



## UvA-DARE (Digital Academic Repository)

### A semantic model for complex computer networks : the network description language

van der Ham, J.J.

**Publication date**  
2010

[Link to publication](#)

#### **Citation for published version (APA):**

van der Ham, J. J. (2010). *A semantic model for complex computer networks : the network description language*. [Thesis, fully internal, Universiteit van Amsterdam].

#### **General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

#### **Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

# Chapter 1

## Introduction

### 1.1 Computer Networks

Communication over computer networks is a very important part of our society today: we make phone calls, send emails, and surf the web. All these processes are enabled by the physical infrastructure of wires, and fibers in the ground, combined with networking devices that communicate electronically over these cables.

There are two different types of network services: packet and circuit switched network services. To describe the difference we can make an analogy to traffic. Packet switched networks are like regular highways that everyone can use, and may encounter traffic jams. Circuit switched networks on the other hand are more like dedicated highways between certain origins and destinations. Cars on these dedicated highways may not drive faster, but it is guaranteed that they do not encounter congestion.

The public switched telephone service is an example of 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 constant quality of service.

More and more scientific research applications require better quality of services than the regular packet-switched Internet can offer. Such applications may produce so much traffic that if they use the regular Internet, they cause conges-

tion, fail to run smoothly and disrupt other Internet traffic. These applications require dedicated network connections, such as the circuits in the above analogy.

Creating connections in a circuit-switched networks, and especially optical networks, is a complex task. A complete end to end circuit must be created before the connection can be used. Provisioning such a circuit requires knowledge of the network topology, capacities and capabilities.

This thesis describes the Network Description Language (NDL), an ontology for describing complex network topologies and technologies. The language defines a clear terminology for describing network topologies that can be linked to other descriptions, to other kinds of resources, but also to other networks, creating a distributed interoperable description of the global topology.

Network operators tend to be protective of detailed topology information, because of scalability, security, or policy reasons. It is also possible to share an aggregated view of the topology, so that only the most important details are published. In part two of this thesis we examine what kind of effect aggregation has on inter-domain pathfinding.

## 1.2 e-Science Applications

Scientific research has grown in step with faster computers and more ubiquitous computer networks. In eScience it is now common to see experiments with very large data-sets or requiring high speed or large amounts of bandwidth[8]. As stated above, if these transmissions use the regular Internet, they can disrupt other traffic. Some experiments do not require large bandwidth, but require very controlled jitter and delay on the data transmissions. Below we provide some examples where scientists use dedicated connections in optical networks, *lightpaths*, for their experiments.

For example *lightpaths* are used by scientists to perform large-scale screening for lung cancer[9]. Large, high resolution radiological images are transmitted to a central repository. An expert can then access the repository and quickly review the new and archived images to come to a diagnosis.

Besides sharing data sets, scientists can also use the network to share specialized equipment such as electron microscopes[10]. With high-speed optical networks such a microscope can be connected to the network, and scientists can remotely control the microscope and view the results real-time.

An extreme example of a specialized instrument is the Large Hadron Collider (LHC) at CERN. Once operating, the experiments in this particle accelerator will generate over 300 GByte/s of data, which is filtered locally to about

300 MByte/s. This data is then immediately globally distributed to 10 Tier 1 institutes through dedicated optical connections. These institutes distribute the data further to Tier 2 institutes. Together these Tier 1 and 2 institutes form the Worldwide LHC Computing Grid[11, 12] where the data of the experiments is stored and analysed.

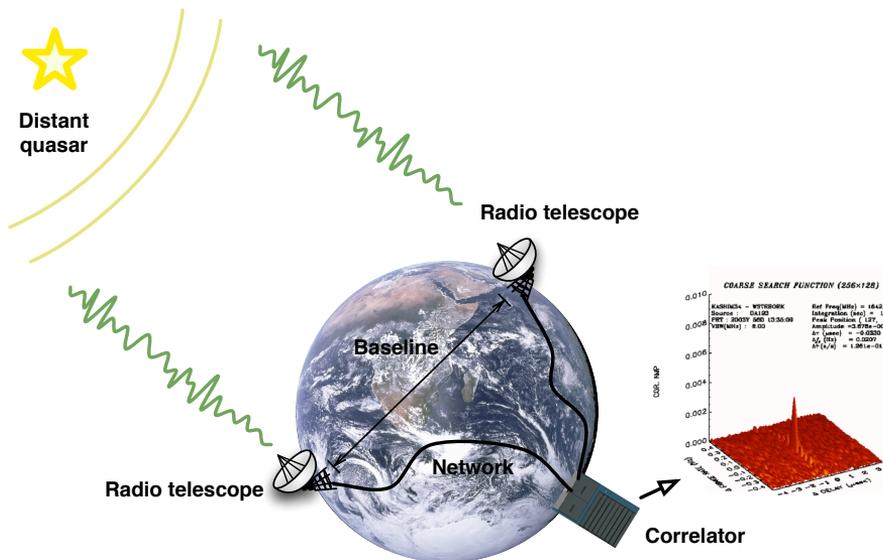


Figure 1.1: Schematic diagram of e-VLBI (Image courtesy of F. Dijkstra[13])

Another e-science application that requires dedicated network connections is Very Long Baseline Interferometry (e-VLBI). Two or more radio telescopes, that are far apart, pick up signals from the sky, as shown in picture 1.1. The received data is directly 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, data is shipped on tape from the telescopes to the correlator. Experiments in 2004 have shown that the data can be transmitted over networks [14, 15, 16]. Transmitting this data in real time requires a bandwidth of 1 to 10 Gbit/s. Since the raw measured signal is nearly white noise, it can not be compressed. Typical observation times for telescopes are in the order of several hours. Depending on what astronomical source is being observed, e-VLBI observations use different sets of telescopes.

The StarPlane project[17] provides researchers a dynamic network topology for computing. The StarPlane network connects together the five different clusters of the DAS-3[18] distributed supercomputer. This network is a dedicated part of the SURFnet6 network. The topology interconnecting the clusters can be dynamically changed, optimising the configuration for the demands of the application being run on the distributed supercomputer.

Dynamic high-speed connections also provide a way to quickly migrate virtual machines to other locations[19]. At SuperComputing 2007 it has been shown that virtual machines can quickly be migrated between different sites using lightpaths with minimal downtime of the virtual machine: in the order of 1–2 seconds. This migration can for example be used to migrate the computation to the data set instead of vice versa, or to conserve energy or carbon footprint.

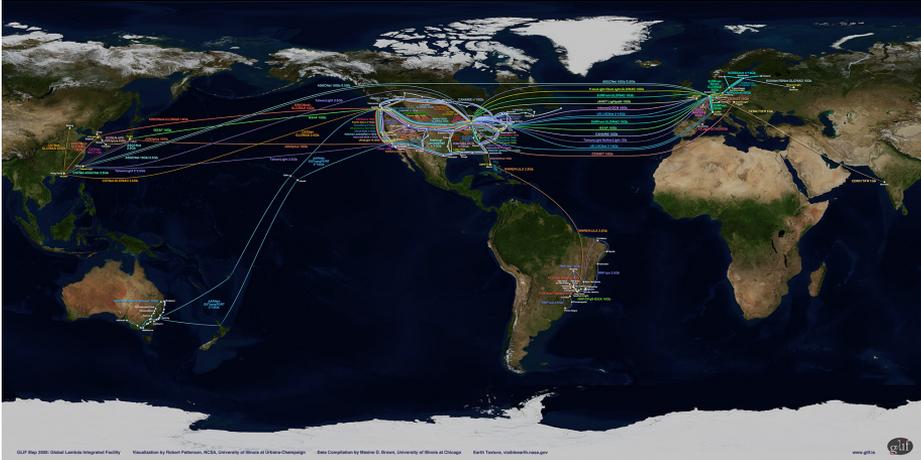
In reality creating such a dedicated network connection is a long process with multiple parties involved. The scientist, his provider network, possible intermediate networks and the destination network all have to agree on the characteristics and details of the connection. The connection is then manually configured by experienced operators in each of the networks. The characteristics and progress on the configuration are then communicated through phone and email. We examine these steps more closely in section 1.6.

The process of creating a network connection can be improved by having a clear and well-defined description of the network. This allows all the parties involved to clearly express their requests and intentions. Having such a language that can also be processed by applications is a first step towards automating this procedure.

### 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)[20]. This community provides a global network to support data-intensive scientific research, and also supports middleware development for optical networking. The ideas in this community led to the concept of *hybrid networking*, the offering of packet switched (IP) services and circuit switched connections over the same physical network infrastructure[21].

Since most e-science applications operate in a large scale environment, with collaborators at different universities, the networks required for these applica-



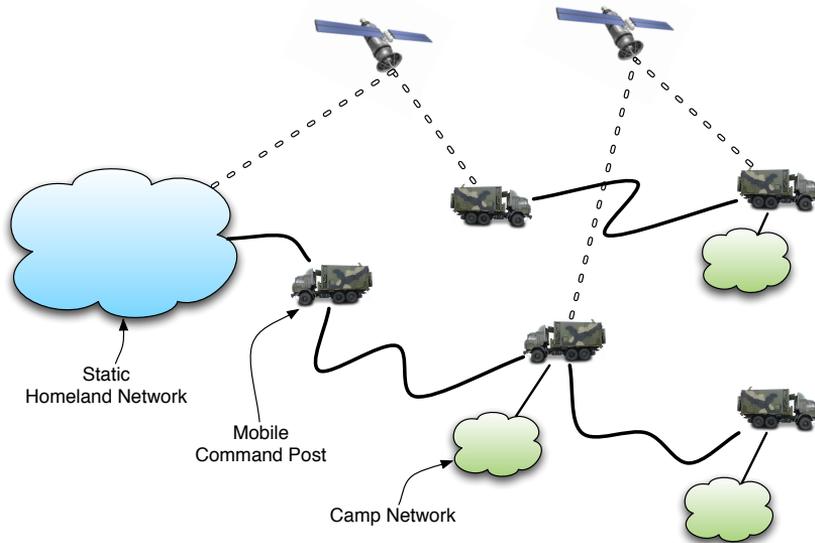
**Figure 1.2:** *GLIF world map of May 2008, with all network connections offered by its participants. Source: Patterson, Brown [22].*

tions are nearly always multi-domain networks. De Laat estimated in 2000 that a typical network connection for a physics experiment crosses seven domains [23]. To achieve inter-domain operation, the different networks have to collaborate. For dedicated network connections, this collaboration is done in the GLIF community. In few years time a number of international network connections have been established to provide the inter-domain connectivity. Figure 1.2 shows a collection of the interconnections provided by partners in the GLIF community as of May 2008.

The GLIF community is working hard at improving the lightpath provisioning process by exchanging experiences, documenting processes and developing middleware.

## 1.4 Military Networks

A large part of the research described in this thesis was performed at TNO, the Dutch Research Laboratory[24], where we also observed a similar complexity in the management of military networks. The military is moving towards more and heavier use of computer networks through network enabled capabilities (NEC)[25]. This way of operating intends to enhance military effect through



**Figure 1.3:** *The abstract architecture of the TITAAN network*

better use of information systems. The data from these information systems must be distributed in order to create a shared awareness among the participants in the operations. The goal is to have the right information at the right place at the right time.

The communication network is a key component in achieving the goal of true network enabled capabilities. It is important that all the actors involved have a clear understanding of the capabilities and state of the network. The network can then be used as optimally as possible, given the current situation and possibilities.

An example of such a military network is TITAAN[26], the Theatre Independent Tactical Army & Airforce Network. It is a robust network, configured in such a way that it can be packed up and deployed quickly, anywhere in the world. The network can then be used to exchange data, email, telephone and video-conferencing.

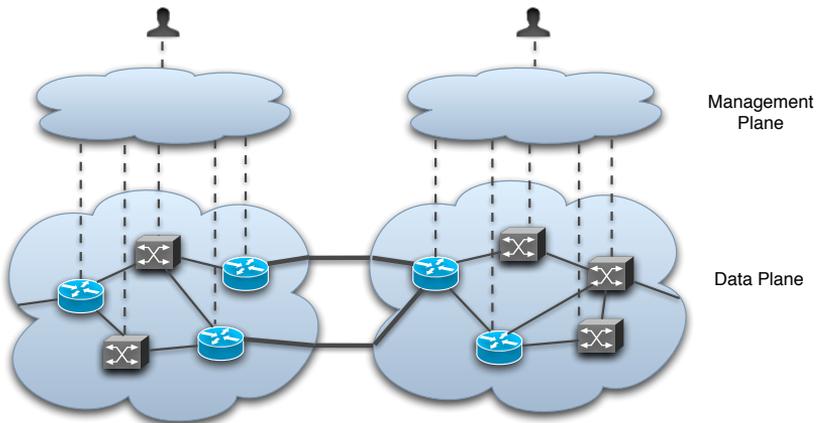
Figure 1.3 shows a high-level abstract view of the TITAAN network. On the left side is the fixed tactical network in the Netherlands. In the middle and right

there are mobile command-posts, or bases. Each has a local network. These networks are connected either through radio, satellite, or fixed links.

Currently the TITAAN network operates independently, but the global trend is that military operations are multi-national operations. Combined with the trend towards more network enabled capabilities, we see that nations are trying to create tighter couplings between their networks. In order to make full use of these networks specific information about the network topologies and its properties must be exchanged.

## 1.5 Management of Computer Networks

Before we present the research overview, we first describe the current architecture for management of computer networks. Figure 1.4 shows a schematic overview of this architecture.



**Figure 1.4:** *The management plane (top) and the data plane (bottom).*

At the bottom is the *data plane*. This is the physical network over which the signals are sent using data communication protocols, electric or optical pulses. The data plane takes care of moving the data from source to destination. Examples of data plane implementations are Ethernet, TCP/IP, or fiber infrastructure.

The *control plane* is used by the network to manage the topology of the data plane. The routers and switches in a network communicate over the control

plane with routing and switching protocols in order to get an overview of the data plane topology. The control plane is not shown in the figure, as it can either be implemented as a small dedicated part of the data plane (in-band), or as an independent network (out-of-band). Examples of control plane protocols are the Spanning Tree protocol (STP) and the Open Shortest Path First (OSPF) protocol, which we will discuss in more detail later.

Finally there is the *management plane*, which is used by the network engineers and operators to manage and monitor the network. The management plane can use the control plane network, or a separate network. Examples of management plane protocols are Simple Network Management Protocol (SNMP) and NetConf. Vendors of networking equipment often also have separate management software which is used to monitor and manage the network. Examples of these applications are HP Openview[27] or Nortels DRAC[28].

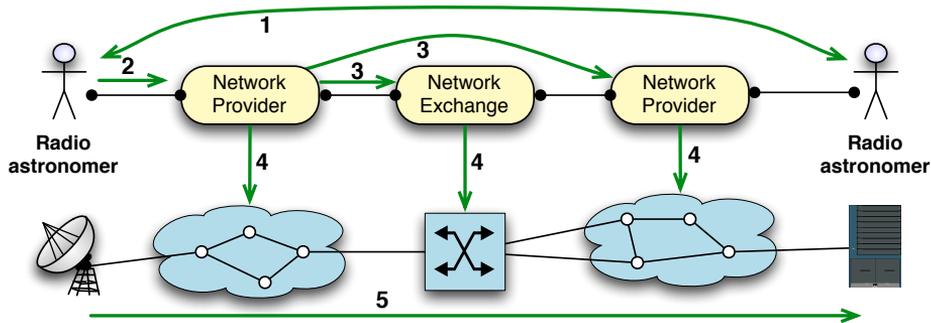
Note that the management planes of different networks are shown separated in the figure above. Detailed topology or management data of different networks is not shared between them. Some limited information exchange between networks happen so far as it is required for the operation of the network, for example the announcement of prefixes and connectivity to other Autonomous Systems in case of the Border Gateway Protocol (BGP)[29].

## 1.6 Research Overview

The goal of our research is to provide a first step towards automatic pathfinding and provisioning of inter-domain lightpaths. Figure 1.5 shows the steps that currently need to be taken to establish a network connection for any e-science application, in this example between a radio telescope and a correlator. If we examine this procedure in more detail, we see that it is broken up in the following underlying steps:

1. The user formulates 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, in our case the national research and education network (NREN).

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.



**Figure 1.5:** Steps required to set up a network connection between a radio-telescope and correlator.

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

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

4. Once the reservations are all confirmed, the reserved resources must be configured in the networks.

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

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).

5. The user runs the applications.

Currently, this whole process of acquiring a (working) lightpath across multiple domains can take several weeks, a lot of emails and phone calls and extensive testing. It is clear that the whole process needs to be improved and automated in order to scale.

The first step is where the user formulates his requirements. Users can have very different use-cases for lightpaths. For example, a user who wants to do live video streaming has very different requirements than a user who wants to transfer a very large data set. Often users do not know which basic settings they need, or are not able to communicate these clearly to the operator.

Sobieski and Lehman have proposed ‘Common Service Definitions’ [30], a set of common services with associated values for common parameters for lightpath performance. These definitions make it easier for users to pick the right values. The requirements are also clear to the network operators, allowing them to provide a measurable and reproducible service.

Fault detection and isolation are processes that often occur when operators are trying to bring up a new lightpath. Pure optical sections are hard for fault isolation, it is easy to detect that no light is coming through, but very hard to isolate the section where the light is stopped. Of course faults can also occur during normal operation. Good communication between domains is required for being able to detect and isolate problems. The GLIF community is currently working on this problem[31] by introducing inter-domain lightpath identifiers to facilitate communication between domains[32].

One of our students has studied the problem of fault-detection and isolation in optical networks[33]. This study provides an analyses of the problem and proposes an expert system that can help users and administrators to detect and isolate problems.

If we examine the intermediate steps taken by the network operators, 2–4 in Figure 1.5, then it becomes clear that this involves a lot of communication between the operators. In order to determine a path, they have to exchange topology and capability information. Once a path has been determined an operator must communicate the specifics to the other operators involved. This process is hampered by the lack of an interoperable way of describing and exchanging network topologies.

The first research question of this thesis is the following: *Is it possible to create a distributed information model for the description of topologies and technologies for inter-domain pathfinding?*

When answering the above research question, I assume complete openness about network topologies. The goal is to provide an information model that is as complete as possible.

However, network operators do not always want to provide a full description of their topology, for example because of scalability, or security reasons. A solution for this is to apply topology aggregation, publishing only an aggregated view of the network topology of a domain. However, leaving out information means that pathfinding will not always find the optimal path.

It should be noted that in some cases aggregation of graphs can be reversed, by using directed queries. The original graph can then be inferred from the aggregated topology and the responses. However this is hard to do because the original graph changes due to provisioned connections, and the management

system of the network could also detect this behavior and either stop answering or provide misleading answers.

When requesting a path based on an aggregated topology, it is also possible that a path is found in the aggregated graph, which turns out not to be available in the actual network. With that in mind, in the second part of this thesis, I answer a second research question: *What impact does topology aggregation have on inter-domain pathfinding?*

### 1.6.1 Thesis Outline

The rest of this thesis is structured as follows.

**Chapter 2** reviews the basic ideas of network information and data models.

We examine the current state of the art, and discuss related research.

**Part I** describes the Network Description Language:

**Chapter 3** presents our information model, the Network Description Language, and the reasoning behind the design decisions that we have made to come to this model.

**Chapter 4** validates our model by implementing components in the context of real life networks. It further discusses the strengths and limitations of our approach.

**Part II** examines the aggregation of network topology, and the impact of aggregation on pathfinding:

**Chapter 5** presents different ways of aggregating network topologies, and discusses related work.

**Chapter 6** describes the experiments that we performed to determine the impact of aggregation on pathfinding.

**Chapter 7** summarizes the overall research work presented in this thesis, and answers to the research questions.