Framework for path finding in multi-layer transport networks

Dijkstra, F.

Publication date
2009

Document Version
Final published version

Chapter 1

Introduction

1.1 Computer Networks

Communication networks are ubiquitous in our society: we use them to make phone calls, send e-mails, and browse the web. In all cases an underlying physical infrastructure of wires, fibres, switches and routers provides network services for applications on the network.

Network services can be classified as circuit switched and packet switched network services. The public switched telephone service (PSTN), also known as plain old telephone service (POTS), is a circuit based switching technology. The Internet on the other hand is largely based on packet switched technology. Both technologies have their merits. Packet switched networks are very robust against failures, but can not guarantee a certain quality. Packets switched networks are like highways where all cars use the same roads, and may end up in traffic jams. The analogy for a circuit switched networks would be a dedicated highway between certain origins and destinations. Cars on these dedicated highways may not drive faster, but it is guaranteed that they do not encounter traffic jams.

The work described in this thesis focuses on networks for applications that require better quality of services (QoS) than the regular Internet can offer. Such applications may require so much bandwidth that if they use the regular Internet, it causes congestion and they fail to run smoothly or they disrupt other Internet traffic. These applications require dedicated network connections, such as the circuits in the above analogy. This work masters and manages the complexity of the networks that are offered to these demanding applications.

1.2 e-Science Applications

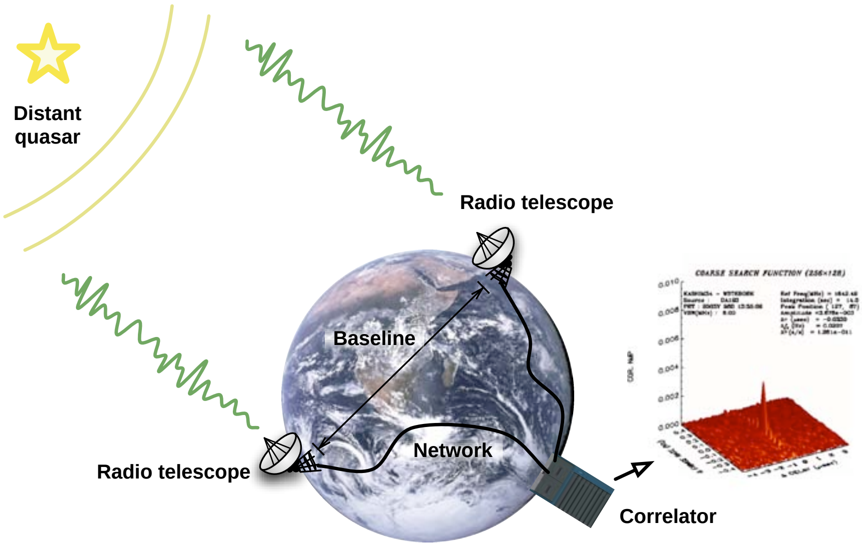

Figure 1.1: *Schematic diagram of very large baseline interferometry (VLBI).*

One of the applications that requires a better service than the regular Internet can offer is very large baseline interferometry (VLBI). Two or more far apart radio telescopes pick up signals from the sky, as shown in figure 1.1. The received data is sent to a correlator for processing. The resolution of the correlated signal improves with the distance between the telescopes. Ideally, the telescopes are located on different continents.

Historically, the data is shipped on tape from the telescope to the correlator. Experiments in 2004 have shown that the data can be transmitted over networks [p38, p42, p34]. Transmitting the data in real time requires a bandwidth of 1 to 10 Gb/s. Since the raw measured signal is nearly white noise, it can not be compressed.

VLBI with data transport over a network (e-VLBI) is an example of an *e-science application*, a scientific application that heavily relies on computer networks [s54]. Sending the data over the regular Internet is not always possible for e-science applications [p23, a11]. This means that the data must be sent over a dedicated network connection.

Typical observation times for telescopes are in the order of hours, and different experiments may link different telescopes together. It is undesirable to change the hardware to reconfigure the network topology for each experiment. Ideally, radio astronomers create a dedicated network connection in software for each experiment with the same ease as it takes to establish a telephone connection.

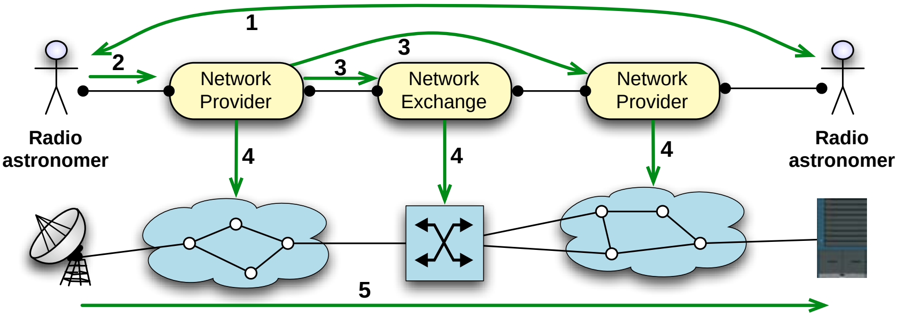

Figure 1.2: *Steps required to set up a network connection between two radio astronomers.*

Figure 1.2 shows the steps that needs to taken to establish a network connection between a radio telescope and a correlator.

The first step is to decide on the end-points, and the characteristics of the required network connection (e.g. the amount of bandwidth required). Secondly, the astronomer makes a request to his or her network provider. The network provider then must find a path that falls within the specified parameters. If the source and destination of the required connections fall in different administrative network domains, the involved domains must collaborate in finding a path. After a valid path is found, the fourth step is to provision this path in the network. Finally, the astronomer can use the network connection to transport his or her data from the telescope to the correlator.

The third step in this process, finding a valid path from source to destination, is covered in this thesis.

1.3 Hybrid Networking

The idea of providing e-science applications with deterministic point-to-point connections was fostered by a community of research networks, later organised in the Global Lambda Integrated Facility (GLIF) [u8]. The ideas in this community led to the concept of *hybrid networking*, the offering of both packet switched (IP) services as well as circuit switched connections over the same physical network infrastructure, the transport networks.

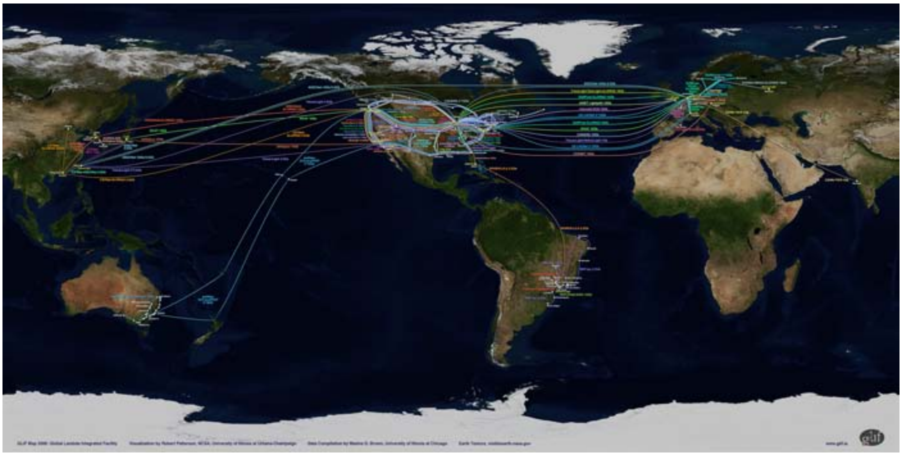

Figure 1.3: *GLIF world map of May 2008, with all network connection offered by NRENs participating in the GLIF. Source: Patterson, Brown [u4].*

Since most e-science applications operate in a large-scale environment, with collaborators at different universities, the networks required for these applications are nearly always multi-domain networks. De Laat estimated in 2000 that a typical network connection for a physics experiment crosses seven domains [p22]. To achieve interdomain operation, the different networks have to collaborate. For dedicated network connections, this collaboration is done in the GLIF community. In a few years time a few dozen international network connections have been established to provide the interdomain connectivity. Figure 1.3 shows a collection of the interconnections provided by partners in the GLIF community as of May 2008.

1.4 Research Overview

The concept of circuit switched networking potentially puts users and applications in the driver seat. Grid applications already treat computing and storage as dynamic resources, and they may want to treat the network as a dynamic and manageable resource.

The paradigm shift to offer dedicated network services to specific applications has lead to many new questions, including questions on usage models, implementation models, manageability and interoperability of multi-layer networks. Common models, shared by all parties are especially important, given that applications must now interface with the network, and network connections cross multiple domains. All parties involved must somehow share information and act accordingly.

Rather than reinventing the wheel, network engineers turn to existing solutions to model and manage hybrid networks and their circuits.

This brings us to the main research question in this field: **Is there a fundamental difference between hybrid networks and the Internet or the telephony network? Can existing models and algorithms be reused or should new models and algorithms be developed?** Section 1.5.2 shows which part of this research is covered in this thesis.

The final goal of this research is to fully automate the use of dedicated network connections, as sketched in the radio astronomy example of section 1.2.

The GLIF community provides useful input to break down the research question in smaller questions. In particular the issues analysis provided by Bos *et al.* [t6] gives an excellent overview of the problems encountered in the field.

In addition, we can draw from our own experience in setting up and using dedicated network connects across the globe [p13]. The Dutch national supercomputing centre SARA [u13] also provided us with insights [p29]. From these experiences, we can categorise these problems by simply looking at the steps that are typically taken when the user or an e-science application wants to get a dedicated network connection:

1. The user must formulate the requirements, including the end points and the network characteristics like bandwidth, latency, jitter, minimum packet size (if applicable), reliability, etc.

2. These requirements must be communicated to their upstream network provider, usually the national research and education network (NREN).

3. The network provider must gather information about available resources, including the resources in other networks, as the two end-points are typically in different networks (the multi-domain aspect of the question).

4. The network provider must, in collaboration with the other network owners, determine a valid path that uses available resources, and is within the specs of the user.

5. The resources needed for the path must be reserved for all involved networks.

6. The reserved resources must be configured in the networks.

7. The end-to-end path must be tested, and in case of faults the faults must be examined and resolved.

8. The network provider informs the user, and the user must configure the end nodes (e.g. configure the IP addresses and set the routing table).

9. The user runs the applications.

This short overview is not complete and makes a few assumptions. First of all, it assumes that the network provider of the user takes a central role in orchestrating the available network resources across domains. This is called the *master contractor* role, and is similar to the role of an agent who acts on behalf of the user [t11, s14]. Secondly, this process only deals with the provisioning (set up), not with the reconfiguration and deprovisioning (tear down) or the service monitoring after it is in service. For instance, it does not deal with protection, recovery and restoration. Finally, this overview is focused at the administrative control of the network, not at technological developments of the data transport itself.

We can organise the above steps in three categories:

Architectural work, including the design of the network, modelling of the data plane and control plane, and network descriptions to share the information between each entity. This is required before a user can even formulate a question for a network connection.

Path configuration, the actual path finding and provisioning of the network connections. This includes the process of debugging, and monitoring the state of the network and sharing that state with neighbours as input for a scheduling mechanism or for debugging purposes.

Application usage, the use of a network connection by an application. This includes the transport protocols, node addressing and co-allocation of different kind of resources (network, CPU, storage, etc.)

Figure 1.4 gives a schematic overview of the different topics in our field of research. Solid arrows represent dependencies between the topics. The boxes represent topics and subdivisions of a larger topic in subtopics.

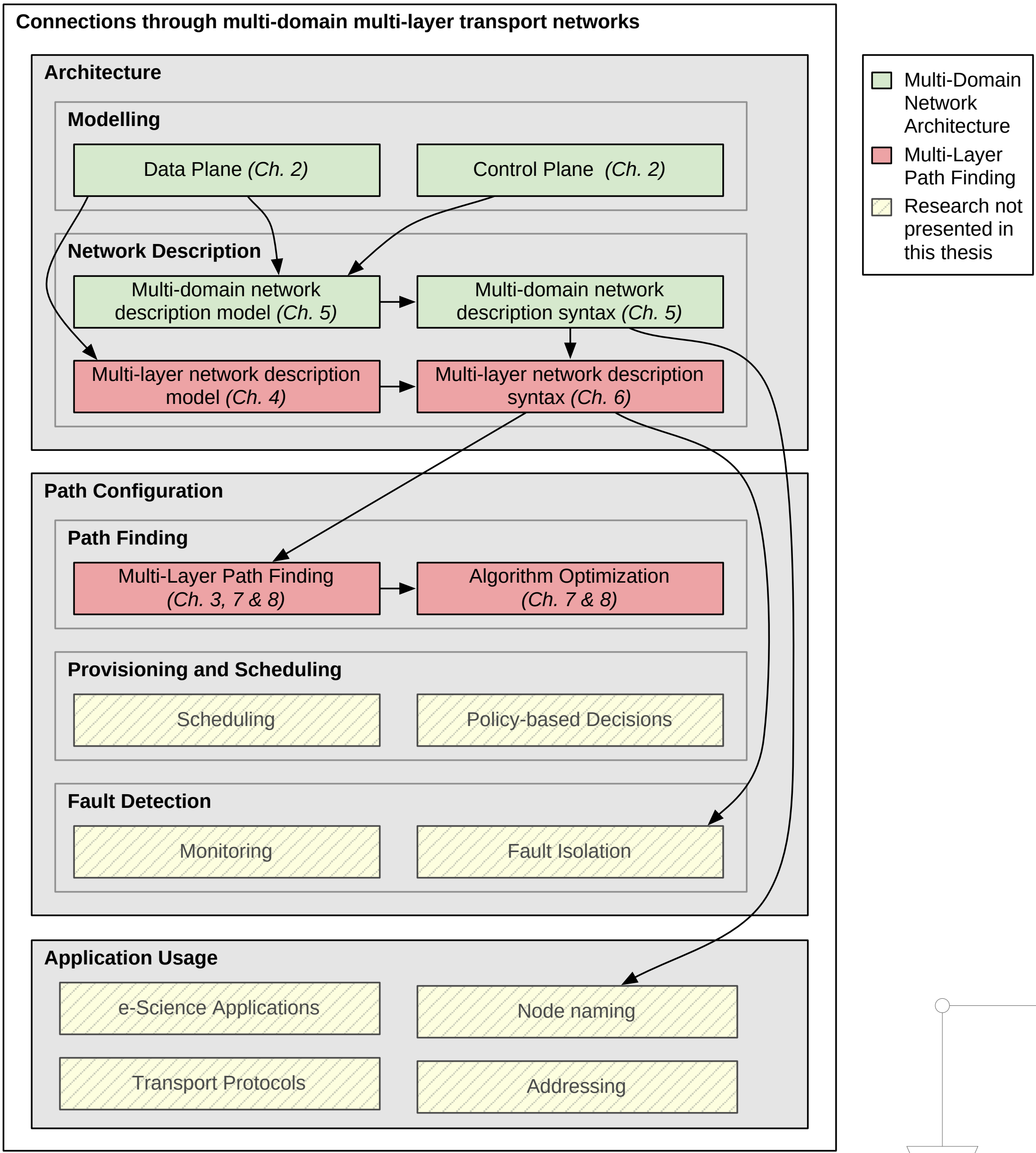

Figure 1.4: *Topics covered in this manuscript compared to the research field, and dependency relations between the various topics.* Ch. *denotes the chapters.*

1.5 Thesis Overview

1.5.1 Papers and Topics Covered

Figure 1.4 not only gives an overview of the research field as a whole, but also shows the contribution of this thesis to the field.

The colours in figure 1.4 denote the sections in this manuscript:

Slanted Light Yellow boxes represent work not covered in this thesis.

Light Green boxes represent topics on *Multi-domain Network Architecture*. This part contains mostly the architectural work.

Dark Red boxes represent topics on *Multi-layer Path Finding*. This part includes path finding work, as well as the architectural work that is related to creating model for path finding in multi-layer networks.

The reason I focus myself to architectural problems and multi-layer path finding is twofold. First of all, there has been no common architectural model used by the different researchers and network providers, and this has hampered interoperability across domains, even though this interoperability is required for most applications. Current tools for provisioning circuit switched networks found out the hard way [t13], and only recently is there a move towards standardisation and interoperability in communities like the GLIF, OGF and GÉANT [u2, t17, t9]. During the course of our research, we were surprised to find that there was only very little research on modelling multi-layer computer networks. The only model in broad use is graph theory, but it turns out that this can only be used for single-layer computer networks.

The second motivation is that there currently are not many multi-layer path finding algorithms, so there is no baseline to begin with. Rather than trying a heuristic approach (which may be the best solution in the long run), we attempted to establish a baseline by first giving an exact algorithm.

Table 1.1 gives references to all my co-authored papers, including topics not covered in this thesis. In addition, this table lists which articles have been at the basis of each chapter.

The work in chapter 5 has been done in collaboration with Van der Ham [a3, a9, p16] and the work in chapter 7 in collaboration with Kuipers [a12].

1.5.2 Research Question

The goal of this thesis is to answer the following research question:

Is it possible to use the same path finding algorithms in multi-layer transport networks as those in use for the Internet and telephony networks? If not, what kind of algorithm is required?

This question will be covered in detail in chapter 3.

Architecture	Papers	Thesis Chapter
Modelling		
Data Plane	[a1]	Chapter 2
Control Plane	[a4]	Chapter 2
Network Description		
Multi-domain network description model	[a3]	Chapter 5
Multi-domain network description syntax	[a9]	Chapter 5
Multi-layer network description model	[a10]	Chapter 4
Multi-layer network description syntax	-	Chapter 6
Path Setup		
Path Finding		
Multi-Layer Path Finding	[a12, a13, o8]	Chapters 3, 7 and 8
Algorithm Optimization	[a12, o8]	Chapter 7 and 8
Provisioning		
Scheduling	-	-
Policy-based Decisions	[o3, o6]	-
Signalling	-	-
Fault Detection		
Monitoring	-	-
Fault Isolation	-	-
Application Usage		
e-Science applications	[a11]	-
Transport protocols	[o5, o4]	-
Node naming	-	Chapter 5
Addressing	[o7]	-

Table 1.1: *Relation between research topics, published papers and chapters in this thesis.*

1.5.3 Methodology

The goal of our work is to find fundamental differences between the architecture of circuit switched networks and other network models, such as graphs or models in use for the Internet, and telephony network. This means that we first must make a model to describe circuit switched networks.

The first step we take in this modelling is to define a concise terminology. If possible, we re-use terminology. If that fails, we turn to network standards. We validate our work by comparing our model with practical applications and actual networks. In particular, we use the networks in the GLIF community to validate our models and principles. If our model is not consistent with the use in the GLIF community, we modify our model until it is.

The creation of a model is partly a logical deduction from relations between terminology, and partly an engineering craft. The scientific value lays in the understanding of the subject at hand – in this case a certain type of computer networks. To make the scientific output more clear we make series of claims closely related to the research question defined in section 1.5.2.

We will prove our claims using any of the standard logical means available to use. For example, a negative claim can be proven using a counter example. Positive claims are harder to prove. In general, we prove those using an implementation that shows it can be done. The risk of such proof is that we inadvertently only prove a special case, rather than the general case. We solve that by using strict boundary conditions.

1.5.4 Chapter Outline

The first chapter of this thesis describes the larger context of the transport networks, and gives the basic model for these networks.

Chapter 2, Optical Exchanges looks at the design of optical exchanges and the difference between Internet exchanges. This chapter makes two claims: **Exchanges can only be ignored during pathfinding if the connections through an exchange are modelled as direct connections, and the exchange does not define a usage policy on its own**. This chapter also defines the terminology used later in this thesis.

The main body of this thesis (chapters 3 to 8) builds upon the observation in chapter 2 that transport networks are inherently multi-layer networks.

Chapter 3, Going in Loops: Path Finding in Multi-Layer Networks introduces the problem of path finding in multi-layer networks, and makes the claim: **Link-constrained algorithms are not sufficient for path finding in multi-layer networks, if links are 1:1 mapped to edges**. This is interesting, since most solutions use such a heuristic link-constrained algorithm. This claim is proven by a counter example. In addition, this chapter claims that for all practical purposes, **graphs can not be used for path finding in multi-layer networks**, even though it is very common to use graphs to describe single layer networks.

Chapter 3 also makes the claim that **Path-constrained algorithms *are* sufficient for path finding in multi-layer networks**. This is proven by an implementation, which is covered in chapters 4, 6, 7 and 8.

Chapter 4, Multi-Layer Network Model introduces a model to describe multi-layer networks, based on ITU-T recommendation G.805 and the label concept in Generalized Multi-Protocol Label Switching (GMPLS). It thus provides an alternative to graphs, which can not be used for multi-layer networks.

Chapter 5, Network Description Language builds upon the terminology in chapter 2 and forms the basis for the research presented in chapter 6. This chapter claims that **it is possible to create a distributed network description, without a central repository**. It shows this by building on the semantic web technology. En passant, this work solves the issue of naming of end-nodes by using uniform resource identifiers (URIs). This work is a collaboration between Jeroen van der Ham and myself.

Chapter 6, Multi-Layer Network Description Syntax describes a syntax and implementation of the model described in chapter 4, and examines if the model is sufficient to describe all technologies that are in use in the networks within the GLIF community.

Chapter 7, Path Finding Algorithms introduces two path-constrained algorithms that can be used to find path in multi-layer networks.

Chapter 8, Path Finding Implementation shows how one of the algorithms of chapter 7, applied to any network with the technologies of chapter 6 indeed is capable of finding paths through a multi-layer network, proving the claim that **Path-constrained algorithms *are* sufficient for path finding in multi-layer networks**.

Finally, this thesis discusses and summarises the results in a concluding chapter.

Chapter 9, Discussion and Conclusion summarises the claims and research question in the individual chapters and concludes this thesis.