Delft University of Technology

Faculty of Electrical Engineering, Mathematics and Computer Science

Network Architectures and Services

A simulation and capacity management tool for KPN's Ethernet network

Ruud van de Bovenkamp

(1289233)

Committee members:

Supervisors:	Dr. F.A. Kuipers
	Ing. J.A. Hoffmans
Members:	Prof. P.F.A Van Mieghem
	Dr. A.C.C. Lo

September 2010

M.Sc. Thesis No: PVM 2010-066

Contents

List of Figures

List of Tables

Abbreviations

BB	**B**ack**b**one
CDN	**C**intent **D**elivery **N**etwork
DSLAM	**D**igital **S**ubscriber **L**ine **A**ccess **M**ultiplexer
Gbps	**G**iga **b**it **p**er **s**econd
MA	**M**etro **A**ccess
MB	**M**etro **B**ridge
Mbps	**M**ega **b**it **p**er **s**econd
MC	**M**etro **C**ore
IPTV	**I**nternet **P**rotocol **T**ele**v**ision
ISP	**I**nternet **S**ervice **P**rovider
VoD	**V**ideo **o**n **D**emand
VPN	**V**irtual **P**rivate **N**etwork

Chapter 1

Introduction

The purpose of a communication network is to allow users to communicate with each other. Traditionally, different types of communication, or services, used a dedicated network to that service. Telephone calls made use of the telephony network while television content was distributed over television networks. This one network per service paradigm has been changing ever since the advent of internet and the growth of data communication. Modern communication networks converge more and more towards general purpose, packet switched networks.

The transport of traffic for very different services makes the capacity planning of that network more involved. Each individual service places different demands on the network. Not only does the usage of a particular service differ from one geographical area to the next also the number of provided services differ from area to area. The variety of services and geographically different usage patterns often lead to a lack of overview on network usage. As a result capacity management usually comes down to upgrading links that are loaded above a certain threshold, while traffic forecasts are based on aggregates.

A solution to these difficulties in capacity planning would be to gain a better view of the network as a whole and the routes of the services that are transported over it. By identifying each individual service, and gaining insight in the traffic flows and volumes forecasts can be made on service level. These forecasts can then be used to find future bottlenecks in the network. In this thesis a tool to visualise, analyse and simulate traffic in KPN's Ethernet network is developed. Visualisation of the network structure and how different services form an overlay over the network topology helps to gain an intuitive understanding of bottlenecks. With the inclusion of traffic statistics for each node, gathered in KPN's monitor system, this intuitive feel for the network is fleshed out even more. Finally the ability to simulate traffic flows for every service makes it possible to translate forecasts of service usage and traffic per user per service into capacity demands in the various parts of the network.

1.1 Structure of this document

In the next chapter the main focus of the capacity planning tool, KPN's Ethernet network, is described. In chapter three this description of the network is continued in the context of complex network studies. Chapter four describes the tool and some of the work that went into developing it. The tool is used in chapter five to find the bottlenecks in the Ethernet for the internet and IPTV services in three scenario over the coming five years. In chapter six recommendations are put forward to optimise the Ethernet topology in order to provide enough capacity to transport the simulated traffic. Chapter seven concludes this thesis.

Chapter 2

KPN's Ethernet Network

The main focus of this thesis will be on KPN's Ethernet network. The Ethernet network is KPN's network of choice to connect both corporate and consumer customers. For the consumer market it is used to transport internet, voice and television traffic to people's homes. For corporate customers it can be used for type of connection the client wants. This can be as diverse from connecting base stations for mobile telephony to connecting offices to a central database. This chapter will discuss KPN's Ethernet network in order to give the reader an impression of the network that will be studied in this thesis.

2.1 Nodes

Apart from its technical properties, a node can also be described by two basic properties. These properties are the node's location and, more importantly, its function. The function of the node basically states on which hierarchical layer the node operates. A node can be described as having, for example, an MC function or, simply, as an MC node. A node's function, then, does not signify *what* a node does, rather *on which layer* it does it. What a node does, is simply to forward traffic from one interface to another. In the following text six functions or layers can be distinguished.

LL Local Loop nodes are the end nodes in the network. The end nodes can be KPN nodes, such as DSLAMs or KPN nodes placed at customer locations. They can also be the property of a customer. As such, these nodes can vary in size and capacity while sharing the property of being an end point of the Ethernet network from KPN's perspective.

MA Metro Access nodes are the access nodes to the Metro network. They form the first aggregation layer in the network. The approximately 200 MA nodes typically connect to multiple LL nodes.

MB Metro Bridge nodes bridge access nodes. The current network contains about 560 MB nodes. In the present network they function like MA nodes in that they connect to different LL nodes on one side and to a single MC node on the other side.

MC These are the Core nodes of the Metro network. MC nodes form the second aggregation level of the network. They connect to some 5 to 25 MA or MB nodes and to two different BB nodes. There are approximately 250 MC nodes.

BB These are the backbone nodes of the network. There are about 60 backbone nodes, divided into two sub networks as will be explained later. The backbone network is the third layer of aggregation, it condenses the 250 MC nodes into about 60 BB nodes.

WAP Wholesale Access Points are the highest-layer nodes in the network. WAP nodes provide access to the network for wholesale parties. An internet provider that uses KPN's ADSL network to serve their customers, for example, will connect its network to a WAP node. From that WAP node every DSLAM in the network can be reached. It is also possible to serve only a part of the country and connect to a regional access point. Such an access point will usually be located at an MC location.

2.1.1 Locations

A node is typically located in a building, and a building typically houses more than one node. A node's location is, therefore, not a unique property. A location can house different nodes with different functions. It is also possible, in areas with a high traffic load, that a location houses several nodes with the same function. A location that houses a BB node usually also houses the MC and MA nodes for that area. In order to distinguish between different locations, a location is also said to have a function. An

MC location, for example, houses an MC node. A location's function is that of the highest-layer node it contains.

2.2 Links

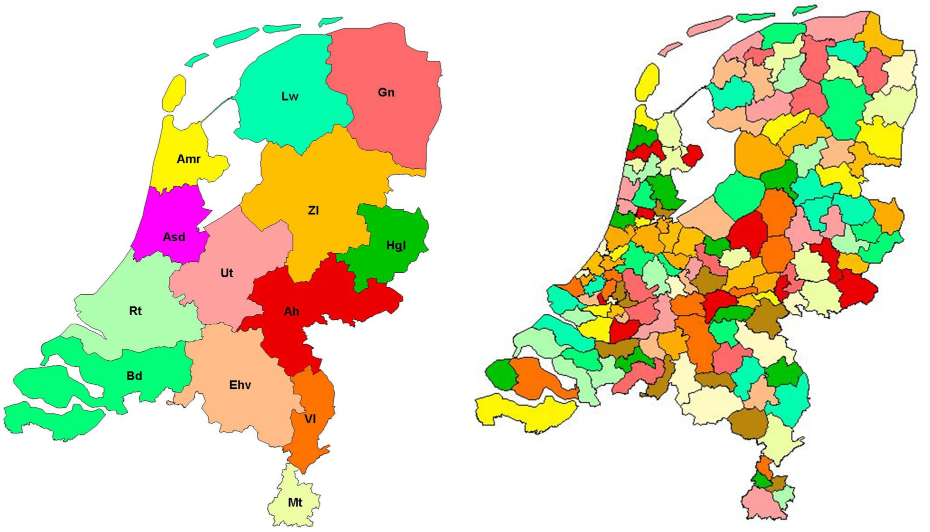
All Ethernet nodes are connected via the optical transport network. The speed of these connections depends on the interfaces in the nodes. The current network implementation consists mainly of 1 Gbps links. Of the total of 17655 links active on the 7$^{\text{th}}$ of April 2010, 16654 are 1 Gbps links. One Gbps links are mainly used to connect local loop nodes to MA and MB nodes and then further to MC nodes. The second most used type of link is the 10 Gbps link. These links are used mainly in the backbone and MC links. In total there are 947 10 Gbps links in the network. Two small minorities of link speeds are 2.490 Gbps and 100 Mbps.

2.3 Topology

The network's main topology is based on the geography of The Netherlands. The country is divided into 13 transport areas as illustrated in figure 2.1(a). These transport areas are further subdivided into metro core areas, as shown in figure 2.1(b).

Every Metro Core Area is governed by a metro core location, while each transport area has two backbone locations. The relationships between the various areas is illustrated in figure 2.2. As can be seen from this figure, the smallest area is the access area. All access areas together provide national coverage of the Ethernet network.

The MC locations are connected to BB locations in ring structures. From the two backbone locations several fibre rings depart to connect a group of MC locations. Because of the ring structure, every MC location is connected to a backbone location by two different paths. When a single MC location passes a lot of traffic over the ring to a BB location a direct connection between that MC location and the BB location is made. This does not mean that a new cable has to be laid between the locations because Ethernet traffic passes over the underlying optical transport network. The backbone locations are, in turn, connected to each other in ring structures. As indicated earlier,



(a) Transport Areas (b) Metro Core Areas

FIGURE 2.1: Transport and Metro Core Areas

the backbone network consists of two separate networks, the A Net and B Net to provide redundancy. Figure 2.3 illustrates this double ring structure.

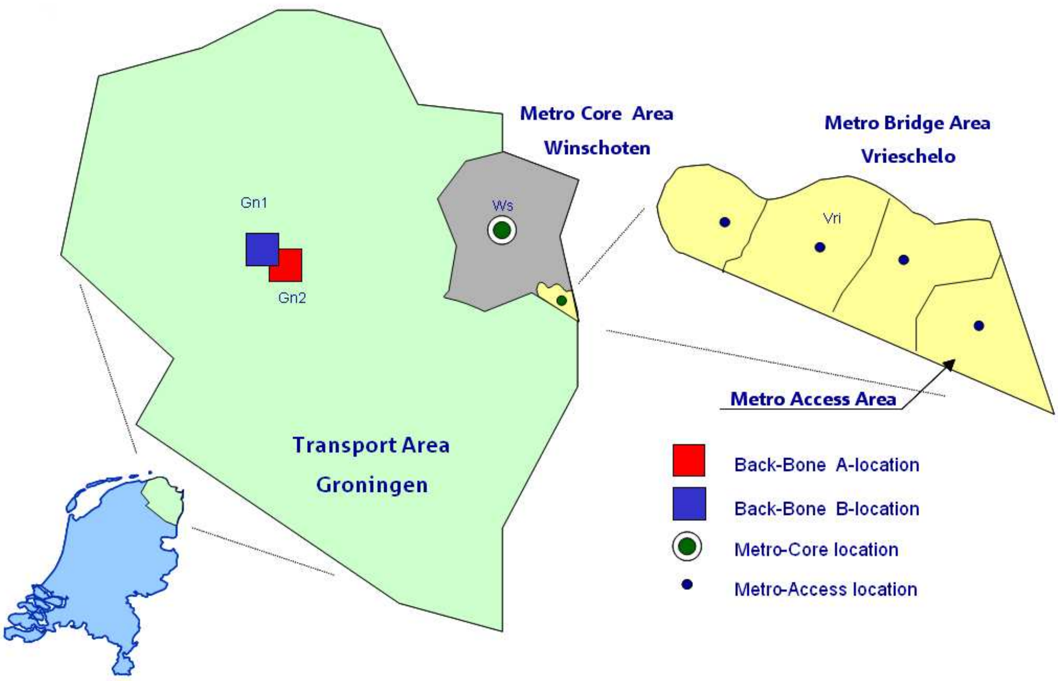


FIGURE 2.2: BB and MC Rings

The subdivision of the country in increasingly smaller areas causes the network to be

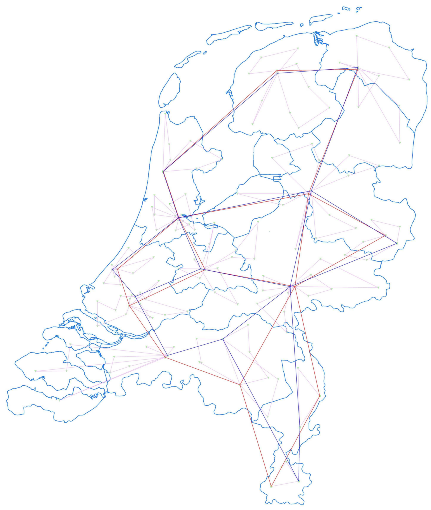


FIGURE 2.3: BB and MC Rings

highly hierarchical. Roughly 750 different access areas connect to four WAP locations. The hierarchical nature of the network is shown in figure 2.4.

This impression of the network displays locations and their logical relations. The locations shown in green are MA locations, they connect to a single MC location. MC locations are shown in blue in this figure, they connect to two different BB locations. The locations shown in red are WAP locations, and are the top nodes in this hierarchy. Every BB location connects to two different WAP locations. At these locations WAP nodes connect to content providers such as ISPs and IPTV providers.

Figure 2.4 does not show connections between nodes on the same hierarchical layer. As a result it does not show, for instance, the backbone ring structures that are shown in figure 2.3. Although this simplifies the picture, the graph is still most relevant for internet and IPTV/VoD connections as they originate at one of the WAP locations. It shows, for example, that if all users request a VoD stream at the same time, the WAP nodes will need to generate an enormous amount of traffic. Other connections such as business market VPN connections do not necessarily go up and down in hierarchy, as they can be made on any level.

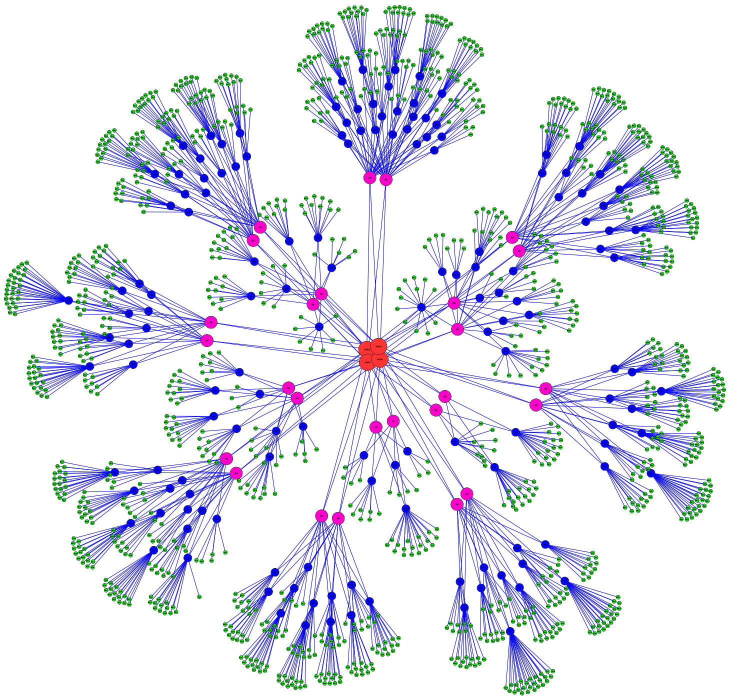


Figure 2.4: Main network structure

2.4 Services

The Ethernet network is used to transport traffic for various different services. These services vary from single point to point connections for companies to broadcast IPTV. The traffic for a particular service is transported over a VPN. This VPN is then statically routed over then Ethernet network. From the MC locations onwards the VPN is routed label switched. A VPN can be part of one of three different traffic classes. For the highest traffic class the bandwidth is guaranteed and the traffic is directed to a priority queue in the switch. The bandwidth for second highest traffic class traffic is only partly guaranteed and is partly overbooked. The lowest traffic class is best effort traffic.

Chapter 3

Complex Networks

In its most abstract form, a network is simply a collection of nodes connected by links. The nodes and links in a network can represent very different real world phenomena. For example, nodes can denote individuals while a link between two individuals represents friendship, kinship, working relations, former classmates or many other things. Other examples are networks of food webs, road and railway systems, flight schedules, but also the interactions of the constituent parts of cells or the monetary flows in economies. Because networks can be used in so many different fields of research, the study of networks has received considerable attention in literature. In this chapter, KPN's Ethernet network is further described and analysed as a complex network.

3.1 Network Metrics

According to Wang and Chen, the study of complex networks took off with the attempts of Erdös and Rényi in the 1950s to describe networks with the help of random graphs [1]. They created random graphs by taking a number of nodes N and connecting each pair of nodes with a certain probability p, resulting in approximately $pN(N-1)/2$ links. A random graph can be used to examine for what connecting probability p in combination with the number of nodes N a particular property of the graph appears. The properties of a graph can be determined by a number of metrics including path length, clustering coefficient, degree distribution, assortativity and algebraic connectivity.

Path length

In a graph distances are measured in the minimum number of links that separate two different nodes. In undirected graphs the difference from node i to node j (d_{ij}) is the same as the difference between node j and node i (d_{ji}), but in directed graphs this is not necessarily the case. The path lengths between nodes express the diameter, eccentricity and size of a network. The diameter is the maximum path length in hops between any two nodes in the network, while the eccentricity of a node is the maximum shortest path length from that node to any other node in the network. The diameter is a global network metric, while the eccentricity is a local metric. The eccentricity of a network as a whole is the average of the node eccentricities. In contrast to the eccentricity of a node, its closeness is the average smallest path length to all other nodes. The node with the smallest closeness value in a network is the most central node, and can reach, on average, all nodes in the smallest number of hops. The size of the network in terms of path length is given by the mean distance between two nodes.

Clustering coefficient

The clustering coefficient C_i indicates to what extent the nodes that connect to a node i are also connected. The familiar example is that of a network of friends. A clustering coefficient close to one indicates that my friends are also friends among themselves. A node's clustering coefficient is determined by the ratio between the number of links between nodes that connect to it to the number of possible links between the nodes that connect to it. From this definition it follows that $C_i \leq 1$. The clustering coefficient of the whole network is the average over all the node's individual clustering coefficients.

Degree Distribution

The degree k_i is defined as the number of links that are incident on node i. The degree of a node is usually taken as a measure of importance of the node in the network. The average degree of a network is the average over all the node degrees. The degree distribution gives information about the occurrence of a particular degree in the network. In a regular lattice all nodes have the same degree, while in a large completely random network the degree follows a Poisson distribution. In real world networks the degree distribution, and in particular the tail of the degree distribution differs from Poisson and is studied widely as will be discussed in section 3.2.

Assortativity

Assortativity in a network describes how nodes connect to each other. Newman calls the way nodes of different types connect to each other mixing patterns and mentions different ways to construct a mixing matrix [2]. The way nodes of various types mix is particularly interesting in social studies. A property that is more generally used in the studies of complex networks is that of mixing according to node degree, this kind of mixing is also called degree correlation. There are various ways to determine the node assortativity r. One is to determine the average degree of the nodes that a node connects to as a function of the degree of the node. If this function increases with the degree the network is positive assortative. The assortativity can also be expressed as a single number by using the following formula taken from Newman [3]

$$r = \frac{L^{-1}\sum_i j_i k_i - \left(L^{-1}\sum_i \frac{1}{2}\left(j_i + k_i\right)\right)^2}{L^{-1}\sum_i \frac{1}{2}\left(j_i^2 + k_i^2\right) - \left(L^{-1}\sum_i \frac{1}{2}\left(j_i + k_i\right)\right)^2}$$

where j_i and k_i are the degrees of the nodes connected by the i-th link. The value of r lies between -1 and 1, where $r = -1$ means that nodes with a high-degree tend to connect to nodes with a low-degree and $r = 1$ means that high-degree nodes tend to connect to other high-degree nodes. In [4] it is shown that Newman's degree correlation coefficient can be formulated in terms of the total number N_k of walks in the graph with k hops. By doing so a relation between the assortativity of graphs and the algebraic connectivity. Rewiring a graph in order to make it more disassortative also increases the algebraic connectivity. This suggests that disassortativity increases the connectedness of graphs.

Algebraic Connectivity

The term algebraic connectivity, denoted as $a(G)$, was coined by Fiedler and denotes the second smallest eigenvalue of the Laplacian of a graph [5]. The Laplacian Q is a matrix formed from the adjacency matrix of a graph according to $Q = \Delta - A$, where Δ is a diagonal matrix with on its i^{th} entry the degree of node i and A is the adjacency matrix. It is a well-known fact that a graph is only connected if the second smallest eigenvalue is non-zero. The value of the algebraic connectivity is often used as a measure for how well a graph is connected.

3.2 Small-World and Scale-Free networks

The Small-World model was proposed by Watts and Strogatz in [6]. They came to their model while looking for a way to construct a graph that was somewhere halfway of being either completely random or completely regular. The graph is constructed as follows. The starting point is a ring lattice with N nodes of degree k. Then with a probability p each link is rewired to a uniformly chosen new node. By changing the probability p a graph can be created that is somewhere midway regularity ($p = 0$) and randomness ($p = 1$). They then characterised the obtained graphs based on the average path length (d) and clustering coefficient (C) and found that the average path length decreases very quickly for low values of p while the clustering coefficient remains virtually unchanged. At the two extreme values for p, d and C are given by

$$d \sim N/2k \gg 1 \text{ and } C \sim 3/4 \text{ as } p \to 0 \tag{3.1}$$

$$d \sim \ln(N)/\ln(k) \text{ and } C \sim k/N \ll 1 \text{ as } p \to 1 \tag{3.2}$$

The small average path length in the graphs gives them the name small-world. When Watts and Strogartz compared their graphs to those of real networks (a network of film actors, the power grid of the western United States, and the neural network of the worm *C. elegans*) they found that these networks display small-world properties. These small-world properties are that the average path length is not much larger than d_{random} while the clustering coefficient is much larger than C_{random}.

While the small-world model of Watts and Strogatz describes the existence of networks of many nodes that have small average path lengths and relatively high clustering coefficients the model does not describe the degree distribution of many real world networks. The difference between the degree distribution of random networks and real world networks became apparent with the study of the structure of the internet [7]. In an Erdös Rényi random graph the degree distribution follows a Poisson distribution. The degree distribution of a small-world graph as constructed by Watts and Strogatz is less well defined but both distributions decay exponentially for large degrees. As a result of the exponential decay it is very unlikely to find nodes with a large number of links. In their study of the structure of the internet, however, Barabási and Albert found that nodes with very high degree do in fact occur. It turns out that the degree distribution in many real networks decays as a power law and not exponentially. The probability $P(k)$ that a node is connected to k other nodes is given by $P(k) \sim k^{-\gamma}$. The power law in the degree distribution accounts for the name scale-free because power laws are the only scale free functions. Barabási and Albert explain the scale-free nature of networks by introducing the concepts of growth and preferential attachment. They state that the number of nodes and links in a real network are not static but grow over time. With the introduction of every new node it must connect with a node already present in the network. According to the preferential attachment principle the new node will connect with a greater probability to a node of higher degree. Barabási and Albert found an exponential degree distribution in the world wide web, and the same actor network and power grid as Watts and Strogatz used. In the case of the actor network the preferential attachment principle manifests itself in the fact that beginning actors are more likely to get a small role in a production that includes some star actors than in a production with a cast of beginners only.

To test whether networks that are subject to both growth and preferential attachment have a power law degree distribution Barabási and Albert ran simulations of growing networks. In their simulations each new node has a chance to connect to a node already in the network that depends on the degree of that node. The simulated networks did indeed show a power law degree distribution. Since the publication of their findings many real networks have been investigated and exponents ranging from 2.1 to 4 have been found. Real networks, however, do not always display a simple power law relation [8]. The fit is usually of the form $P(k) \sim (k + k_0)^{-\gamma} \exp(k/k_x)$, where k_0 is the lower cut-off of the power law region and k_x is the the length scale of the exponential higher cut-off. The power law region lies between these two cut-off values. The most common deviation from the power law distribution is a flattening of the curve for small degrees; a high degree cut-off is less common. Although power law degree distributions are found in many different networks, it is not easy to measure the tail of the distribution due to the fact that the tail does not contain many sample points [2]. Newman offers two ways around this. One way is to make the bin sizes increase exponentially in the histogram. This will make more samples fall into the same bin in the tail. Another way is to make a plot of the cumulative distribution function and plot it on a log-log scale. The cumulative distribution function of a power law distribution plotted on a log-log scale displays a straight line. An exponential distribution gives a straight line when plotted on a log-normal scale. A straight line alone is, however, not sufficient to say that the distribution follows a power law [9]. Clauset et al. show that under certain conditions a power law, log-normal and exponential distribution all show straight lines on a log-log scale. To overcome this problem they introduce a plausibility value (p-value), to indicate how plausible it is that a data set follows a power law distribution. The value is determined by measuring how far samples taken from a generated, true power law distribution vary from the power law form and comparing this to similar measurements performed on the dataset. If the empirical dataset is much further from the power law form than the generated one it is not probable that it follows a power law distribution. A p-value close to one indicates a very plausible power law distribution while a value close to zero indicates a very improbable power law fit. The power law fitting and p-value calculations performed in this chapter are done with Matlab code from the authors of [9].[1]

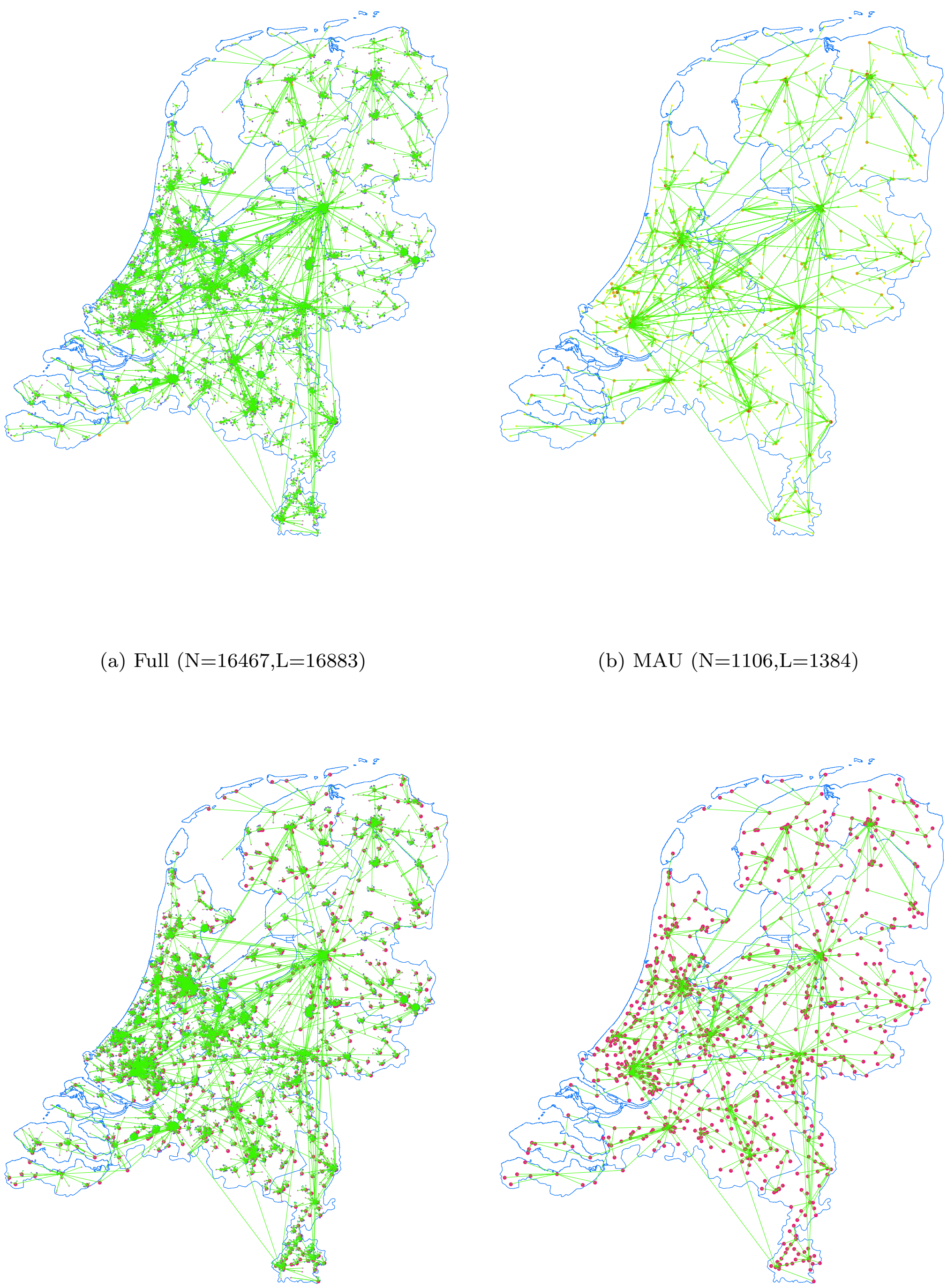
3.3 KPN's Ethernet Network as Complex Network

In this section the results of the complex network study of KPN's Ethernet network are discussed. The network metrics discussed in section 3.1 have been measured in various forms of the network as discussed below.

3.3.1 Studied Networks

For the calculation of the network metrics six different (sub)networks were used. These networks are illustrated in figures 3.1 and 3.2. Figure 3.1(a) shows the complete Ethernet network. This network contains all the Ethernet nodes, including the local loop nodes in DSLAM's and at corporate customer locations. A smaller version of the full network is shown in figure 3.1(b). In this figure only the nodes up from the level of MA nodes are taken into account, hence the name MAU which stands for MA upwards. As the Ethernet network is designed with national coverage in mind, the notion of locations, sometimes, carries more significance than nodes, as is explained in chapter 2. In order to emphasise locations more, all the nodes at a location are merged into a single node in the network shown in figure 3.1(c). The network in 3.1(c) also contains all local loop

[1]The Matlab code can be found at `http://tuvalu.santafe.edu/~aaronc/powerlaws`



(a) Full (N=16467,L=16883)

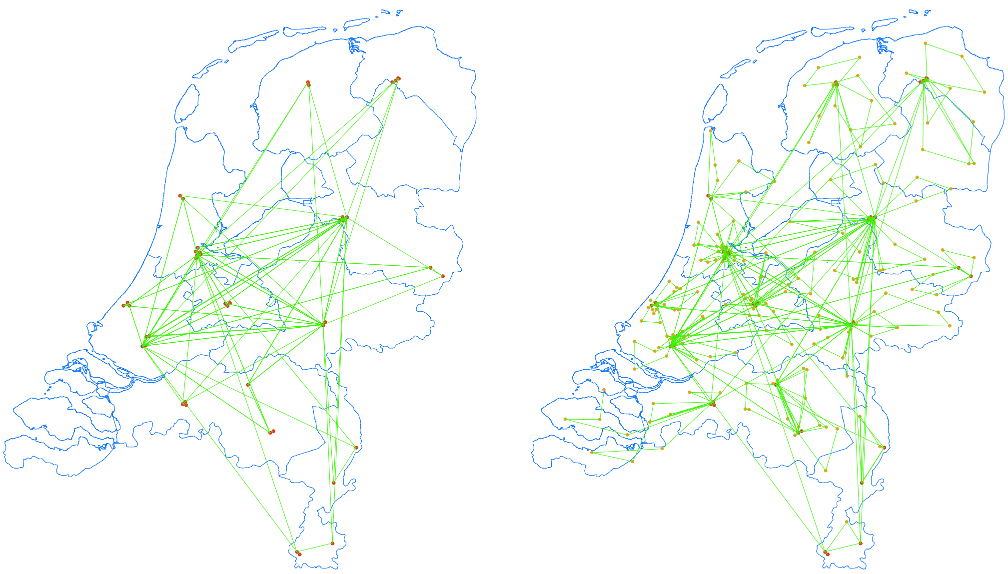
(b) MAU (N=1106,L=1384)

(c) Merged Full (N=16011,L=16222)

(d) Merged MAU (N=886,L=1061)

Figure 3.1: Networks

nodes. The final network in figure 3.1(d) contains only (merged) nodes from MA layer upwards. Figure 3.2 shows two networks that are subsets of the nodes shown in figure



(a) BB (N=93,L=196) (b) BBMC (N=343,L=621)

FIGURE 3.2: Networks Subs

3.1(b). In figure 3.2(a) only the backbone nodes are shown while in figure 3.2(b) both the backbone nodes and the metro core nodes are shown. In all the networks a link between nodes means that there is at least one connection between the two nodes. It is possible that multiple interfaces connect between two nodes to provide higher capacity, however, they still count as one link.

3.3.2 Results

Table 3.1 contains the metrics as calculated for the different networks described in the previous section. As can be seen from this table, the number of local loop nodes is very large as compared to the number of MA and higher layer nodes. There are almost fifteen times as many local loop nodes in the network as higher layer nodes. It is not surprising that there are so many local loop nodes in the network as it is, for a large part, a content delivery network. A few nodes at the top of the hierarchy provide access to content sources such as the internet or video broadcast for all the subordinate nodes.

The merged version of the network contains approximately 400 fewer nodes than the full version while the merged MAU network contains roughly 200 nodes fewer than the normal MAU network. This is because local loop nodes can also be located in buildings that contain higher layer nodes.

	Normal		Merged		Core	
Metric	Compl.	MAU	Compl.	MAU	BB	BBMC
N	16467	1106	16011	886	93	343
L	16883	1384	16222	1061	196	621
$\overline{d}$	7.37	6.02	5.91	4.86	3.01	4.36
$\overline{E}$	9.77	8.18	8.85	7.32	4.42	6.78
$\max(E)$	12	10	11	9	6	9
Cl	4.07	3.23	3.37	2.85	1.86	2.54
LD (%)	0.0124	0.226	0.0127	0.271	4.58	1.06
$\min(LD)$(%)	0.0121	0.181	0.0125	0.226	2.15	0.583
$\overline{C}$	0.0012	0.0157	0.001	0.043	0.075	0.11
A	-0.41	-0.18	-0.13	-0.14	-0.25	-0.02
$a(G)$	0.0015	0.0284	0.0029	0.041	0.463	0.098
$\overline{k}$	2.05	2.50	2.03	2.40	4.22	3.62

TABLE 3.1: Network Metrics. The measured metrics are: total number of nodes N; total number of links L; average distance $\overline{d}$; average eccentricity $\overline{E}$; maximum eccentricity $\max(E)$; minimum closeness Cl; link density LD; minimum link density $\min(LD)$; average clustering coefficient $\overline{C}$; node assortativity A; algebraic connectivity $a(G)$; average degree $\overline{k}$.

One of the most striking things that can be learned from this table is the scarcity of links in the network. This does not mean, however, that there is only 1 or 10 Gbps of available bandwidth between any two nodes in the network, because a link can consist of multiple connected interfaces. What it does show is that a node typically connects to only one other node. This is true for all local loop nodes, and as we saw earlier, they make up more than ninety percent of the nodes in the network. As a result the link density is very close to the minimum link density of a connected graph with the same number of nodes.[2] The networks that contain only MA and higher layer nodes have a slightly higher link density, as can be expected, but only the BB and BBMC networks have a significantly higher than minimum link density. This is because from the MC layer nodes upwards every node connects to two other nodes in order to be able to reach both the A and the B net, as explained in chapter 2.

The average path length of the full graph is approximately 7. In other words, a node can reach any other node in seven hops on average. It can reach its most distant node

[2]The minimum number of links in a connected graph of N nodes is $N-1$, while the maximum number of links is given by $N(N-1)/2$, this leads to a minimum link density of $2/N$.

in approximately ten hops, as can be seen from the average eccentricity. Even the most eccentric node in the full network can reach all nodes in twelve or less hops. The node with the lowest closeness, sometimes called the most central node, can reach all nodes in the network in four hops on average. For the MAU network all path length based metrics are generally one hop lower than for the full network. This is the result of the fact that an MB or MA node is only one hop further than an MC node. The most eccentric node in the MAU network can reach all other nodes in two hops less than the most eccentric node in the full network. This again is the result of two time one added hop to get from an MC to either an MB or MA node. The same line of reasoning holds to explain why the metrics are again about one hop lower for the merged network; most MA and MB nodes are merged with their MC node. With the smaller scale of the backbone networks average path length decreases as well, the only exception is the average eccentricity of the BBMC network which is only one hop lower than that of the MAU network.

The clustering coefficient of most of networks is very low, which is to be expected for networks with a very low link density. If every node connects to just one other node in general it has no neighbours that can connect amongst themselves which results in a zero clustering coefficient. Even nodes that have a higher degree are often connected in a star topology which also results in a zero clustering coefficient. To make the influence of the nodes with zero clustering more pronounced the number of nodes with non-zero clustering coefficients are listed in table 3.2.

	Normal		Merged		BackBone	
	Compl.	MAU	Compl.	MAU	BB	BBMC
N. nodes	16467	1106	16011	886	93	343
No. non-zero cc	150	112	122	112	23	112
Av. non-zero cc	0.13	0.15	0.13	0.34	0.30	0.35

Table 3.2: Number of nodes with non-zero clustering coefficient

The rise in clustering coefficient from Normal to Merged to Core and within those three groups is largely due to the decrease of the number of nodes. The number of nodes with a non-zero clustering coefficient stays more or less the same as can be seen in the table. Table 3.2 makes it clear that the low average clustering coefficient of the larger networks is the direct results of the increased number of nodes with zero clustering. The average clustering coefficient excluding the zeros stays more or less the same until the Merged MAU where it increases to about a third.

Now that we have seen the average distance between nodes in the various networks and the clustering coefficients, the question arises whether the networks are small-world networks or not. By comparing the metrics of the six networks to the limits for the clustering coefficient and average path length of the completely regular and the fully random small world graph proposed by Watts and Strogatz this question can be answered. Equations 3.2 and 3.1 give the clustering coefficient for a completely regular and completely random graph. The average degree is taken as $k = 4$ for the Core networks and $k = 2$ for the other four networks. In table 3.3 the results are summarised. As can be read from this table, the average path length in all the graphs is much smaller than for a regular graph. In all but the BBMC network the average path length is even lower than in a random graph. The clustering coefficient in the networks is, in general, a factor ten greater than that of a small world graph. The combination of a very low average path length and the greater than random clustering coefficient answers the question whether the networks under investigation are small world networks affirmatively.

	Normal		Merged		Core	
	Full	MAU	Full	MAU	BB	BBMC
N	16467	1106	16011	886	93	343
$\overline{d}_0$	4116.25	276.5	4002.75	221.5	11.63	42.88
$\overline{d}$	7.37	6.02	5.91	4.86	3.01	4.36
$\overline{d}_{random}$	14.01	10.11	13.97	9.79	3.27	4.21
$\overline{C}_0$	0	0	0	0	0.75	0.75
$\overline{C}$	0.0012	0.0157	0.001	0.043	0.075	0.11
$\overline{C}_{random}$	0.00012	0.008	0.00012	0.0023	0.043	0.012

TABLE 3.3: Clustering Coefficient and average path length

The node assortativity is negative for all six networks. This negative assortativity is in accordance with Newman's observation that while social networks appear to be assortative other networks appear to be disassortative [2]. The disassortativity is strongest for the full network, because many local loop nodes can connect to a single higher layer node. It is smallest in the BBMC network because half of the nodes, the BB nodes and a part of the MC nodes, connect to similar degree nodes while the other half, most of the MC nodes, connect to high degree BB nodes.

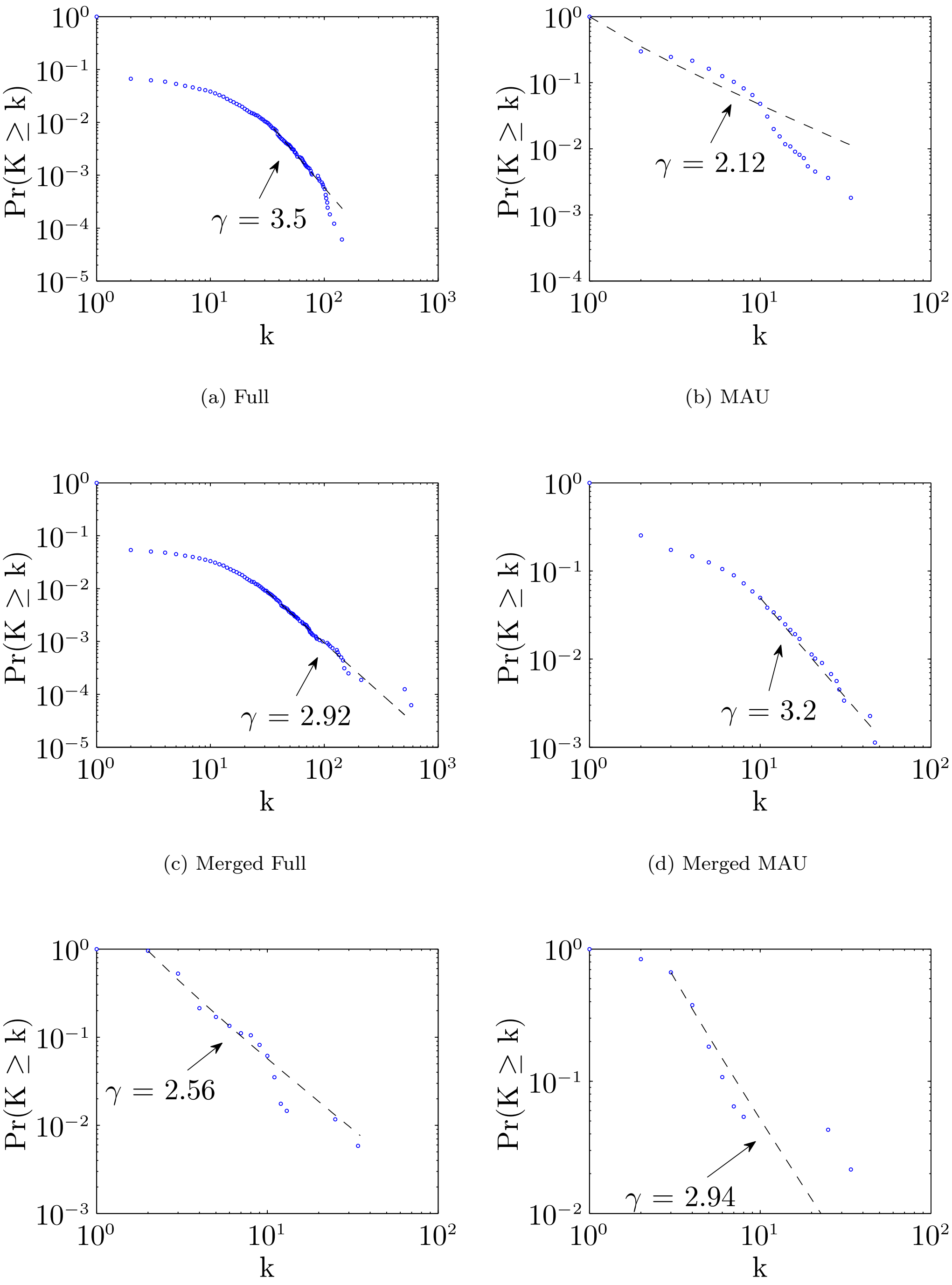
In order to be able to calculate the algebraic connectivity of the full network and the merged full network the Laplacian of the network was saved as a sparse matrix in Matlab. With the Matlab function *eigs* the two smallest eigenvalues were calculated. Because *eigs* is a numerical solution that uses a random starting vector the resulting eigenvalues have a slightly different value every time the function is executed. To overcome this, the average value over 100 runs is used. As can be expected from the network topology, the algebraic connectivity is much larger for the backbone networks. Coincidently the backbone network is also more disassortative. Because of the double backbone network and the fact that every MC node connects to both backbone networks the network is more robustly connected. Both the normal and the merged full networks have so many single connected local loop nodes that the connectivity is very low. The two MAU networks share the fact that they have an $a(G)$ that is fifteen to twenty times as high as the full ones.

As can be seen from the last row of table 3.1, the average node degree lies around two for the larger networks and around four for the core networks. However, these two values do not provide full information about the degree distribution. In order to verify whether the six networks are scale-free or not, a power law is fitted to the cumulative distribution function as described in section 3.2. The results are given in figure 3.3 on page 21. The values for the exponents vary between 2.12 and 3.5. As mentioned in section 3.2, the fact that we can draw a straight line through the data points and calculate a slope for it does not mean that the distribution actually follows a power law. In order to gain confidence in the fits the accompanying p-values are given in table 3.4. The results correspond to

	Normal		Merged		BackBone	
	Compl.	MAU	Compl.	MAU	BB	BBMC
γ	3.50	2.12	2.92	3.20	2.56	2.94
p-value	0.41	0	0.89	0.80	0.21	0

Table 3.4: Results Power Law Fit

what one might induce from a visual inspection of the graphs. The node degree of the full network follows a power law distribution with only a marginal probability and only if the local loop nodes are not taken into account. It not very probable that the BB network follows a power law node degree distribution. The most probable fits are those of the merged networks with p-values of 0.89 and 0.80. This means that the network of physical locations in the Ethernet network is scale-free with and without the local loop nodes. Growth and preferential attachment can partly explain the scale-free property of the merged network but the geography of the network is the dominant factor. When



(a) Full

(b) MAU

(c) Merged Full

(d) Merged MAU

(e) BBMC

(f) BB

FIGURE 3.3: Power Law Fits

the network grows a new node cannot connect to just any node, it has to connect to a node in the vicinity. Another important factor is that the number of locations is more likely to decrease than to increase because buildings are expensive. Only nodes that cannot be placed in an existing location, such as DSLAMs that serve a region furthered removed from a location, will form a new location and link to an existing one. Nodes that are not at a merged location are then almost always DSLAMs and nodes at corporate custumor locations. Especially the placement of DSLAMs is dictated by the population density. The spread of corporate customers is to a large extent also dictated by population density but less so. These nodes connect to the closest higher layer node; which can be seen as preferential attachment. As a result, merged nodes in more populated areas become high degree nodes. This can be seen from the fact that five out of ten merged nodes with the highest degree are in the top 15 of most populated municipalities in The Netherlands.[3] The two merged nodes with the highest degree are located in Amsterdam and Roterdam, the two largest cities in The Netherlands.

3.4 Conclusion

The six different networks that were studied all show small world properties. The average distance between nodes is lower than can be expected from a random network in all but one of the networks. The clustering coefficient, however, is larger than that of a random graph. The combination of these two observations make the graphs small world. At the same time, some of the graphs show a power law degree distribution which makes them scale free. When the network of locations is considered, the power law distribution is a very probable fit. We can, therefore, say with confidence that the merged networks are scale free. The full network seems to be scale free as well, but the power law fit is less certain than in the case of the merged networks. One of the striking characteristics of all the networks, except for the Core networks, is the very low link density. Only a very small fraction of possible links is actually in place, in the full network the fraction is only a hundredth of a percent. As a result of the low link density the algebraic connectivity is low as well. A final observation is that the networks mix disassortatively. This is in accordance with other biological and technological networks.

[3] The population of The Netherlands per municipality as of January 2010 can be found at the website of Statistics Netherlands: http://statline.cbs.nl/StatWeb/selection/?VW=T&DM=SLNL&PA=03759NED&D1=0&D2=129&D3=0-4&D4=(l-1)-l&HDR=T&STB=G2,G1,G3

Chapter 4

Software Tool

In order to get a good impression of the network as a whole and to visualise the flows of traffic through the network a software program was written in Java. It essentially combines various sources of information from KPN to give a geographically correct visual representation of the network, including traffic information. Apart from being a tool to visualise the network and analyse traffic statistics, the tool can also be used to calculate network metrics. Although the tool was written with the specifics of KPN's Ethernet network in mind it is not limited in use to KPN's network. In this chapter the tool is briefly discussed.

4.1 Data Structure

The basic unit of the program is the node. A node has many properties including an id number, name, type, layer, street address and its geographical coordinates. Apart from these node specific properties, a node also has graphical properties such as shape, colour and marker size. Because the nodes in the model together form a network, they can be connected to each other. Every node contains a list of node ids of other nodes it connects to. Apart from the list of nodes a node connects to, it also stores a list of interfaces that facilitate these connections. It is possible that more than one link exists between two nodes. In this case for all individual connections the accompanying interfaces are stored. For each connection between two nodes a link is also stored separately from the nodes. Each link stores, amongst other properties, the source and destination node id as well as the capacity of the link. As the network can be drawn from real network traffic statistics each interface has inbound and outbound traffic fields, for both average and maximum values. The aggregate of the traffic fields of all the interfaces is stored in the nodes traffic field. Nodes, links and interfaces can all be individually turned invisible in order to filter the network for specific properties.

The data structure of the program can be used for generic networks. The KPN specific way to fill the network is described here. The network was filled based on traffic statistics from KPN.

1.	Inbound Throughput (bps)-avg	180083508.93484777
2.	Inbound Throughput (bps)-max	468743398.43342
3.	Outbound Throughput (bps)-avg	1316916819.7537613
4.	Outbound Throughput (bps)-max	2287615330.89458
5.	Inbound Errors (PDUs)-sum	0.0
6.	Outbound Errors (PDUs)-sum	0.0
7.	groupName	nl-ah-c-nge-775012-01
8.	resourceName	nl-ah-c-nge-775012-01(3/2/1)
9.	description	"Connection-to-nl-wtw-c-nge-74507-01-2/1/1"
10.	DisplaySpeed	10 Gbps
11.	AP ifStatus	up:up
12.	totalNoResources	25862.0

Table 4.1: Line taken from statistics

Table 4.1 shows the columns of a network statistics data file as rows. The first four rows contain the traffic information, followed by two rows of error information. Line seven gives the name of the node while line eight also includes the specific port information. In building the network model from these traffic statistics lines eight and nine are the most important lines. From line nine it can be read which node the node named in line seven connects to. The example given in table 4.1 shows node nl-ah-c-nge-775012-01 connected to node nl-wtw-c-nge-74507-01 via a 10 Gbps link. In this case the interface numbers are also given (3/2/1 to 2/1/1), this is, however, not always the case. Constructing the network from these lines was complicated by the fact that the description field is not always of the same make up. Tabel 4.2 records the different formats of the description field found.

A further complication is the fact that not every node's traffic is monitored. This means that some links occur only once in the traffic statistics, while other links occur twice. A link between two monitored nodes occurs twice in the statistics data. Once with node A in line seven of table 4.1 and node B in the description field, and once with node B in line

Connection to nl-gr-hrwk-nge-74507-01
Connection-to-nl-ah-2-nge-775012-01-6/1/1
Connection to nl-ah-laar-ngeb-74501-01:1/1/19
Connection to nl-ah-2-nge-775012-01 5/2/1
Connection to nl-ah-dc1-12c-m320-01 P6/0/0
Connection to nl-asd-dc2-12c-tx-01 ge-4/3/0
Connection to nl-hgl-dc1-12c-m320-01 so-0/0/1
Connection to nl-asd-dc2-ias-arg-47 poort 4/0/0

Table 4.2: Different formats of description field

seven and node A in the description field. When this is the case, it is important that the interface information is included in the description field so that both occurrences can be matched. Unfortunately this is not always the case. As a last resort the two interfaces can then be matched based on their traffic statistics. The traffic statistics, however, do not match exactly, because they are measured in different nodes and the intervals over which traffic is average need not overlap precisely. Non overlapping intervals lead to different measured values in different nodes.

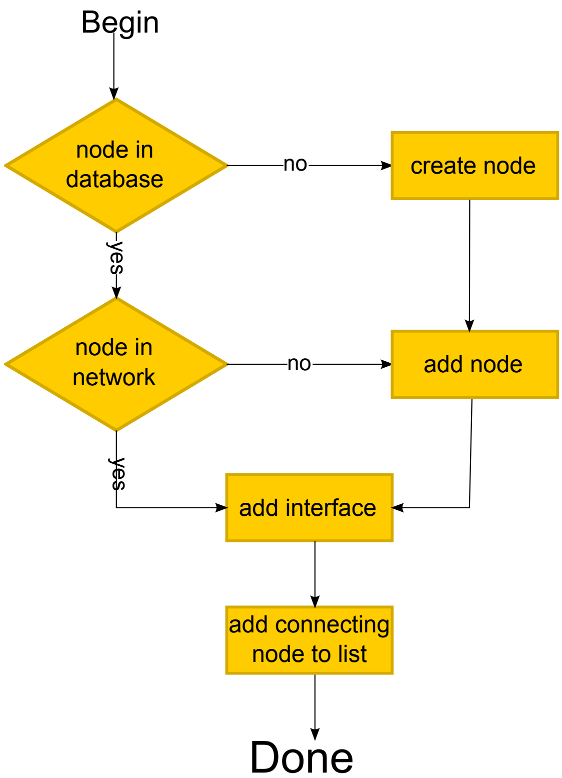


Figure 4.1: Flow Chart for adding a node

The flow chart describing the addition of a node is given in figure 4.1. For each line in the traffic statistics file the node name is read and checked against a database containing all the Ethernet nodes I gathered layer and address information about. If the name is found in that database it is checked whether the node is already in the network or not. In the case that the node already exists in the network a new interface is added to it. If the node is not present in the network, both an interface is added to the node and the node is added to the network. If the node is not present in the database a new node is created without database information, the interface is added, and the node is added to the network. The connecting node, taken from the description field, is then added to a list of connecting nodes that is processed when all the lines in the traffic statistics file are read. The connecting nodes are queued and processed later in order to avoid duplicate links for connections that are monitored in both nodes as explained earlier.

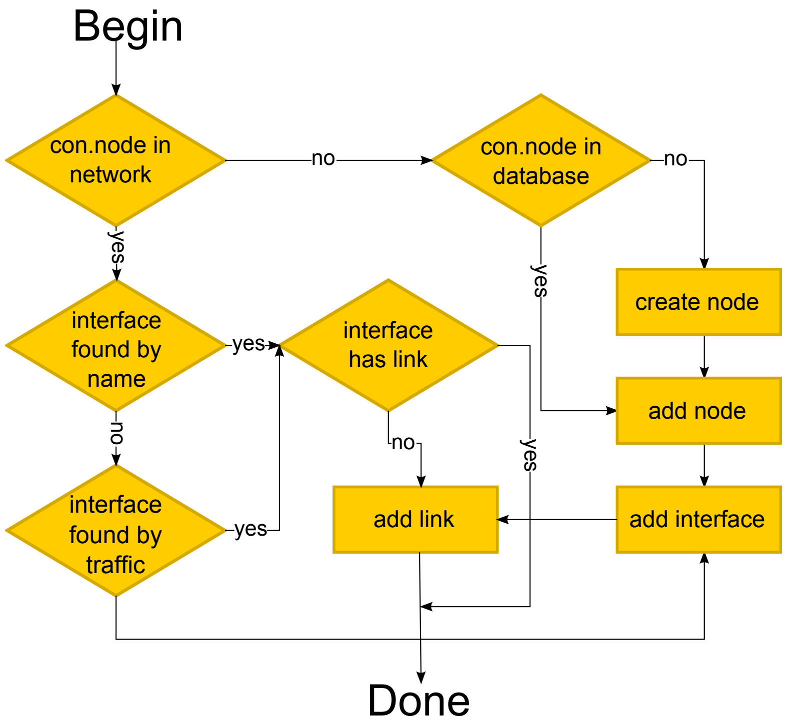


Figure 4.2: Flow Chart for adding a connecting node

The processing of the connecting nodes is slightly more involved than processing the monitored nodes. This is because connecting nodes of monitored nodes can themselves be monitored nodes. The flow chart for processing connecting nodes is given in figure 4.2. For each connecting node it is checked whether the node exists in the network. If it does, the node in the network is searched for the interface name. If an interface cannot be found, the traffic over the network node's interfaces is compared to that of the connecting node interface. If an interface can be found with the same name or with a similar traffic profile as the interface of the connecting node it is checked whether a link for that interface already exists. In the case that no link exists, a new link is created and the processing ends. When the connecting node is not present in the network but can be found in the node database it is added to the network and an interface and link are also added. In the case that the connecting node can neither be found in the network nor in the database a new node is created. As this node is created without database information it has no known street address or coordinates. In order to still be able to place the node in a location that has geographical relevance it is placed at a random position on a circle around the node it is a connecting node of. When the list of connecting nodes is finished the network is completely built up of the monitored and non monitored nodes. All the interfaces contain the traffic that was recorded in the network statistics.

As an overlay over the nodes and links VPNs can be loaded. All the vpns that are defined on KPN's Ethernet network are defined in a MySQL database. In this database the hopsequences of every vpn are stored as well as the booked capacity for that hop. The extra VPN information makes it possible to compare the booked, used and installed capacity of a link. VPNs are also used in the simulation function of the tool.

4.2 Graphical Structure

The program is built around three key concepts: nodes, links and a space to draw. The space to draw is the blank background of both the graph and map windows. The nodes and links are drawn over this background. Besides being a canvas to draw on, the background also catches the user's mouse input. In the network view, a slightly modified form of the Dutch national coordinate system, Rijksdriehoeks coordinaten (RD), is used. The adaptation lies in the fact that the y coordinate of the original RD system is flipped to match the standard Java screen coordinates (rectangular coordinates increasing from the upper left hand corner) and the system is translated to have the left most point of the contour of The Netherlands at $x = 0$ and the highest point of the contour at $y = 0$. The RD system was used because it is a rectangular coordinate system which makes it easy to use on a computer screen and because easy and accurate transformations from WGS84 to RD exist. The transformation from WGS84 to RD is needed as an intermediate step between the node's street address and a position on the screen. The street addresses were geocoded with the Google Maps Api to WGS84 coordinates. These WGS84 coordinates were then transformed to RD coordinates and plotted on the screen.

4.2.1 Contour of The Netherlands

The geographically correct representation of the network is displayed within a contour of The Netherlands. The contour consists of line segments that accurately represent the international borders and coastlines. The contour was obtained from the American National Oceanic and Atmospheric Administration(NOAA).[1]

4.2.2 Markers

The shape, size and colour of the node markers depend on the type and layer of the node. As can be seen in figure 4.3 the markers are all circular and their size increases for increasingly higher layer node. The colour coding of the node also depends on the node layer in normal situations. The colour of the node can be overruled when, for example, different groups of nodes are distinguished from each other based on to which source they connect or to what community they belong to. It is also possible to increase the size of the markers from the command window while maintaining the relative differences in size between the different layers. This can be convenient when saving the network view as a file for printing.

4.2.3 Containers

As already mentioned in chapter 2, many nodes share their physical location with other nodes which makes it impractical to draw individual nodes on the screen. Nodes at the same location would all be drawn on top of one another and only the top node would be clickable. To overcome this, nodes in the model are housed in 'locations' as well. A location in the model can contain one or multiple nodes. After the nodes are loaded into the network they are mapped to locations. This process works as follows. For each node it is checked whether a location with the street address of the node exists in the network. If this is the case the node is added to that location. If no location with the node's street address can be found a new location is created. The newly created location copies the

[1]See the we the website of NOAA for more information: http://www.ngdc.noaa.gov/mgg_coastline

street address and coordinates of the node it is created for. A location drawn on the screen using the marker shape, colour and size of the highest layer node it contains.

Just as nodes will end up drawn on top of each other if they are not grouped into locations so links would be drawn over each other if they run from different nodes at the same locations. A similar solution is used for links as is used for nodes. Links are mapped to link containers in a similar way as nodes are mapped to locations. Between two locations only one link container is drawn. This link container contains the links between all the nodes in the two locations.

4.3 User Interface

Figure 4.3 shows a screenshot of the tool's user interface. It shows the two different ways in which the network can be drawn. In the upper left hand corner a window containing a graph representation of the network is shown. In the upper right hand corner a geographically accurate representation of the same graph is shown, this representation is referred to as the map view. The two windows in the bottom half of the screen are the control window to the left and the date window to the right. The control window, as the name already indicates, is used to control the program. Through the command prompt the user can invoke all the program's functions. The command window also contains slider bars that allow the user to filter on interface utilisation. The sliders can be set up to filter on maximum throughput or average throughput. For each direction, inbound and outbound, two sliders can be used to set a minimum and maximum value for shown interfaces. The interfaces are filtered real-time so that the user immediately sees the interfaces appear or disappear while moving the sliders. Shown nodes and links can also be filtered based on properties such link speed, node type and layer and node name. To this end a tree structure as shown in figure 4.4 can be used.

In the data window various different panels with information can be contained. When a location is clicked with the mouse, for example, a panel containing the names and interfaces of the nodes that are housed in the location is shown including a graph for each node showing the interface traffic. An example can be seen in the lower right hand corner of figure 4.3. The data window can, however, also be used to display the names

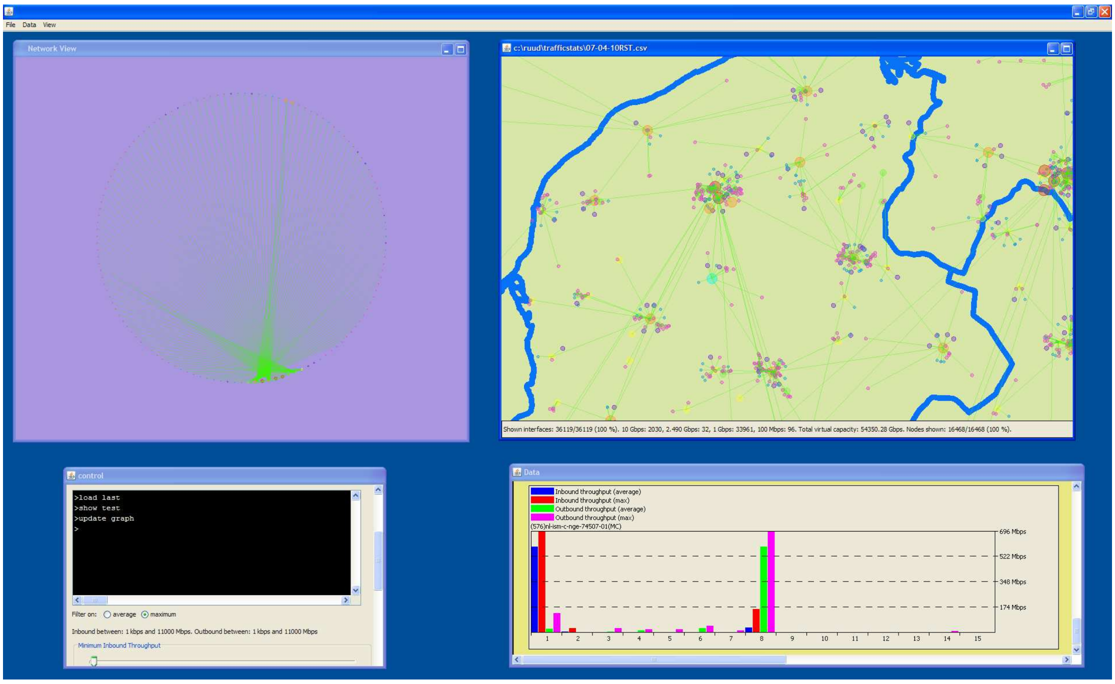


Figure 4.3: Screenshot of user interface

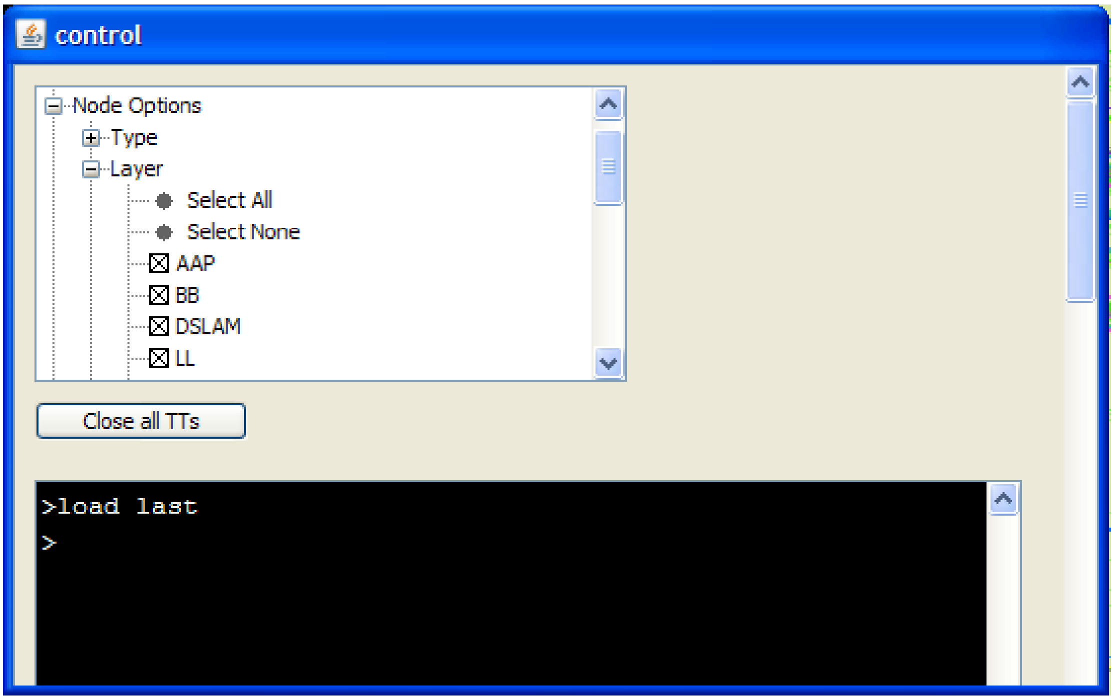


Figure 4.4: Screenshot of control window

and traffic of links within a link container or display a meter that shows the traffic that goes through a node during a simulation.

4.3.1 Mouse control

The user can use the mouse to access different information sources in the network. With a regular left mouse button click on a location, information about the nodes contained in the location are shown in the data window. When the shift key is held down and a location is clicked, the user can add a meter for one of the nodes in the location. The meter shows the simulated traffic flow through the node. Link containers are also clickable. When the map is clicked at a point that is not a location the link container with the smallest distance to that point is selected. A mouse menu is shown on the screen that allows the user to select which link in the container or, if VPNs are loaded, which vpn that is defined over one of the links in the container the user wants information about. The mouse can also be used to drag the map around and to zoom in and out with the scroll wheel.

4.4 Functionality

The tool is built to allow users to visualise and analyse a network. Because the tool can display or hide nodes and links based on several criteria, the network can be shown at every desired level of detail. Apart from visualisation the tool can also be used to calculate various network metrics based on the network as it is shown or even based on that part of the network that is in view. This makes it very easy to compare, for example a network metric calculated for the entire network to the same metric but calculated for only the higher or upper layers of the network. A full list of commands is given in Appendix A.

4.4.1 Visualisation

The visualisation capabilities of the program are perhaps best illustrated by means of an example. Figure 4.6 on page 33 shows five images from left to right and top to bottom illustrating the hierarchical structure of KPN's Ethernet network. The first image shows the two backbone locations in Leeuwarden. The second image shows all the nodes connected to these two backbones that are on a lower layer, these are the MC locations. In the third image nodes that connect to nodes on the same layer in the window are shown. The fact that they need to be in the window prevents the links between other backbone locations to be shown. The fourth and fifth images show increasingly lower layer nodes connecting to nodes that were already shown. These images where created with the command sequence shown in figure 4.5. The first two

```
show node 631
show node 633
save view c:\images\f1.ps
show link lgap 1
show node adjacent
save view c:\images\f2.ps
show link v lgap 0
show node adjacent
save view c:\images\f3.ps
show link lgap 1
show node adjacent
save view c:\images\f4.ps
show link lgap 2
show link lgap 3
show node adjacent
save view c:\images\f5.ps
```

FIGURE 4.5: Command sequence to create images in figure

commands in figure 4.5 show the two backbone nodes. The third command saves that part of the network that is currently in view as a postscript file. The links that connect the two backbone nodes to nodes one layer lower are shown with the fourth command. The 'lgap' is short for layer gap. Should we want to show links going up one layer instead of down we can use 'lgap -1' instead of 'lgap 1'. The fifth command shows all the nodes at the endpoints of every link. Line seven contains the extra option 'v' before 'lgap' this ensures that only links to nodes that are within the viewing area are shown. These commands make it very easy to visualise parts of the network, or certain types of connections.

Next to showing nodes and links it is also possible to show the VPNs that are defined over the links. Because VPNs are directional they are drawn with a colour gradient for each hop. The gradient indicates in which direction the forward path is defined.

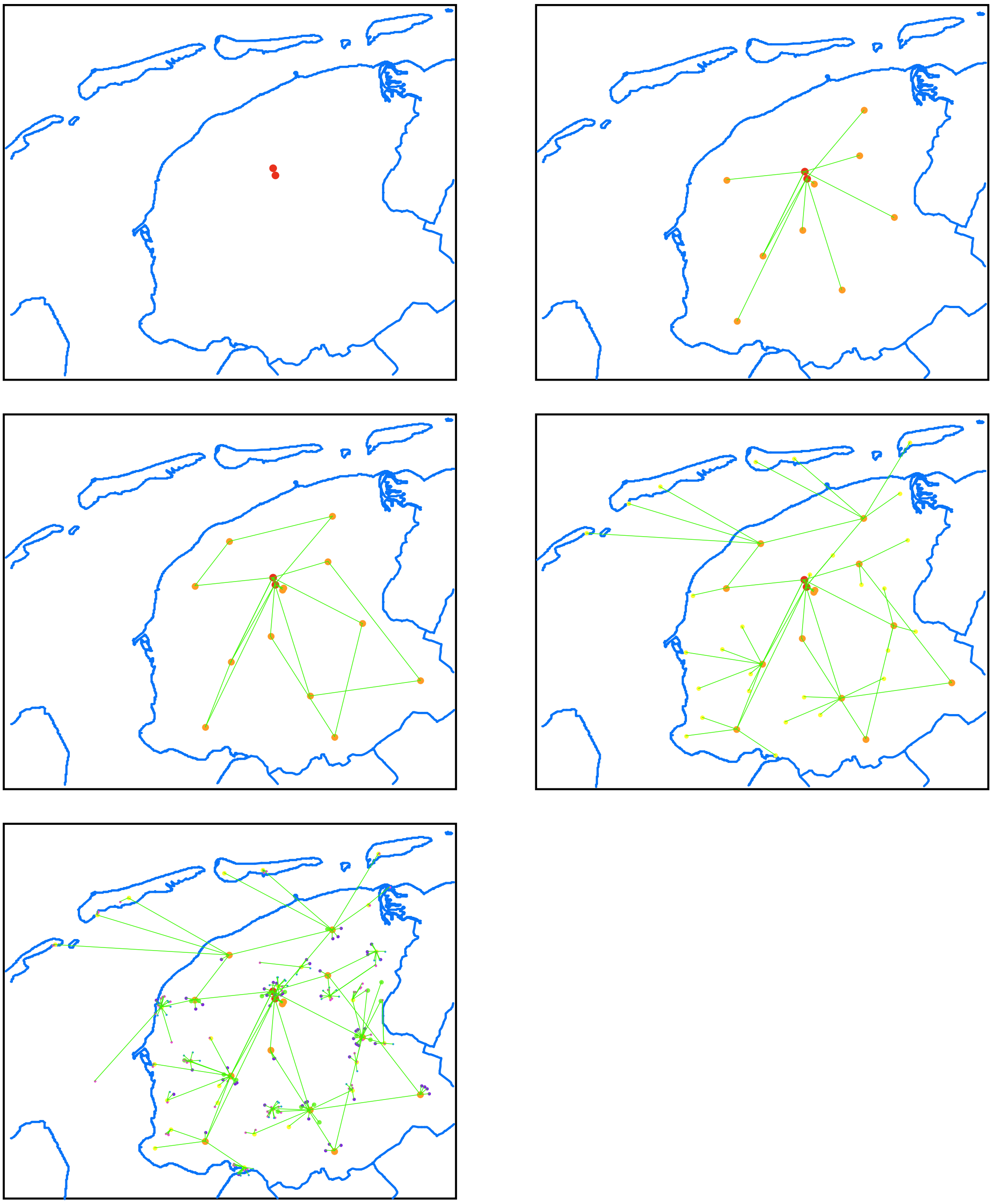


Figure 4.6: Layer relations

4.4.2 Network Metrics

For the network that is shown in the view window various network metrics can be calculated. The following metrics can be calculated in the program directly: path length based metrics such as closeness, eccentricity, diameter and average distance, clustering coefficient, assortativity, node degree and betweenness centrality. The results of all these calculations can be put in the command window or either be written to a file or to the clipboard. The last function makes it easy to combine the tool with for example Matlab for plotting or other functions (one drawback of Matlab is that it cannot handle large amounts of data on the clipboard very well). To calculate other metrics such as the algebraic connectivity the program can create the adjacency matrix or list or the Laplacian of the network. Just as for the calculations of the network metrics, the matrices can be saved to file or to the clipboard.

4.4.3 Simulation

The tool also features an integrated simulation part. For every VPN that is loaded in the network a generator can be attached to one or more of the nodes in the hop sequence. These generators can be added via the command window but this quickly becomes tedious for larger numbers of vpns. To overcome this a graphical approach is implemented. A window containing three scrollable list is displayed as is shown in figure 4.7. The first contains the first endpoints of the VPNs while the second contains the second endpoints, the third list contains the VPN ids. All three lists are linked and scroll together. In the first two lists the elements are also linked together. Only one can be selected, selecting the corresponding element in the other list will deselect the element. The selected element in the first two list acts as the source of the traffic while the other element is the destination. If the element in the third row is deselected no generator will be added to that VPN.

When the generators are started simulated traffic will start to flow over the VPNs that have a generator added to them. By clicking a location in the map view a meter can be added for a particular node. The meter will be displayed in the data window. Checking the 'log' box in the meter will tell it to save the values to a file. It is also possible to log all the nodes that are part of a VPN. This log file has a column for each hop and

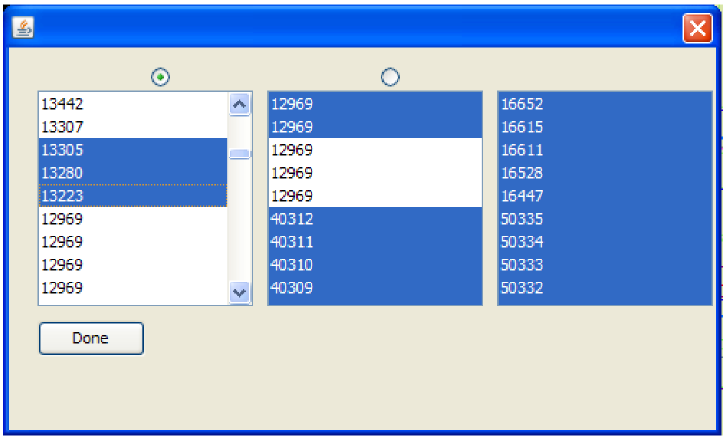


FIGURE 4.7: Window for adding generators

will log the traffic per hop. For both the meters as the overall log the simulated traffic is aggregated over all the VPNs that are defined between two nodes.

Chapter 5

Simulation

In order to have a good view of the demands that will be made on various parts of the network, it is necessary to be able to simulate future traffic loads. In this chapter the tool described in chapter 4 will be used to simulate traffic for three different scenarios over the coming five years.

5.1 Scenarios

The scenarios that are used in this chapter are based on KPN's traffic forecast for the consumer market. In these forecasts internet traffic and unicast IPTV traffic is considered. In this chapter the term IPTV is used to denote video traffic as requested by a customer's set top box. Although video content, together with peer to peer file sharing traffic, currently makes up for most of the internet traffic, IPTV does not include video web content.

The forecasts are based on four parameters: the number of internet customers, the number of TV customers, the peak bandwidth usage per internet customer and the peak bandwidth usage per TV customer. These four parameters are a function of both time and the type of scenario. For each year from 2011 up to 2015 a value for the four parameters is given for three different scenarios: low, medium and high. The three scenarios can be set for both the customer development and the bandwidth development, leading to nine different combinations. In order to simulate both extremes, while minimising the number of simulations, the scenario for customer development and bandwidth development are taken to be the same. Tables 5.1, 5.2 and 5.3 record the parameters for the three scenarios.

	2011	2012	2013	2014	2015
Customers Int.	2,656,298	2,507,238	2,358,179	2,209,119	2,060,060
Customers IPTV	341,667	621,250	943,688	1,266,125	1,588,563
BW Int.	115	149	194	252	328
BW IPTV	220	264	308	440	550

Table 5.1: Scenario Low

	2011	2012	2013	2014	2015
Customers Int.	3,020,298	3,053,238	3,086,179	3,119,119	3,152,060
Customers IPTV	569,444	1,035,417	1,572,813	2,110,208	2,647,604
BW Int.	133	187	261	366	512
BW IPTV	275	330	462	660	990

Table 5.2: Scenario Medium

	2011	2012	2013	2014	2015
Customers Int.	3,379,736	3,714,103	4,012,277	4,186,921	4,449,755
Customers IPTV	813,492	1,479,167	2,246,875	3,014,583	3,782,292
BW Int.	153	230	344	516	775
BW IPTV	440	660	990	1,320	1,650

Table 5.3: Scenario High

The simulations are based on the current routing of VPNs over the Ethernet network and the current distribution of end connections over the country. The customer development is based on a market share that is averaged over the country. For each simulation the total expected traffic per VPN link is determined. When multiple VPNs are defined between the same nodes, this traffic is combined. The expected traffic is compared to the currently installed capacity. The result of this comparison is graphically displayed in the tool. For each link in the VPN the total needed capacity is indicated next to it, together with the installed capacity in brackets. The link is drawn either in green, orange or red, indicating that the installed capacity is sufficient, marginally sufficient or insufficient respectively. In figure 5.1 an example of this output is given.

In addition to the visual information, the number of links that do not have sufficient capacity in each simulation is stored. For each stored link it is also recorded between which node layers the link is made, i.e. whether it is a link between two MC nodes or

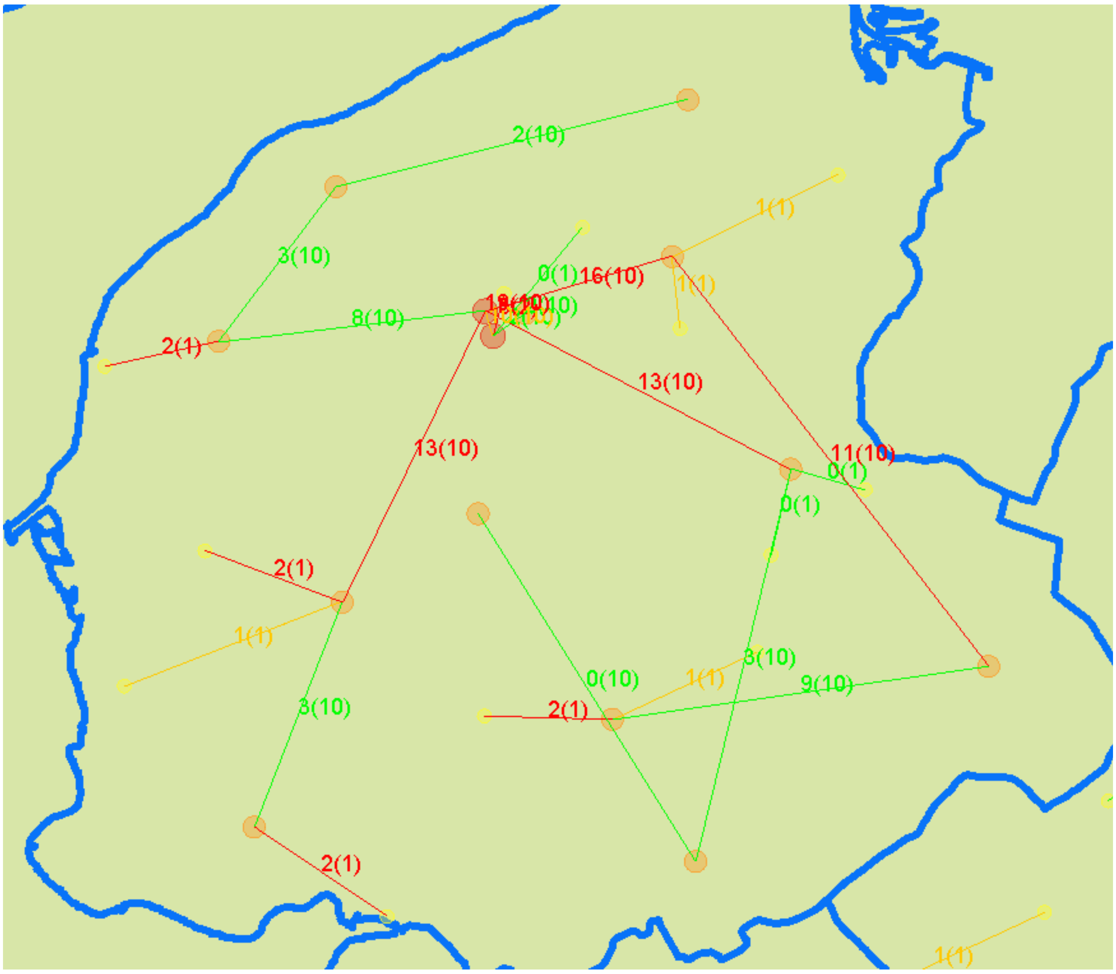


FIGURE 5.1: Graphical output

between an MC and MB node. The total traffic between each backbone location and the WAP location is also recorded.

5.1.1 Network

The network that is considered in these simulations consists of those nodes that are part of the VPNs that route traffic to customers. Not all nodes are taken into consideration, but only those nodes at the MB layers and upwards. Table 5.4(a) records the number of nodes of each layer, while table 5.4(b) records the different links. It should be noted that links from the BB locations to the WAP locations are treated differently from the rest.

Node type	Number of nodes
MB	478
MC	230
BB	23
WAP	2

(a) Number of nodes per type

Link Type	Number of links
MB-MC	478
MC-MC	146
MC-BB	85

(b) Number of links per type

Table 5.4: Node and link types

5.2 Simulation Results

Figure 5.3 on page 45 shows the number of links that get more traffic in the simulation than is currently installed. The sub-figures 5.3(a), 5.3(b) and 5.3(c) show the same data for the three different scenarios. The overloaded links are grouped by type. In each figure a line indicates the percentage of link types that are overloaded. By overloaded we mean that the traffic that goes over a link in the simulation is larger than the currently installed capacity.

5.2.1 Scenario Low

5.2.1.1 MB-MC links

The links connecting MB nodes to their MC nodes are the least affected by the increasing traffic load. In the low scenario the first overloaded link is to be expected in 2013. From 2013 onwards, the number of overloaded MB-MC links will increase quickly from thirty in 2014 to triple that in 2015. Although the number of overloaded MB-MC links will reach almost a hundred in 2015 this is still only twenty percent of the total number of MB-MC links. Figure 5.4(a) shows the average and maximum capacity shortage for the various types of links in the low scenario. From this figure it can be read that by 2015 the average capacity shortage on MB-MC links will be around 1.6 Gb/s while the maximum shortage will be around 5 Gb/s. From these figures it can be concluded that the MB to MC part of the network can cope with the expected customer and bandwidth developments. A number of links have to be upgraded from 2013 onwards but the needed extra capacity can be satisfied with extra 1 Gb/s links or an occasional 10 Gb/s link.

5.2.1.2 MC-MC links

Links between MC nodes generally develop similarly to MB-MC links. The main differences are that the increase in overloaded links from 2013 onwards is far more regular. Each year approximately twenty links will have insufficient capacity compared to the present situation. By the end of the five year period some thirty-six percent of the MC-MC links will have to be upgraded. The average extra capacity that is needed lies around the five Gb/s, the maximum extra capacity that is needed on a single link steadily increases from seven to thirty-seven Gb/s between 2013 and 2015. The ratio between the average and maximum needed extra capacity is greater for the MC-MC links than for the MB-MC links. This is mainly caused by the fact that most MC nodes are connected in a open ring structure. As a result, all traffic from the nodes towards the end of this opened ring has to pass through the nodes closer to the BB location. This situation can also be seen in the top part of figure 5.1 where three MC nodes (the orange dots) are connected via each other to a BB node. Although some links will need quite a few extra interfaces, the extra capacity can be installed using the same ten Gb/s interfaces that are used now.

5.2.1.3 MC-BB links

For these links the same situation applies as for the MC-MC links. Most of the MC-BB links are the first link in a string of MC nodes and as a result have to transport all the traffic for the other nodes as well. The number of overloaded MC-BB links increases faster than the number of overloaded MC-MC links. The average extra capacity that is needed per overloaded MC-BB link is also higher than for the MC-MC links. By the end of 2015 over sixty percent of the MC-BB links will have to be upgraded by, on average, more than ten Gb/s. The most heavily used MC-BB links need an extra capacity of forty Gb/s. These capacity improvements can be made by installing extra ten Gb/s interfaces, just as in the MC-MC link case.

5.2.2 Scenario Medium

5.2.2.1 MB-MC links

In the medium scenario the number of overloaded links between MB and MC nodes increases far more quickly than in the low scenario. From 2012 onwards almost a hundred links will become overloaded per year. At the end of the five year period almost eighty percent of all the MB-MC links will have to be upgraded. However, the extra capacity that is needed on average per link remains below five Gb/s. Only after 2014 some MB-MC links will have a capacity shortage of more than ten Gb/s compare to the currently installed links.

5.2.2.2 MC-MC links

The number of MC-MC links that will become overloaded over the years in the medium scenario is more or less constant from 2012 onwards. Each year thirty more links will have insufficient bandwidth. The average capacity deficit will stay below the 10 Gb/s until 2014 and will then increase till twenty-three Gb/s in 2015. The most heavily used MC-MC links will, however, be needing capacity upgrades in the order of 120 Gb/s by the end of 2015. It will require a technology change to 100 Gb/s Ethernet to provide these kinds of capacity boosts. Apart from a switch to higher speed interfaces it will be worthwhile to investigate changes to the network topology. The ring structures that are efficient at lower traffic loads are no longer efficient at these kinds of traffic loads. Just as the MB-MC links, eighty percent of the MC-MC links will have insufficient capacity by 205 when compared to the current situation.

5.2.2.3 MC-BB

The traffic load on the MC-BB links will increase so much over the five years in the medium scenario that ninety-five percent of them will be overloaded by 2015. The increase is rather steady. From 2011 onwards some twenty to twenty-five MC-BB links will become overloaded per year until they are almost all overloaded. That fact that almost all MC-BB links are overloaded towards the end of the simulation is the result of the previously mentioned ring structure.

5.2.3 Scenario High

5.2.3.1 MB-MC links

Because of the quick pick up of IPTV customers in the high scenario the MB-MC links will overload quite quickly. By 2013 almost seventy percent of them will be overloaded. The total number of overloaded MB-MC links will reach ninety-four percent by 2015. The average lack in capacity increases from one to three Gb/s in the first three years, while the maximum capacity lack increase over that period ranges from two to fifteen Gb/s. The majority of these capacity increase can be done with one Gb/s links while some MB nodes will need ten Gb/s interfaces. After 2013 the capacity demands will, however, increase to such an extend that in 2015 the average capacity demands will be around the ten Gb/s per link. The most heavily used MB node will need a capacity of about fifty Gb/s. Towards the end of the five year period almost all the MB nodes will have to be connected through ten Gb/s interfaces.

5.2.3.2 MC-MC links

Just as for all the other link types in the high scenario, the MC-MC links will almost all be overloaded by 2015. The more interesting factor in this scenario is how much mroe bandwidth is needed. From 2012 onwards the average needed capacity increase will be more than ten Gb/s, increasing steadily to almost sixty Gb/s. The capacity boost needed for the most heavily used links will increase from almost fifty Gb/s to nearing three hundred Gb/s. As already mentioned in the medium scenario, this kinds of bandwidth increments are only feasible with 100 Gb/s Ethernet interfaces and topology changes.

5.2.3.3 MC-BB links

For the MC-BB links the situation is comparable to that for the MC-MC links. From 2013 onwards almost all the MC-BB links need more capacity than is installed at the moment. The maximum bandwidth increase that will be needed for the MC-BB links is a steady ten percent more than for that of the MC-MC links. This can also be explained by the ring structures that have been discussed earlier. The average increase for the MC-BB links is a little lower than for the MC-MC links in the first year but will increase to about sixty-five percent more than the MC-MC links towards 2015. Just as for the MC-MC links this kind of capacity can only be feasibly guaranteed by higher interface speeds and more efficient topologies.

5.2.4 Backbone traffic

Figure 5.5 on page 47 shows the total traffic from the backbone locations to the WAP location. Towards the right hand side of the figure the total traffic is given. The total traffic is the sum of the traffic from each backbone location individually. Note that the vertical axis at the right hand side of the figure indicates the scale for the total traffic while the axis at the left hand side of the figure indicates the scale for the individual backbone locations. As can be seen in these figures, the maximum traffic load between an individual backbone location and the WAP location varies from a few hundred Gb/s in the low scenario to more than a Tb/s in the high scenario. The topology of the optical transport network is given in figure 5.2. Since the capacity of the optical transport network is considered to be sufficient to facilitate the Ethernet network and no changes to the ring structure of the network are to be expected this network can be considered as a given. Depending on the scenario, the total amount of traffic for IPTV services will be greater than that of internet services by the year 2012 or 2014. Because of this, it can be hugely beneficial to place content servers at multiple backbone locations in the network. For internet traffic the load on the backbone rings can be eased by introducing caching servers.

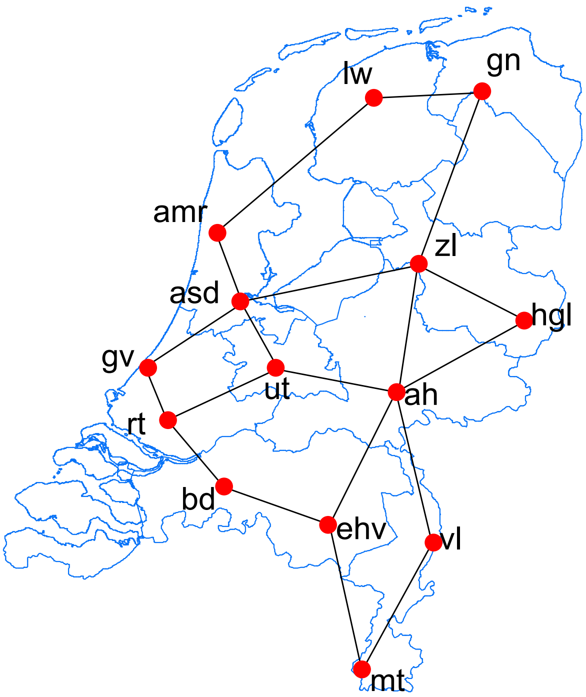
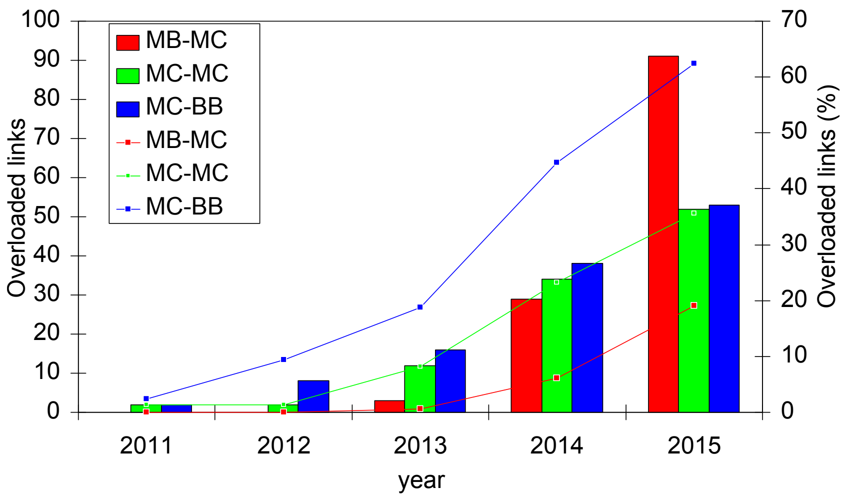
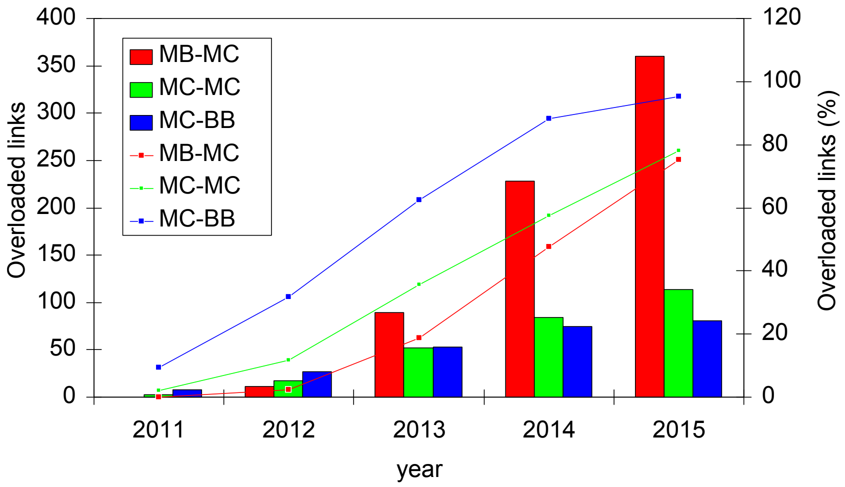


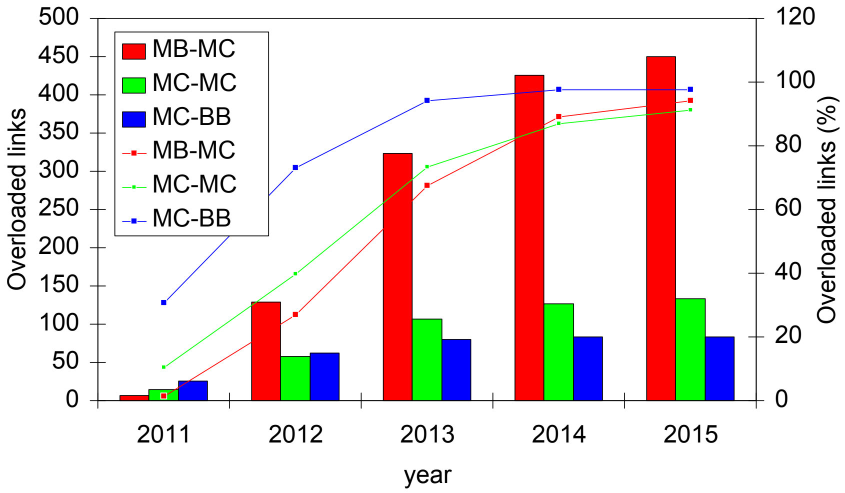
Figure 5.2: Optical Transport Network



(a) Low

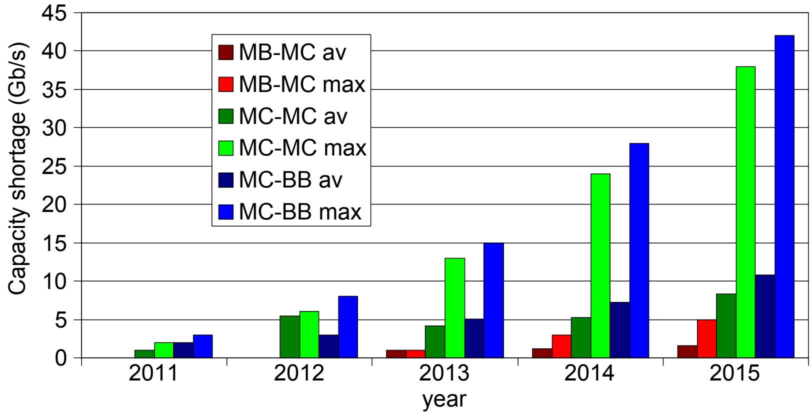


(b) Medium

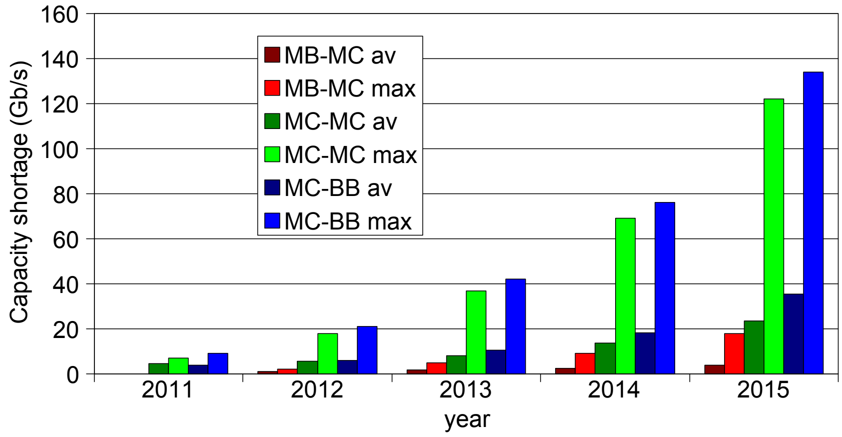


(c) High

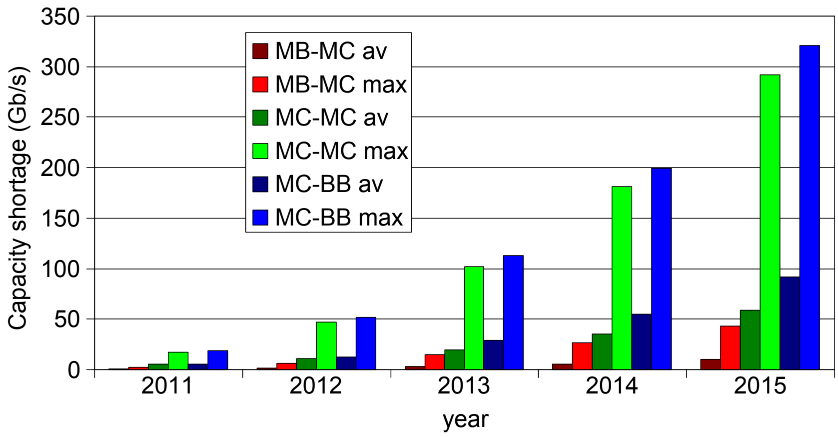
FIGURE 5.3: Number of links that are overloaded for three scenarios. The lines in each figure show the percentage of links of that type that is overloaded.



(a) Low

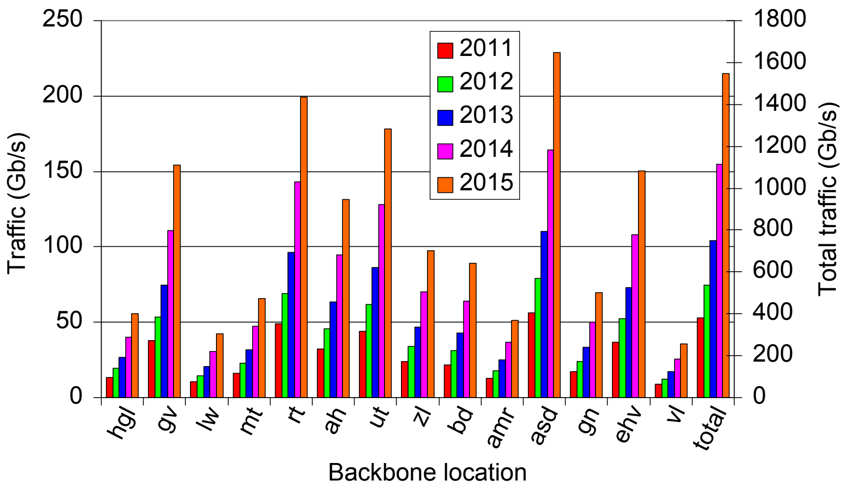


(b) Medium

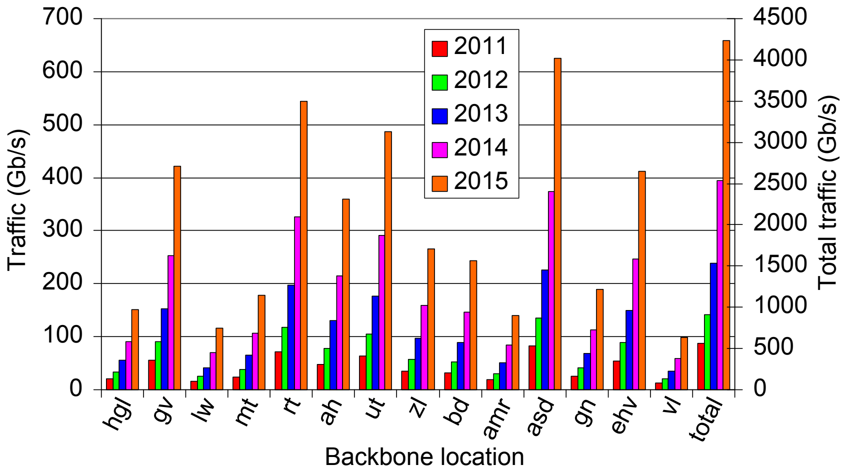


(c) High

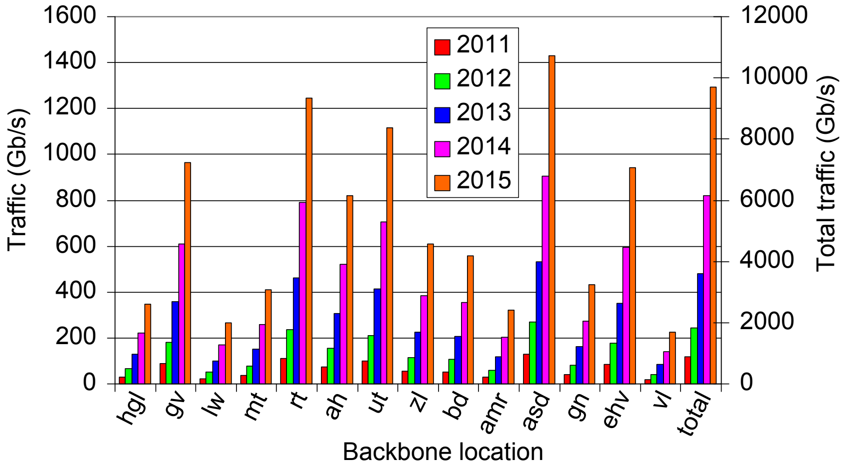
FIGURE 5.4: Average and maximum capacity shortage per link type for three scenarios.



(a) Low



(b) Medium



(c) High

FIGURE 5.5: Traffic from backbone locations to WAP location over a period of five years for three scenarios. The total traffic for each scenario is given at the right hand side of the figure.

Chapter 6

Topology Optimisation

The number of links that have to be upgraded is, in fact, larger than the number of links that have become overloaded in the simulations. This is because in the Ethernet topology every VPN has both a primary and a backup path assigned to it. During the simulation the VPNs transporting consumer market Internet, TV and telephony traffic were used. Because the simulation reproduced the normal operational state of the network, only primary paths were included. When the capacity of the links in the primary path is not large enough, it is likely that this will also be the case for the links in the backup path. As a result, both the links in the primary as well as the links in the backup path have to be upgraded.

Upgrading the link capacity is not the only way to ensure that the traffic demands can be satisfied in the future. It is also possible to alter the Ethernet topology. Because the Ethernet network is a layer on top of the optical transport network, the Ethernet topology is only partly dictated by the physically optical network. In the optical network add-drop multiplexers and photonic cross connects make it possible to define light-paths that can connect Ethernet switches directly to each other that are not physical neighbours. In other words, the topology as seen by the switches is not bound to the physical topology of the fibre optic cables. In [10] the authors distinguish three types of networks: opaque, transparent, and optimised. In an opaque network the overlay topology, in our case the Ethernet topology, follows the optical topology. Only switches that are directly connected by an optical fibre are neighbours. In the transparent network all switches are connected to each other, whether they share a direct fibre connection or not. The optimised network is a combination of opaque and transparent links. Only those transparent links that divert transit traffic from the Ethernet layer to the optical layer are used. Bypassing Ethernet switches for transit traffic decreases the load on the switch and diminishes the number of Ethernet ports that are needed. The optimisation is a trade-off between extra light-paths on the optical level and fewer ports on the Ethernet level.

The topic of optimising multi-layer networks is well studied in literature. In [11] the authors optimise a two-layer network in three steps. First the opaque network is optimised to include only paths that carry traffic. In the second stage switches that carry more transit traffic than a certain threshold are bypassed by adding a link in the lower layer topology between the previous and next switch. In the third and final stage the network derived in step two is cost optimised by varying the threshold for transit traffic. In [12] multiple logical topologies are compared for throughput and costs. The topologies include the opaque, transparent and combinations of opaque and transparent topologies created by three heuristics. It shows that in terms of throughput the transparent topology is best but that it comes at a high price in the optical network. It is also not very scalable because of the many light-paths that are needed. All the authors used Integer Programming or Linear Integer Programming or Mixed Integer Programming techniques to solve their problems. In [13] the authors use an Analytic Hierarchy Process to design a topology that can be optimised for multiple hierarchically ordered parameters. The analytic hierarchy process is a method the make rational decision concerning a problem by evaluating multiple criteria that are placed in a hierarchical structure. Their approach allows them to use multiple optimisation parameters with relative importance.

6.1 Optimisation of Ethernet Topology

In the simulations performed in the previous chapter the traffic flows had a rather simple shape. All traffic originates at the WAP location in Amsterdam and flows to one of the MB or MC locations in the network. The simulation results are basically a traffic matrix stating the demands of every node from the MB nodes upward to the central WAP location. This traffic matrix can be used to optimise the Ethernet topology for the traffic demands taken from the simulations of different scenarios. In this section we will use a linear integer program to find link disjoint primary and backup traffic routes for each node while minimising the total link capacity. In theory it is possible to make a transparent Ethernet network, but since the number of Ethernet nodes is rather big this would place an infeasible demand in terms of light-paths on the optical network. Apart from the demand on the optical network, the number of possible combinations of paths in the Ethernet network would also be very high and, as a result, it would take very long to solve the linear program. A further complication to optimising the whole Ethernet network in one go is that the scope of the tool described in this thesis is the Ethernet layer and does not include the optical layer. This makes it difficult to include reasonable costs and resource limitations of the optical network in the optimisation. In the light of these limitations the optimisation proposed here optimises the capacity usage of the individual MC rings. Because of the rather simple shape of the traffic flows in the simulation, optimising all the MC rings individually would lead to a reasonably optimised network. The possibility to include bypasses between different MC rings of the same backbone location is not considered. Bypasses between different rings are ignored because of the shape of the traffic matrix. Since traffic flows from a backbone location to a location on a ring and rings only join in backbone locations, a bypass between rings would only lead to a loop on the optical layer. If the traffic demands also include traffic between MC locations, however, inter ring bypasses can optimise the overall network.

6.1.1 Model

In the optimisation process the MC rings have been slightly simplified. The structure of the actual fibre rings is illustrated in figure 6.1.

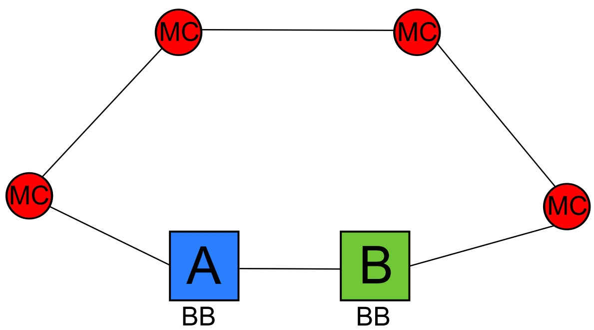


Figure 6.1: Fibre MC Ring

As can be seen in figure 6.1, the fibre rings connect the MC locations (red circles) to two backbone locations (blue and green squares), one backbone location is connected to the A-net, the other is connected to the B-net. In the optimisation the two backbone locations are placed in the same node. This leaves only one backbone location as the origin of the traffic and various MC locations as destinations. Each MC location has a certain traffic demand that was determined during the simulation. For a each ring and traffic demand all possible links (the set L contains all the links in the network) on the Ethernet layer are taken into account. The integer program takes the following form.

Constants:

d_i	Demand of node i
M	Capacity granularity

Variables:

$x_{i,u,v}$	Binary variable. Is one when link (u, v) is in primary path i, zero otherwise.
$y_{i,u,v}$	Binary variable. Is one when link (u, v) is in backup path i, zero otherwise.

Constraints:

$\sum_{i,(u,v)\in L} x_{i,u,v} = 1,\ v = \text{BB}$	Guarantees a primary path to BB is allocated.
$\sum_{i,(u,v)\in L} y_{i,u,v} = 1,\ v = \text{BB}$	Guarantees a backup path to BB is allocated.
$\sum_{v:(u,v)\in L, u\neq i} x_{i,u,v} \leq \sum_{v:(v,u)\in L, u\neq i} x_{i,u,v}$	No more primary paths leave a node than enter it.
$\sum_{v:(u,v)\in L, u\neq i} y_{i,u,v} \leq \sum_{v:(v,u)\in L, u\neq i} y_{i,u,v}$	No more backup paths leave a node than enter it.
$x_{i,u,v} + y_{i,u,v} \leq 1$	Primary and backup path are disjoint.
$\sum_{i} x_{i,u,v} d_i + x_{i,u,v} d_i \leq M C_{u,v}$	Link is not overloaded.

Objective:

$$\text{Minimise}(F = \sum_{(u,v)\in L} C_{u,v})$$

6.1.2 Results

This program was solved with the open source mixed integer linear programming solver LpSolve[1]. As an example two MC rings are included here. One is a heavily used ring centred at the Amsterdam backbone location, the other one is a fairly lightly used ring around Leeuwarden. For the three scenarios simulated in the previous chapter the optimal link capacities are determined for the year 2012 and the year 2015. The demands

[1] see http://lpsolve.sourceforge.net/5.5/ for more information and download of LpSolve.

for the three scenarios per year per node in Gb/s are given in table 6.1 and 6.2 for the rings around Amsterdam and Leeuwarden respectively.

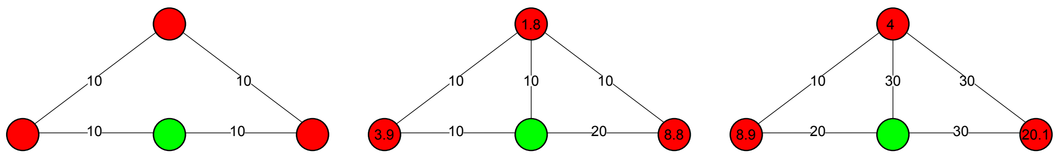
	Low		Medium		High	
Node	2012	2015	2012	2015	2012	2015
1	4.8	14	8.2	38.2	16.5	87.4
2	2.7	7.9	4.6	21.5	9.3	49.2
3	1.6	4.6	2.7	12.6	5.5	28.9
4	3.3	9.5	5.6	25.9	11.2	59.3

TABLE 6.1: Demands for the Amsterdam ring nodes in GB/s for the different scenarios.

	Low		Medium		High	
Node	2012	2015	2012	2015	2012	2015
1	0.5	1.4	0.8	3.9	1.7	8.9
2	0.2	0.6	0.4	1.7	0.8	4
3	1.1	3.2	1.9	8.8	3.8	20.1

TABLE 6.2: Demands for the Leeuwarden ring nodes in GB/s for the different scenarios.

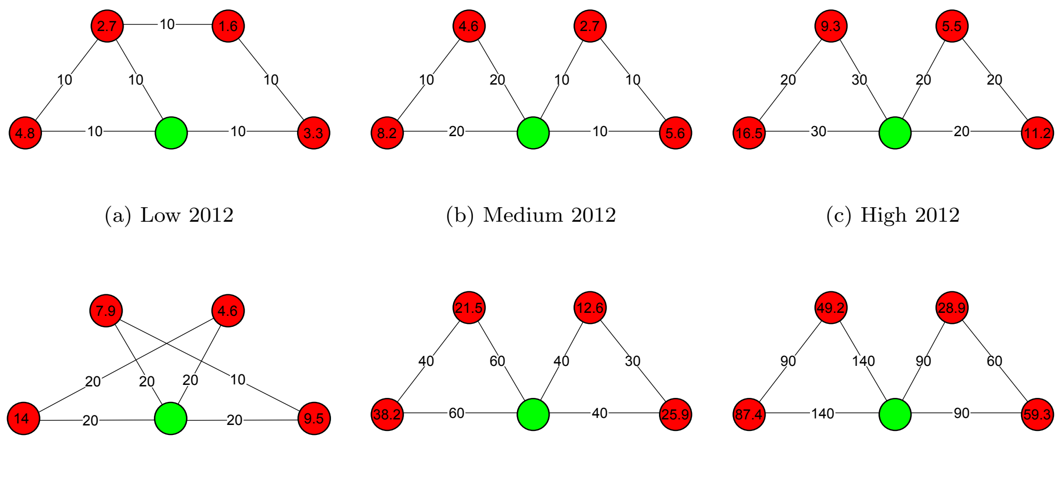
The results of the optimisations are illustrated graphically in figures 6.2 and 6.3. In these figures the backbone locations is always indicated by the green circle. Metro Core locations are indicated by the red circles and their traffic demand in Gb/s is drawn in the circle. The Ethernet links that are used are indicated by the lines connecting the nodes. The numbers in the links indicate the capacity that is needed. As the linear program was solved for ten Gb/s links, the capacity assignment is in multiples of ten.



(a) Standard (b) Medium 2015 (c) High 2015

FIGURE 6.2: Capacity assignments for Leeuwarden ring.

As can be seen from figure 6.2, an Ethernet topology that follows the underlying ring structure is best for low traffic loads. As Leeuwarden is a sparsely populated province, the current ring topology will be sufficient until 2015 for the low scenario. Figure 6.2(a) is called standard because it is the original ring structure and is sufficient in all situations except the following. In the medium and high scenarios extra capacity will be needed towards the year 2015. In both scenarios a bypass from MC node 2 (counting clockwise beginning at the backbone) to the backbone location is more efficient in terms of installed capacity than reinforcing the entire ring.



(a) Low 2012 (b) Medium 2012 (c) High 2012

(d) Low 2015 (e) Medium 2015 (f) High 2015

FIGURE 6.3: Capacity assignments for Amsterdam ring.

For the heavily used ring around Amsterdam the situation is quite different from that of the ring around Leeuwarden. Figure 6.3 shows the capacity assignment that optimises the Amsterdam ring. In the low scenario the ring is still the most efficient topology in 2012, albeit that a bypass from node two to the backbone location is needed. By the year 2015 the optimal topology is, however, completely changed. The topology almost exclusively consists of bypasses. The topology shown in 6.3(d) uses two interfaces less than the next best topology which uses the ring structure and two bypasses. Because the topology is optimised on link capacity, the number of bypasses has no influence on the selected topology. It is possible to assign a higher cost to bypasses than to additional capacity in the ring structure to avoid a high number of bypasses. The medium and high scenarios show that for increasing capacity demands the optimal topology consists of triangles. Two sides of the triangles are formed by links over the ring while the third side is a bypass. These triangles are basically small ring structures, and can also be recognised in 6.3(d). The rule for assigning the link capacity to the triangles is quite simple. The two sides adjacent to the backbone location should have enough capacity to accommodate the sum of the traffic for the two MC nodes while the side connecting the two MC nodes should have enough capacity to accommodate the traffic of the most demanding node.

6.2 Backbone

Because the traffic in the simulated scenario originates at a WAP location in Amsterdam, the Ethernet topology of the backbone network with its A-net and B-net is very well suited for this kind of traffic pattern. Each backbone location has a direct link to the WAP location. On the optical layer, however, many of these direct links will come together on the ring connecting the locations Amsterdam, Zwolle, Arnhem, and Rotterdam. As the contribution of IPTV content to the total amount of traffic will be more than half in the medium and high scenarios from 2013 onwards it can be beneficial to place multiple VoD servers in the network instead of serving all users from one location. Apart from cloning the VoD server to multiple locations a Content Distribution Network (CDN) can be used [14, 15]. The structure of a CDN often takes the shape of a central content server and multiple replica servers. The replica servers are located close to the end users and are filled with content by the central servers. Many techniques have been studied on how to efficiently select which content is stored at which replica server and where to place the servers. Just as a CDN structure can relieve some of the traffic from the backbone rings, web content can be cached [16, 17]. In the case of the Ethernet network this is, however, not directly applicable because the web caching takes place at the IP layer. In the Ethernet network studied here the IP layer is only available at the WAP location. The introduction of web caching would then also imply bringing the IP layer closer to the edge of the network.

6.3 Source Placement

Suppose we want to place an extra VoD server in a network. In the case considered here, the placement of the server does not involve adding a node to the network, it merely adds a function to an existing node. Which node we select as the information source has an impact on the flow of traffic through the network. It might be a good idea to minimise the total average distance from every node in the network to the source node in order to balance the traffic flow. Consider the network given in figure 6.4.

All links in figure 6.4 are bidirectional and have unit length, the node numbering is arbitrary. If we place the source at node number four, the traffic intended for nodes ten and twelve, for example, would have to cross the entire network. Indeed, node

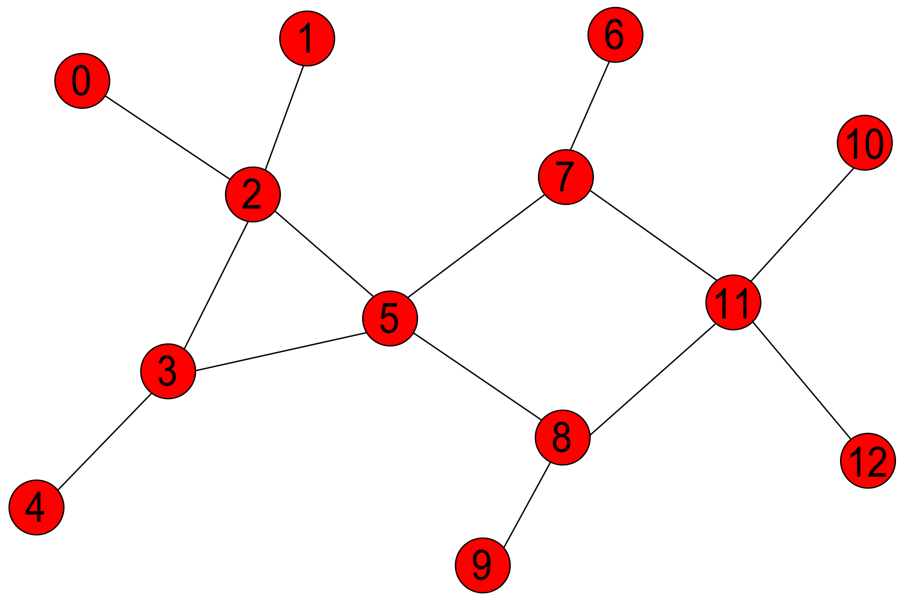


FIGURE 6.4: Sample Network

number four has the largest average distance to all the other nodes in the network and is, therefore, not a suitable candidate for a source node. The most suitable candidate for a source node is the node that has the lowest average distance to the other nodes in the network. In Figure 6.5 the distance matrix of the network shown in figure 6.4 is recorded. Element $d_{i,j}$ in a distance matrix contains the length of the shortest path from node i to j. The first separated column in figure 6.5 lists the node numbers, for easy reference. The final column lists the sum of all the entries in the row. In order to select the node with the smallest average distance to all the other nodes in the network we must find the row in the distance matrix that has the smallest sum of its elements. In the case of our sample network that would be row number six, indicated in blue. It turns out that node number five is the most suitable candidate for the function of source node.

Now suppose that we want to place two identical sources in the same network. We now need to minimise the average distance from every node to its nearest source node. If we now select those two nodes that have the smallest average distance to all the other nodes in the network (nodes five and seven, for example) we can determine the average distance to the source for the entire network by summing the minimal column values for row six and eight. The total distance would be $\min(2,3) + \min(2,3) + \min(1,2) + \min(1,2)+\min(2,3)+\min(0,1)+\min(2,1)+\min(1,0)+\min(1,2)+\min(2,3)+\min(3,2)+$

$$\left[\begin{array}{c|ccccccccccccc|c}
0 & 0 & 2 & 1 & 2 & 3 & 2 & 4 & 3 & 3 & 4 & 5 & 4 & 5 & 38 \\
1 & 2 & 0 & 1 & 2 & 3 & 2 & 4 & 3 & 3 & 4 & 5 & 4 & 5 & 38 \\
2 & 1 & 1 & 0 & 1 & 2 & 1 & 3 & 2 & 2 & 3 & 4 & 3 & 4 & 27 \\
3 & 2 & 2 & 1 & 0 & 1 & 1 & 3 & 2 & 2 & 3 & 4 & 3 & 4 & 28 \\
4 & 3 & 3 & 2 & 1 & 0 & 2 & 4 & 3 & 3 & 4 & 5 & 4 & 5 & 39 \\
5 & 2 & 2 & 1 & 1 & 2 & 0 & 2 & 1 & 1 & 2 & 3 & 2 & 3 & 22 \\
6 & 4 & 4 & 3 & 3 & 4 & 2 & 0 & 1 & 3 & 4 & 3 & 2 & 3 & 36 \\
7 & 3 & 3 & 2 & 2 & 3 & 1 & 1 & 0 & 2 & 3 & 2 & 1 & 2 & 25 \\
8 & 3 & 3 & 2 & 2 & 3 & 1 & 3 & 2 & 0 & 1 & 2 & 1 & 2 & 25 \\
9 & 4 & 4 & 3 & 3 & 4 & 2 & 4 & 3 & 1 & 0 & 3 & 2 & 3 & 36 \\
10 & 5 & 5 & 4 & 4 & 5 & 3 & 3 & 2 & 2 & 3 & 0 & 1 & 2 & 39 \\
11 & 4 & 4 & 3 & 3 & 4 & 2 & 2 & 1 & 1 & 2 & 1 & 0 & 1 & 28 \\
12 & 5 & 5 & 4 & 4 & 5 & 3 & 3 & 2 & 2 & 3 & 2 & 1 & 0 & 39
\end{array}\right]$$

FIGURE 6.5: Distance matrix of Sample Network. One source indicated in blue.

$\min(2,1) + \min(3,2) = 17$. This is, however, not the smallest possible value. Figure 6.6 shows that the smallest value is actually fourteen and is obtained when nodes two and eleven are used.

$$\left[\begin{array}{c|ccccccccccccc|c}
0 & 0 & 2 & 1 & 2 & 3 & 2 & 4 & 3 & 3 & 4 & 5 & 4 & 5 & 38 \\
1 & 2 & 0 & 1 & 2 & 3 & 2 & 4 & 3 & 3 & 4 & 5 & 4 & 5 & 38 \\
2 & 1 & 1 & 0 & 1 & 2 & 1 & 3 & 2 & 2 & 3 & 4 & 3 & 4 & 27 \\
3 & 2 & 2 & 1 & 0 & 1 & 1 & 3 & 2 & 2 & 3 & 4 & 3 & 4 & 28 \\
4 & 3 & 3 & 2 & 1 & 0 & 2 & 4 & 3 & 3 & 4 & 5 & 4 & 5 & 39 \\
5 & 2 & 2 & 1 & 1 & 2 & 0 & 2 & 1 & 1 & 2 & 3 & 2 & 3 & 22 \\
6 & 4 & 4 & 3 & 3 & 4 & 2 & 0 & 1 & 3 & 4 & 3 & 2 & 3 & 36 \\
7 & 3 & 3 & 2 & 2 & 3 & 1 & 1 & 0 & 2 & 3 & 2 & 1 & 2 & 25 \\
8 & 3 & 3 & 2 & 2 & 3 & 1 & 3 & 2 & 0 & 1 & 2 & 1 & 2 & 25 \\
9 & 4 & 4 & 3 & 3 & 4 & 2 & 4 & 3 & 1 & 0 & 3 & 2 & 3 & 36 \\
10 & 5 & 5 & 4 & 4 & 5 & 3 & 3 & 2 & 2 & 3 & 0 & 1 & 2 & 39 \\
11 & 4 & 4 & 3 & 3 & 4 & 2 & 2 & 1 & 1 & 2 & 1 & 0 & 1 & 28 \\
12 & 5 & 5 & 4 & 4 & 5 & 3 & 3 & 2 & 2 & 3 & 2 & 1 & 0 & 39
\end{array}\right]$$

FIGURE 6.6: Distance matrix of Sample Network. Two sources indicated in blue.

This example illustrates that the optimal way to place two sources in the network cannot be found by simply selecting those two nodes that have the smallest average distance to the rest of the nodes. In fact, the solution to the single source problem is not even part of the solution to the two source problem. In order to find the set of nodes that minimises the average distance to the nearest source, all the possible combinations of nodes have to be tried. The total number of combinations that have to be tried depends on the number of nodes n and the number of sources, s, that have to be placed. The number of different ways in which s elements can be chosen from a set of n elements without repetition is given by $\frac{n!}{(n-s)!}$. In this case, however, the solution set is not ordered and we have to divide by the number of different ways to order a set of s nodes, which is $s!$. Consequently, the total number of combinations that have to be tried before the minimum average distance can be found is given by $\frac{n!}{(n-s)!s!}$. Obviously it is not feasible to try all possible combinations to place a number of sources for a network that is larger than a few dozen nodes. For a network of a hundred nodes there are almost four million ways to place four sources, while for a network of a thousand nodes there are more than forty billion ways. Placing five sources in a network of a thousand nodes can be done in more than eight trillion ways.

The problem of placing sources or servers in a network is an example of the much studied k-median problem [18]. In the k-median problem the goal is to find a set of k nodes in a larger set of nodes that minimises the average distance from a node to the closest selected node. The k-median problem is a version of the facility location problem. The goal of the facility location problem is to find, for example, the locations of factories to most effectively serve a group of customers. Given a set of facilities F and clients D two different types of costs are considered in the problem. These are the cost to serve a client i from location j, given by $c(i, j)$, and the cost of opening facility j, given by $f(j)$. Many versions of the facility location problem exist, such as the capacitated facility location problem where each facility can serve a limited number of clients, and the the capacitated facility location problem with client demands where not only the facility can supply to a limited demand, but clients have varying demands as well. Often the service costs are assumed to be symmetric and satisfy the triangle inequality, i.e., $c(i, k) \leq c(i, i) + c(j, k)$ for any $i, j, k, \in D \cup F$. A much used technique to find approximations to the k-median problem is to formulate the problem as an integer program. The linear program that can be obtained from the integer program by dropping the integer constraint can then be solved in polynomial time. Since the original problem was an integer problem the results from the linear program have to be rounded in such as way that the integer solution is as close as possible to the optimum linear solution. Apart from rounding a linear program, local search and greedy algorithms can also give good approximations to the problem. In [19] the authors list several approximation algorithms for the uncapacitated facility location problem with approximation factors ranging from 3.16 to 1.52.

In this section we follow a different approach to the k-median problem. One way to at least curb the staggering number of combinations in larger networks is by not taking all nodes into account. Not every node is an equally likely candidate to be a source node. In the network of figure 6.4, for example, nodes zero, one, four, six, nine, ten and twelve are not likely to be good source candidates, because they are endpoints. If they are left out in the search process, the number of combinations that has to be checked is greatly reduced. We could filter candidate nodes on their degree before testing the different combinations, but this method has a few drawbacks. It is quite possible that a good candidate node has a low node degree and will therefore not be taken into consideration. It is a better idea to rank candidate nodes by their betweenness centrality. The betweenness centrality is a measure for how many of the shortest paths from all nodes in the network to all other nodes go through a certain node. A node that has a high betweenness centrality can be thought of as being positioned in the middle of the network or a region of the network. As such, a node with a high betweenness centrality is likely to minimise the average distance from the nodes it lies in the middle of to a source.

In order to test this idea the following experiment was carried out. Three different types of random networks were generated. Of these networks the nodes were sorted based on their betweenness centrality. Subsequently various numbers of sources were placed in the network in such a way that the average distance from all the nodes to their nearest source was minimised. For all the nodes that were chosen as source their position on the sorted betweenness centrality list was recorded. If there were multiple sets of nodes that were equally good solutions the nodes from all these sets were taken into account. To avoid distortion, each generated network was tested for connectedness. Only connected networks were taken into consideration. The following random networks were tested:

- Erdös Rényi (ER) graph of a hundred nodes with a link probability of 7 percent.
- Watts and Strogatz Small-World (SW) graph of a hundred nodes with an average node degree of two and a rewire probability of fifty percent.
- Barabási and Albert (BA) scale-free graph with an average degree of two.

For all the networks the position in the sorted betweenness centrality list of the source nodes was recorded for source sets of $s = 2$, $s = 3$ and $s = 4$ over three hundred iterations.

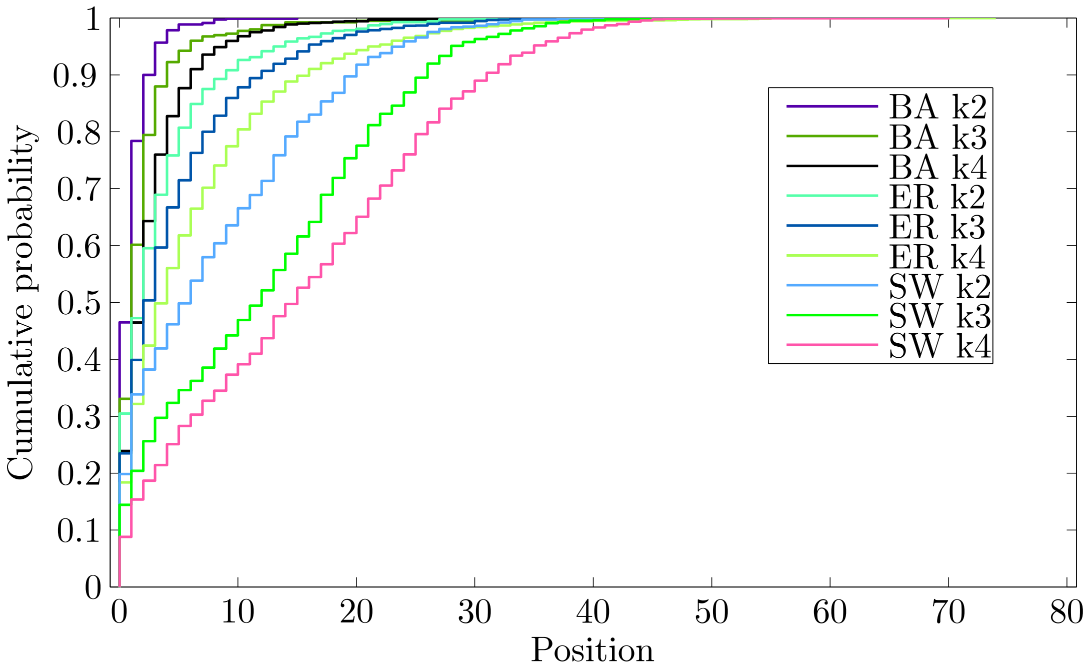


FIGURE 6.7: Cumulative Probability for three different graph types and two, three and four sources.

Figure 6.7 shows the cumulative distribution function of all three graph types for the various number of sources. The curves indicate for each graph what the probability is that the selected sources are ranked higher than a certain position on the sorted betweenness centrality list. It can be seen that for the small-world graphs the positions of the sources are less concentrated near the top of the list. This can be a property of small-world graphs, although it might be better explained by the possibility that the graphs that were generated still contained much of the original lattice structure in them. In a regular lattice each node is equally well suited to be a source. At the same time all nodes in a regular lattice have the same betweenness centrality score. As every node in a regular lattice is equally likely to be a source the positions of the sources in the betweenness centrality list is more or less evenly distributed. For the ER graphs it can be seen from figure 6.7 that the source nodes are far more likely to come from the top of the betweenness centrality range than from the bottom. Almost all sources rank thirty or higher. For the scale-free graphs sources are even more likely to come from the nodes with a high betweenness centrality. Figure 6.8 again shows the cumulative distribution functions for the ER and BA graphs but zoomed in. The lines for the same graph type but different numbers of sources follow the same path yet are displaced by a number of positions. The displacement is the result of the fact that in order to place more sources more nodes are needed. For the scale-free graphs the displacement is about two positions

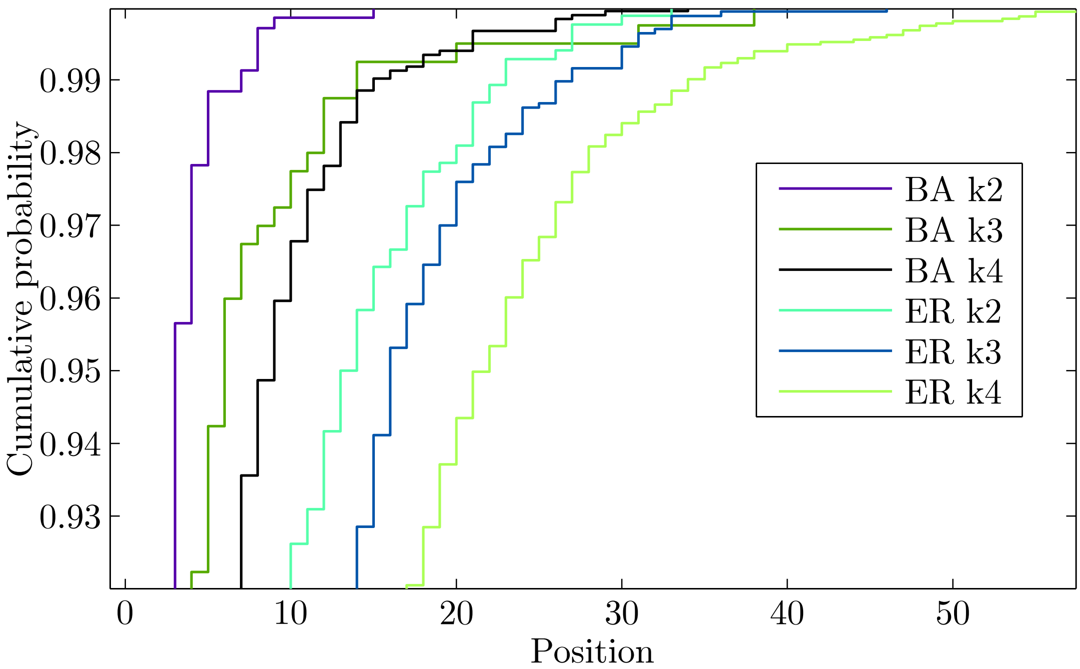


Figure 6.8: Cumulative Probability for ER and BA graphs

per extra source node. For the ER graphs the displacement is three positions between two and three sources and five positions between three and four sources. In order to have a chance of 98 percent that the optimal source locations will be found we only have to select the twenty to twenty-seven nodes with the highest betweenness centrality in the case of an ER graph for two to four sources. For a BA graph the best four to twelve nodes are sufficient. In the case of a graph of a hundred nodes this means that instead of the almost four million combinations we have to try only between 500 and 17500 combinations depending on what graph type we are considering. This is a significant improvement. Further tests show, however, that the number of nodes we need to select as candidates does not depend on the size of the network but only on the number of sources we wish to place. That is, if the number of sources is small in comparison to the number of nodes. This can be seen in figure 6.9. In this figure the cumulative distribution functions for two ER graphs with the same properties besides the number of nodes. For both graphs the positions of two sources on the sorted betweenness centrality list are recorded over a hundred iterations. It can be seen that the curve for the network with a thousand nodes lies slightly above the curve of the one hundred node network. This means that the same number of candidates are needed for a network of a thousand nodes as for a network of a hundred nodes.

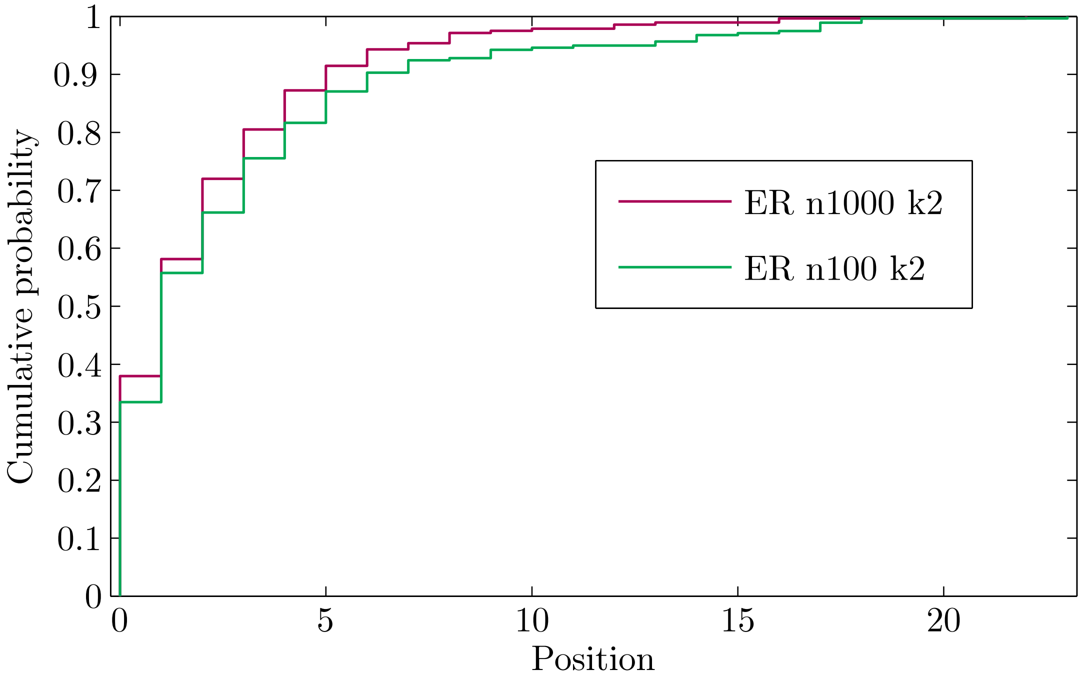


Figure 6.9: Cumulative Probability for ER graphs of 1000 and 100.

Chapter 7

Conclusion

In this thesis the KPN's Ethernet network was described. The basic structure is based on fourteen backbone locations that are connected to each other in ring structures. From these backbone locations rings depart which connect the Metro Core locations to the backbone locations. The Metro Core locations connect to multiple Metro Bridge and Metro Access locations which in turn connect to the local loop nodes in the Ethernet network. When this network is analysed in the light of complex network studies the first thing that springs to the attention is the very low link density. The full graph of the Ethernet is a very sparse graph. As a result, the algebraic connectivity is very low. For sub-graphs such as the backbone network this is not the case. For these graphs the link density is higher and the algebraic connectivity is so too. Especially the network of locations displays small-world and scale-free properties. All sub-graphs show disassortative degree mixing, which coincides with Newman's observation that man-made networks tend to have a negative assortativity constant.

In order to get a better overview of the network and a better understanding of the demands made on the network by the various services transported over it, a software was developed. This tool can display a geographically correct representation of the network. This representation is augmented by traffic statistics for all links and routing information for the VPNs that are defined over the network. For each VPN carrying the traffic for a certain service it is possible to generate simulated traffic and log the load that is placed on all the nodes and links. During a simulation the load on the links is compared to the installed capacity and the links are shown in either red orange or green depending on whether the installed capacity is sufficient or not.

With the help of the tool three different internet and IPTV forecast scenarios for the coming five year were run. The results showed that the expected data growth will first start to overload the rings connecting MC locations to backbone locations. Towards the end of the five year period the links between MB and MC locations will also be overloaded if they are not upgraded. Because the Ethernet topology is a logical topology laid over the physical optical network, bypasses can be used to divert traffic from the Ethernet layer to the optical layer. A linear integer program was used to determine the optimal topology to transport the traffic demands that were obtained in the simulation. The analysis of the network structure and the ability to simulate traffic scenarios for individual services form a great asset in the capacity management of KPN's Ethernet network. The extra intuitive feel for the network that can be gained by displaying the network in the many different viewpoints offered by the tool's filtering capability is another attraction of the tool developed during the project.

Appendix A

Commands

In this appendix all the commands that the tool recognises are listed.

add	generator	open dialog to add generators
	generator <*vpn*> <*element*>	add generator for element <*element*> in vpn <*vpn*>
	monitored element <*element*>	add monitor for element <*element*>
count	link all	count all the links in the network
	link shown	count the number of shown links
	node all	count all the nodes in the network
	node shown	count the number of shown nodes
	overloaded	count the number of overloaded links during simulation
filter	[inbound,outbound,none] [av,max] [gt,st] <*value*>	filter links in direction [inbound, outbound, both] for an [average, maximum] value <*value*>
find	node <*node*>	highlight node <*node*>
gen	log off	stop logging simulated traffic
	log on	start logging simulated traffic
	start	start traffic generator
	stop	stop traffic generator
get	adj matrix	get the adjacency matrix
	adj matrix i	get the adjacency matrix with index array
	adj matrix w	get weighted adjacency matrix with traffic as weight

get	assort	get the assortativity coefficient
	av degree	get the average degree of all shown nodes
	av eccent	get the average eccentricity
	av pl	get the average path shortest path length between any two nodes
	booked traffic from <*node*>	show booked capacity leaving node <*node*>
	booked traffic to <*node*>	show booked traffic entering node <*node*>
	cc	get the clustering coefficient
	closeness	get the closeness of all shown nodes
	closeness i	get the closeness of all shown nodes with ID
	connected	check whether graph is connected or not
	degree	get the degrees of the nodes
	degree i	get the degrees of the nodes and node ID
	eccentricity <*node*>	get the eccentricity of node <*node*>
	laplacian	get the Laplacian
	link density	get the link density
	links node <*node*>	get the links adjacent to node <*node*>
	max pl	get longest shortest path in the network
	most central	get the most central node in the network
	most eccentric	get the most eccentric node in the network
	network eccentricity	get the eccentricity of every shown node
	node id <*node name*>	get node ID of node with name <*node name*>
	node layer <*node*>	get the layer of node <*node*>
	node name <*node*>	get name of node <*node*>
	node type <*node*>	get the type of node <*node*>
	sp	get a matrix with all shortest paths (distance matrix)
	sp <*node 1*> <*node 2*>	get a shortest path between <*node 1*> and <*node 2*>
	subgraph count	get the number of disconnected subgraphs in the network
	vpn <*node*>	get all shown VPNs that pass through node <*node*>
	vpn <*node*> all	show all VPNs that pass node <*node*>
hide	node layer <*layer*>	Hide all nodes of layer <*layer*>
	group <*group*>	

list	contents location	List all nodes in a location
	locations in view	List all locations in view
	links	List all links
	links in view	List all links in view
	links node <*node*>	List all links adjacent to node <*node*>
	links shown	List shown links
	network stats	List the number of nodes per layer and number of links per type
	nodes	List all nodes
	node in view	List all nodes in view
	node layer count	
	nodes shown	List shown nodes
	unconnected nodes	List all nodes without links
	vpn path <*vpn*>	List the hops in vpn path <*vpn*>
load	last	load the last opened network
	network	show file chooser to open network
	vpn <*vpn*>	load vpn <*vpn*>
	vpn cust <*id*>	load all vpns for customer <*id*>
	vpn cust <*id*> m	load all vpns for customer <*id*> and merge
macro	play	play macro
	start	start recording macro
	stop	stop recording macro
merge	locations	Merge all nodes in a location to a single node
	locations inv	Splits all nodes in a location into separate locations
remove	node <*node*>	remove node <*node*> from network
save	view <*filename*>	save the view as postscript file under <*filename*>
	view full <*filename*>	save the complete network as postscript file under <*filename*>
set	line width <*width*>	set line width to <*width*>
	marker size <*size*>	set marker size to <*size*>
show	link adjacent	show all links adjacent to shown nodes
	link all	show all links
	link connecting	show all links connecting shown nodes
	link lgap <*lgap*>	show links that connect nodes with layer distance <*lgap*>

show	link v lgap <*lgap*>	show links that connect nodes with layer distance <*lgap*> that are in view
	link l <*layer*> lgap <*lgap*>	show links that connect nodes at layer <*layer*> with layer distance <*lgap*>
	link none	hide all links
	node <*node*>	show node <*node*>
	node adjacent	show all nodes adjacent to a shown link
	node all	show all nodes
	node layer <*layer*>	show all nodes of layer <*layer*>
	node none	hide all nodes
	node v layer <*layer*>	show all nodes of layer <*layer*> that are in view
	node vpn adjacent	show all nodes adjacent to a VPN
	scenario	show the dialog to set the simulation scenario
	vpn all	show VPNs
	vpn none	hide VPNs

Bibliography

[1] Xiao F. Wang and Guanrong Chen. Complex networks: small-world, scale-free and beyond. *Circuits and Systems Magazine, IEEE*, 3(1):6–20, 2003.

[2] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.

[3] M. E. J. Newman. Assortative mixing in networks. *Phys. Rev. Lett.*, 89(20):208701, Oct 2002.

[4] P. Van Mieghem, H. Wang, X. Ge, S. Tang, and F. A. Kuipers. Influence of assortativity and degree-preserving rewiring on the spectra of networks. *The European Physical Journal B - Condensed Matter and Complex Systems*, pages 1–10–10, July 2010.

[5] Nair Maria Maia de Abreu. Old and new results on algebraic connectivity of graphs. *Linear Algebra and its Applications*, 423(1):53 – 73, 2007.

[6] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 6 1998.

[7] A. L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[8] A. L. Barabási. The architecture of complexity - from network structure to human dynamics. *IEEE Control Systems Magazine*, 27(4):33–42, 2007.

[9] Aaron Clauset, Cosma Rohilla Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009.

[10] T. Engel, A. Autenrieth, and J.-C. Bischoff. Packet layer topologies of cost optimized transport networks. *Optical Network Design and Modeling, 2009. ONDM 2009. International Conference on*, pages 1 –7, feb. 2009.

[11] M. Schlosser and E. Patzak. Method for placing bypass capable nodes in two-layer networks. *Transparent Optical Networks, 2009. ICTON '09. 11th International Conference on*, pages 1 –4, jun. 2009.

[12] A. Kern, I. Moldovan, P. Hegyi, and T. Cinkler. Ethernet over wdm: Optimization for resilience and scalability. *Design and Reliable Communication Networks, 2007. DRCN 2007. 6th International Workshop on*, pages 1 –8, oct. 2007.

[13] N. Kamiyama and D. Satoh. Network topology design using analytic hierarchy process. *Communications, 2008. ICC '08. IEEE International Conference on*, pages 2048 –2054, may. 2008.

[14] T. Wauters, J. Coppens, B. Dhoedt, and P. Demeester. Load balancing through efficient distributed content placement. *Next Generation Internet Networks, 2005*, pages 99 – 105, apr. 2005.

[15] Jing Sun, Suixiang Gao, Wenguo Yang, and Fei Huang. Replica placement algorithms in content distribution networks. *Information Engineering and Electronic Commerce (IEEC), 2010 2nd International Symposium on*, pages 1 –4, jul. 2010.

[16] Lili Qiu, V.N. Padmanabhan, and G.M. Voelker. On the placement of web server replicas. *INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, 3:1587 –1596 vol.3, 2001.

[17] S. Zaman and D. Grosu. A distributed algorithm for web content replication. *Network Computing and Applications, 2009. NCA 2009. Eighth IEEE International Symposium on*, pages 284 –287, jul. 2009.

[18] Roberto Solis-Oba. Approximation algorithms for the k-median problem. In Evripidis Bampis, Klaus Jansen, and Claire Kenyon, editors, *Efficient Approximation and Online Algorithms*, volume 3484 of *Lecture Notes in Computer Science*, pages 292–320. Springer Berlin / Heidelberg, 2006.

[19] Mohammad Mahdian, Yinyu Ye, and Jiawei Zhang. Approximation algorithms for metric facility location problems. *SIAM J. Comput.*, 36(2):411–432, 2006.