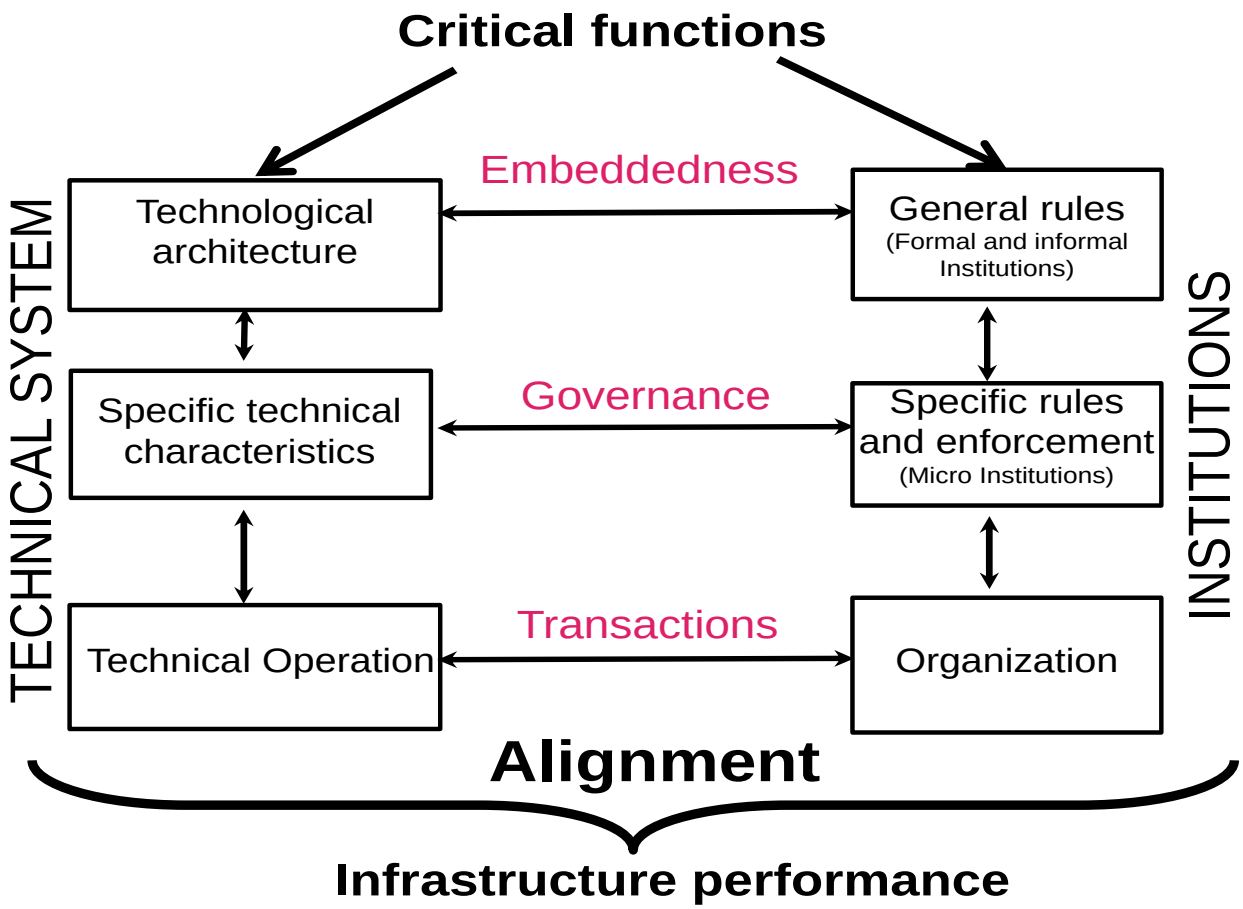

Figure 1: Alignment framework

3. Our framework.

Our driving motivation in this book can be expressed as a quest to answer the following question: *How to align the institutional and technological coordination in infrastructures in order to derive system performance that meets societal expectations?*

In the coming chapters, we investigate under which conditions alignment between institutional and technological features of infrastructures can be ensured so that their performance meets societal expectations. The challenge we face is to connect the complex engineering systems of infrastructures to their economic organization and their institutional embedding. We would like to find out what institutional features are required in support of specific infrastructure technologies and their development; and vice versa, what impact have specific technologies and their associated technical functions on the relevance of existing or planned institutional rules. The underlying assumption of our analysis is that the relation between institutions and technologies in infrastructures is not at random; it is based on regularities that we aim to unravel. In elaborating our framework and in order to better substantiate its empirical significance, we refer to examples from different network infrastructures with some emphasis on examples coming out of the energy sector, the railway sector, the water sector, and the internet.

Our analysis relies on three key hypotheses that correspond to the three layers in the organization of network infrastructures identified in figure 1: (1) the level at which the technological architecture of a network infrastructure interacts with the general rules of the game; (2) the level at which the specific technical characteristics of a network interact with the micro-institutions that define and implement specific rules of governance derived from laws, regulations, customs; (3) the level at which organizations in charge of providing the network and its services interact with the technology, making it operational.

These key hypotheses, to be substantiated later on, are the following ones.

> **H 1:** *The alignment (or misalignment) between the technological architecture and the general rules of the game depends critically on the forms and functions determining open or closed access to network infrastructures*.

Networks can be designed in a way that offers open access or limited access to the different actors potentially involved. The degree of openness is tightly related with the technological architecture of the system and the general rules framing its usage. So far, electricity networks have been built and developed through an architecture that imposes very tight constraints on how power is produced and used, which defines a limited access to the system. On the opposite, the internet owes its success to the very open characteristics of its architecture. These two examples illustrate how the different design among networks command distinct modalities of access and usages.

H 2: The general rules of the game are translated into specific rules, defined and monitored by micro-institutions, in order to meet the technical characteristics of specific networks.

By micro-institutions, we mean devices that develop at the intersection between the general rules defined within a certain institutional environment (e.g., laws established through the political system) and the actual conditions under which operators develop their activities (e.g., standards that they must endorse). These micro-institutions hold their legitimacy from the existing institutions (e.g., the role of the NERC is framed by the *Energy Act* of 2005): they act as subsidiaries. Their role is to adapt the general rules to the specific characteristics of the network at stake. For example, the regulation of the electricity network in the Canadian province of Quebec, which depends almost entirely on hydro-power and is highly centralized, differs substantially from the organization of the American network, which relies on various sources of energy and on the coordination of numerous independent companies and utilities. Similar observations could be done when comparing the organization of the electricity system in France with that of Germany. The alignment or misalignment between the micro-institutions that monitor these systems (e.g., the NERC in the U.S.) and the specific technical characteristics of the network depends critically on the governance implemented and its capacity to shape, monitor and discipline the behavior of participating entities.

H 3: The fundamental units of analysis for understanding how a specific network infrastructure works are the critical transactions that must be organized in order to make a system operational.

The operation of the critical technical functions needs to be coordinated throughout the different nodes and links that frame an infrastructure. Assuming the general rules as given and the governance of the sector guided by micro-institutions, there are transactions that must be carried out through the coordination of relevant actors in order for the network to be operational. Let us illustrate with the case of load balancing in the electricity sector. If there is for instance an unexpected failure of a power production unit, immediate action needs to be taken; otherwise large parts of the system will break down. Several technical activities need to be coordinated instantly, alternative sources of power supply need to be identified and connected to the grid, etc. The interactions between operators and the technical requirements, which is embedded in general and specific rules, depends on two aspects: (i) the technical acceptable time period to react, and (ii) the technical scope of control, since some technical functions are related to the control of the entire network whereas others might be restricted to a sub-system or even to specific units.

4. Underlying premises.

Our argument builds on the assumption that there are relevant features of network infrastructures and the associated technology that make them different from other economic activities and also that make a difference from an institutional perspective.

First, network infrastructures are socio-technical systems. Infrastructures are engineering systems that function in a specific social context. They perform intended functions, like for instance the safe and reliable provision of energy to households. Human actors purposefully design these systems;

they monitor and adjust them in order to meet expectations that encompass values. There is a close inter-relation between the technical design and functioning of the infrastructures and social interactions to safeguard and support these functions. To be sure, not everything in infrastructures is purposefully planned. Technology develops according to path dependent trajectories. The fact that the Netherlands has one of the most advanced and well-developed gas networks in the world is an important precondition for the use of sustainable biogas in residential areas, but the interest of farmers to provide biogas can certainly not be planned in every detail and the willingness of consumers to adjust their behavior or of environmentalists to accept the deep changes in agricultural practices that biogas requires cannot be pre-determined. If we want to understand and analyze what is happening in infrastructures and where this might lead us, we have to approach them as socio-technical systems.

Second, coordination is essential to network infrastructures. The different components of infrastructures are not operating in isolation. The local production of energy needs to be attuned with the available network capacity, the required energy quality, possibly storage, and of course the energy needs of the final customers. There is a necessity to coordinate these technical processes in order to produce the expected services to the final customers. Hence there is the need for a physical technical coordination as well as economic coordination of the different entities and organizations involved in the network. The technical activities in an infrastructure are complementary to each other and coordination is required for delivering expected services that have a social dimension (e.g., the 'right' to clean and safe drinkable water).

Third the technology of network infrastructures imposes technical functions that must be accommodated by rules and regulation in order to meet performance expectation. As already mentioned, four technical functions are critical in that respect: (i) System control: the system needs

to be operated according to certain technical requirements. For gas this would be a certain caloric value and chemical composition, for electricity voltage (230 volt) and frequency (60 hertz), etc. (ii) Capacity allocation: The physical capacity of the system needs to be balanced in such a way that the production capacity meets the actual demand. For the case of electricity this physical balance (called 'load balancing") is a central feature, otherwise the systems might collapse. In the railroad industry, the allocation and scheduling of slots is a determining factor. (iii) Interconnection: networks need to be connected with each other in order to improve the technical functioning and/ or guarantee the delivery of certain services. For example, the reliability of an infrastructure network may depend on the existence and quality of interconnection between the local network and the regional or national network, as in the railroad or airline industry. (iv) Interoperability: the different parts of the network need to be technically equipped in order to fit the technical needs of the system. For instance, solar panels need to fulfill certain technical requirements in order to be connected to the electricity network and signal systems on board of trains need to be adapted to differences in the network.

5. Expectations, objectives, performance

As suggested in figure 1, our book focuses on the issue of alignment (or misalignment) between technology and institutions at different levels of analysis. These levels partially connect with the layer scheme proposed by Williamson (2000: 597). However, we adapt Williamson's framework to the specific problems raised by network infrastructures and we extend it to include technology. Therefore, our goals require looking both ways: we intend to contribute filling the gap pointed out by this author (Williamson 2000: 600) regarding the neglect of the organizational and institutional dimension that plague studies on technology, but we also intend to introduce technology into transaction costs consideration. In doing so we need to explore the devices that connect

expectations and objectives to actual performance, with one guiding question in mind: how can a network infrastructure be technically reliable and institutionally feasible? Other criteria could be introduces, such as sustainability, affordability, etc. that we do not consider here.

We are not after understanding the process of technological and institutional change, but after the question of the alignment or misalignment between institutions and technologies at the three levels identified in figure 1. These levels connect to what Williamson identified as the level of institutional environment (L 2 in his terminology) and the level of governance (L 3) so that we consider other levels only marginally. Last, we do not plan to perform a process analysis, but to deliver a comparative static analysis aiming at understanding the requirements for appropriately matching institutions and technology.

Through the different layers of our framework, we explore the potential factors of alignment or misalignment that determine the capacity of a specific infrastructure to meet expectations and fulfill objectives assigned to the network. For example, in the railroad industry passengers expect trains to be on time and to provide a certain level of comfort. These expectations translate into technical requirements as well as the need for rules to be appropriately implemented and enforced in order for transactions to be accomplished efficiently. We identify three categories of performance criteria: societal, technical, and economic objectives. Societal objectives include social services obligations like universal service or sustainable infrastructures. Technical objectives are related to issues like reliability and technical safety. Typical economic objectives include efficiency, affordability, and innovativeness.

The *societal dimension* reflects the expectations of the users of infrastructures. In that respect we pay particular attention to contributions focusing on socio-technological systems (e.g., Ropohl, 1999) although we depart from the social constructivist approach (Bijker, Hughes and Pinch 1989)

as well as from the 'constrained optimization approach' (Bauer and Herder 2009). We differ from the former, which looks at technology as well as institutions as 'social constructs', in that we do not adopt a historical process perspective aiming at understanding their co-evolution (Callon 1989: 77). We differ from the latter in that we diverge with its assumption that there are 'immutable underlying regularities in the socio-technological system' and that technology determines institutional choices. For instance, if sustainability is considered a priority, then minimizing lost might command institutional choices that will frame technological ones, e.g., adopting the costly transformation of wind energy into hydrogen gas as a mean of energy storage. However, if the societal expectation is to minimize the costs of power supply, another trade-off is appropriate, like turning-off wind power and using more cost efficient fossil-fueled power plants as a backup.

The *technical performance* of a network infrastructure must be assessed through its reliability, which raises the question of whom can have access under what condition. Roughly, we can distinguish between open access and closed access systems, which determines different types and levels of investments to make the system reliable. Closed access is a situation in which only dedicated actors or agencies are entitled to monitor and control critical technical functions within an infrastructure. Typically, in that context the provision of network services is strictly regulated and primarily the task of dedicated network companies operating as centralized hubs that monitor and control the critical technical functions, thus determining the reliability of the infrastructure. At the other end of the spectrum, an open access system characterizes infrastructures that are accessible for all actors and agencies that are willing and able to contribute to its services. Imagine a household has invested heavily in decentralized energy production and would like to deliver power to his neighbors. In the backyard, cables are dogged to supply the nearby houses with electricity. However, even in this extremely decentralized option, the critical technical functions need to be

safeguarded and appropriate investments must be done, so that coordination can be ensured if the infrastructures are to remain reliable. The main point is that by contrasting these two options, we show that there will be different design and different tasks for the institutions in charge of making the rules needed for adequate coordination. In open access infrastructures coordination relies on protocols, standards, or procedures that firms or agencies have to adhere to if they want to participate. In limited access we are likely to find more 'command-and-control' types of coordination, with reliability depending on the capacity of micro-institutions to issue appropriate rules and the capacity of operators to enforce them in coherence with the technical requirements.

Last, *economic performance* depends on the capacity to match a specific mode of organization with the delivery of goods or services through transactions that are in accordance with the critical technical functions. Here we join the Williamsonian criteria of the fitness between the mode of organization and the attributes of transactions, but with a qualification: these attributes, particularly uncertainty and the type of assets required are embedded in and depending on the technical characteristics at stake. This critical interaction between technology and institutions reinforce the transaction cost economics perspective, which argues that costs-benefits analyses must go beyond considerations limited to the production function. Organizational considerations become essential for assessing the performance of the infrastructure. Allocating rights to a specific mode of organization (a private firm? A network of private firms? A publicly owned operator? A local cooperative?) is a strategic issue that determines not only who is going to perform the critical technical functions, but the level of acceptable transaction costs if the system is to be reliable. In the energy sector we observe that different arrangements have been adopted in different countries; and in some cases they even co-exist within the same country, with the same

institutional environment. This diversity raises the questions of the nature, efficiency, and costs of alternative mechanisms of coordination in network infrastructures.

6. Outline of the book.

Our book does not intend to deliver answers to all the problems already pointed out. We rather focus on the development of our framework as a way to provide tools for analyzing these issues. Our arguments are exposed in five chapters.

In chapter one (*Aligning Institutions and Technologies*), we unfold our theoretical framework through a discussion of the key concepts we are using and the different layers we are distinguishing. We submit that with respect to infrastructure networks, the allocation of property rights and decision rights shapes the institutional dimension while fundamental technical functions to be met frame the technological architecture and how it works.

Chapter two (*Infrastructures features and characteristics*) comes back to the nature of network infrastructures in order to identify which characteristics matter and why and to discuss in what sense they can be understood as socio-technical systems. In doing so, we depart from the traditional approach to network industries that focuses on the role of economies of scale and scope and the resulting externalities, which in our view put too much emphasis on technological determinants, thus neglecting the social goals involved, how they are entrenched into institutions, and how critically they connect with the technological architecture.

Chapter three (*The critical access issue*) focuses on the upper level of our framework as summarized in figure 1. It pays special attention to the conditions of access to network infrastructures and how they connect to reliability, which poses the problem of the links between the general rules embedded in formal as well as less formal institutions and the technological architecture of the networks. In this chapter, we put some emphasis on the political economy at work behind these links and we explore the possible relations this critical access could have with the broader conception of open versus limited access societies.

In chapter four (*The critical transaction issue*), we turn our attention to the alignment or misalignment between the technical operation of the critical functions characterizing the technology of an infrastructure network and the way transactions needed to ensure these functions are organized. We do so by developing the concept of critical transactions, understood as a subset of the more general concept of transaction, with a particular attention to the scope of control and the speed of adjustment the organization of these transactions requires for being efficient.

Chapter five (*Critical governance*) develops the concept of micro-institutions, which are organizational arrangements implemented to transform general rules such as laws and regulatory guidelines into specific ones, with also the possibility of being responsible for their implementation. We understand these micro-institutions as governance devices monitoring and aligning specific rules with technical functions specific to the network infrastructure at stake and its specific institutional environment. We argue that this intermediation between the general rules framing the domain of action of a network and the operators in charge of producing and delivering the services is a key factor in determining the technical performance of the related infrastructure.

In our *conclusion,* we provide an overview of what we have learned as well as indications about some key issues that would require further development, e.g. taking into account more in depth the strategic behavior of actors or the factors of change in the alignment between institutions and technology. We argue that it is a bilateral interaction, so that technology is not the sole driver of change: there are situations when institutional change gives an impulse to innovation, either positively by providing adequate support or through the new constraints it imposes.

CRITICAL NETWORK INFRASTRUCTURES:

ALIGNING INSTITUTIONS AND TECHNOLOGY

TABLE OF CONTENT