universität wien

DISSERTATION

Titel der Dissertation

"Operations Research in action:
a project for designing
telecommunication access networks"

Verfasser

Mag. rer. nat. Bertram Wassermann

angestrebter akademischer Grad

Doktor
der Sozial- und Wirtschaftswissenschaften
(Dr. rer. soc. oec.)

Wien, im Oktober 2011

Studienkennzahl lt. Studienblatt:	A 084 136
Studienrichtung lt. Studienblatt:	Statistik
Betreuer:	Univ.-Prof. Dr. Immanuel Bomze

Everything should be made as
simple as possible, but not simpler.

This statement is ascribed to Albert Einstein. Ian Morris quoted this statement in his extremely interesting book *Why the West Rules - For now.* He noted that the source of the quote is unknown. But, he added: *Probably it is the most important thing, Einstein never said.*

Acknowledgments

My greatest thanks belong to my advisor Prof. Immanuel Bomze. It was he who motivated me to take on the challenge to write a thesis while holding down a full time job in the private sector. I was sure of his support, even if the time to write the thesis took longer as normal due to the project I had to attend to. After being away from the world of academia for some years already, I needed lots of advice which he willingly gave. Also, coming from the algebraic branch of Mathematics I needed some extra introduction to topics of optimization. Moreover, he introduced me to people who were not only helpful to me but also for the project.

There was Dr. Ivana Ljubic to whom I owe many thanks. She had finished her PhD thesis on two network design problems which was awarded with a prize from the ÖGOR. I had the pleasure at the prize giving. Ivana, Immanuel and I initiated a cooperation between Telekom Austria and the University of Vienna which was funded by the FFG and Telekom Austria. During many project meetings Ivana had lots of opportunities to support me with her working knowledge of discrete optimization and her organizational skills. It was she who coordinated and organized most of the tasks concerned with the cooperation and the project group.

I should consider myself lucky to work for an employer who supports and in fact encourages further education. I was allowed to complete thirty hours of course work during regular office hours. Four months of educational leave helped me to speed up writing my thesis. Since it is a little bit abstract to pay tribute to a company I would like to thank my bosses: The initial commitment came from Dipl.-Ing. Michaela Barta-Müller and Dipl.-Ing. Thomas Riedl, MBA. Then I changed the working group and as a consequence the topic of the thesis, from statistical analysis based on support vector machines to network design. I have to thank Immanuel again for allowing and making this change possible. My new bosses who continued to support me were Dipl. Ing. Dr. Georg Kern, MBA and Andreas Engel. Finally, the merger of the two companies Telekom Austria and A1 brought many changes including new superiors: Mag. Astrid Wallisch and Mag. Gunther Oswalder. Thanks to all of them.

There are several colleagues to whom I would like to express my gratitude. Although, they were just doing their work, I would not have been successful without them. Ing. Roman Redler — a colleague from the Operations Research group — introduced me to the technical world of telecommunications. On the one hand he taught me all I had to know about terms and concepts of network design and operation. On the other hand he introduced me to the people who were living in this world. Ing. Franz-Josef Vögel provided the data for the first phase in an extraordinarily precise, professional and yet fast way. He never answered a request with friendly small talk and the perspective that a solution to the request might follow sometime the following week. He answered the same day by sending the solution. There were many planners and other people involved in discussing and defining the problem at hand. As representatives for all of them I would like to mention Ing. Günther Bamer and the two planners Ing. Andreas Kriesel

and Ing. Roman Schmidt. We spent a lot of time together and had one or two heated discussion and disagreements and we found solutions. Last but not least, there was Gerhard Prasch — the project leader of the first phase of SARU. We formed a well tuned team. We collected all necessary specifications for optimization and visualization software which was to be produced. We also spent some time on trains which took us to training sessions for regional planning groups. Cheers!

Mag. Annemarie Schiffbänker — a colleague and good friend — provided support for the analysis and description of the national and international market situation in Section 1.1. She also did the proofreading of this section and gave valuable feedback.

My two dearest friends from the furthest corners of the world helped me to perfect my written English. Helen Innerhofer, Bsc, an Australian with 100% Chinese roots teaching English and Science in Vienna, and Kevin Marks, Cert.ed, an English Gentleman having lived all over the world from France, Australia to Vienna where he teaches and examines English seemed to me to be the best people to check my German English. If the more mathematical parts of the thesis with their succession of definitions and propositions, theorems and proofs show some stylistic or grammatical problems, then it is due to the fact that I wanted to spare them. Thank you.

Mag. Claudia Lasser, a good old friend and a colleague in matters of PhD studies, helped me to write my first line and to fill the first page by explaining to me that I could always throw it away later, if I did not like it. After overcoming writer's block she stayed my coach until I had finished my work. Thank you, too.

Of the total amount of time and effort which people were generously willing to support me with I received by far the largest contribution from my deep love and wife Veronika. By now she knows more about discrete optimization and network design than I every will about accounting. And she was there, when I did not want to speak. Therefore, I planned to dedicate this work to her. However, she preferred an invitation to a very good dinner and the promise that I would start a new project soon, so that she still may have some spare time for her beloved garden.

It is quite likely that everybody I could dedicate this piece of work to would rather prefer a very good dinner. Therefore, I dedicate my thesis to my advisor Prof. Immanuel Bomze, since without him this piece of work would not exist.

Thanks to all of you who accompanied me through this fascinating, exhausting, challenging and occasionally disappointing project. What's next?

Contents

Chapter 0

Introduction

According to an article in the daily newspaper Der Standard [15] the time of internet started in Austria on 10^{th} of August 1990. Since then the community of people accessing the internet has been constantly growing, and so has the demand for speed of data transmission and size of transmitted data. More and more people are sending and downloading bigger and bigger files. Access has to be immediate. Delays during data transmission have to be reduced and for some applications they even have to be avoided.

The data load on the physical telecommunication networks has increased enormously. Backbone networks — the internet highways between local access points, so to speak — are meanwhile all based on optical fiber technology. Light is used to encode the information which has to be transmitted through optical fiber cables which function like small tubes of mirrors. This technology is a lot faster and more efficient than the old technologies which utilize copper wires and electrical impulses.

Local access networks — the city roads and driveways — which connect the internet user to the local access points and from there via the backbone network to the entire world wide web still live in the "old", electrical world. It is a question of time, that transmission demands will supersede the abilities of these networks. New local access technologies are needed and with mobile telephony an appropriate candidate arises.

The arrival of fast mobile internet access on the telecommunication market increased the pressure on fixed line access providers to plan and install new access networks. The fixed line telecommunication industry started to discuss a whole bundle of strategies to lay out new access networks. These strategies became known under the abbreviation FTTx which stands for Fiber To The x. In the center of all these considerations is the substitution of copper wires by optical fiber. They differentiate themselves by the degree of how much they still utilize existing copper infrastructure.

FTTH, for example, which stands for fiber to the home, abandons copper completely by connecting the homes of customers directly to the internet by optical fiber cables. In FTTC, where

the C is short for curb, the last few hundred meters of the access lines — so to say the driveway — are still copper cables. The optical fiber network ends somewhere at the curb in front of several houses and is connected to the remaining copper net by so called access remote units.

In any case, these strategies are very generally formulated and not useful to actually plan such a network. Moreover, since building access networks is quite expensive because of construction work and necessary equipment, questions of how to cost-optimally plan new networks arise.

For this purpose, Telekom Austria — the incumbent on the fixed line telecommunication market in Austria — started a project to develop a software which supports a cost optimal planning process for FTTC access networks in 2006. The project was called Situating Access Remote Units, in short SARU. The Operations Research group of Telekom Austria was assigned with the realization of the task. This thesis is about project SARU.

The thesis documents the main steps of the project, describes the selected approaches and some of the major flaws, repeats the discussion of the central issues and embeds them in appropriate mathematical optimization theory. Finally, the solutions which were chosen and applied are presented. Although, the application of mathematical theory to a real world problem was the ultimate goal of this project, the success of an Operations Research undertaking needs more then good mathematics. A good understanding of the environment of the problem, communication and project management are also needed. Therefore, this work goes beyond a mere introductory description of the real world problem and elaborates more extensively the background of the problem and the steps of the project themselves.

The first two chapters organize their topic in three steps. First, a real world problem is presented and discussed. The corresponding events which were important for the course of the project are documented. A basic solution strategy is outlined. Second, a literature review of appropriate mathematical theory is listed. The theory is used to translate the previously described real world problem formally into a mathematical problem. The formulation is checked whether it is in compliance with all requirements of the real world problem. Finally, the solution as it was developed during the project and applied to the problem is stated and discussed in detail.

In Chapter 1 the FTTC (fiber to the curb) planning strategy is translated into a facility location problem which is already thoroughly studied especially for telecommunication problems. There even exists a book on this topic from H. Yaman: Concentrator location in telecommunications networks [60]. The chapter starts with an analysis of the telecommunication market of the last 10 to 15 years with a special focus on the period from 2006 to 2009. The market is portrayed from an economical as well as a technological perspective to give a well funded background for the problem which has to be solved. Then the initial solution strategy is described, the most important data problems and project requirements as they were collected during the specification process are documented and the reasons are stated why it was necessary to depart from the initial strategy.

The theory part gives a collection of the most important variants of the facility location problem and screens their compliance with the previously stated specifications. Some adjustments are necessary and are carried out. This part closes with a formal definition of a structural planning process which is basically an iterative combination of two of the variants of the facility location problem.

The last section comprises a collection definitions which originate from the analysis of the database of the problem. It concludes with the major result of this chapter which is a dynamic programming algorithm that solves the facility location problem for the FTTC planning strategy. Under certain conditions the algorithm achieves this goal to optimality.

The second chapter is based on two problems which came to light after first approaches to solve the facility location problem for the FTTC planning strategy were presented to a group of experienced planners. One problem resulted from a miserable visualization, i.e. communication, of the found solutions, the other from the fact that solutions contained facilities of low utilization: the number of customers for whom these facilities were supposed to be installed was too small. The first part of the second chapter reports the events and considerations which took place during this period of the project.

The first of the two problems — the visualization — was to be solved by an adequate visualization tool. This tool was made available by the Operations Research group. A detailed report is not part of this thesis. The second problem leads directly to questions of network quality. In this context network quality has to be understood as the speed of internet access which can be granted to individual customers by a given network. A discussion of what network quality may mean and how it should be measured is also contained in the first part of Chapter 2.

Basically, the underutilization problem — too few customers are assigned to a facility — can only be treated by reducing the number of facilities which in consequence lowers the overall quality of the network by increasing the distance between customers and supplying facilities which leads to a reduction of transmission rates. From this a new question arises: how much network quality should be abandoned for how much saving? The answer is a priori not clear. To answer this question decision support is suggested as a form of investigating the parameter space of a mathematical optimization problem and making the results accessible to the practitioner, rather than delivering a single optimal solution.

The k-median problem which is also a well studied optimization problem appears as a possible mathematical answer for the combination of solving the underutilization problem and providing decision support. The k-median problem is introduced in the theory section of Chapter 2 as a facility location problem with a fixed number of facilities, i.e. a fixed number of k access remote units has to be situated. By solving the k-median problem for several values of k an appropriate number of facilities which provide an acceptable network quality can be found. Additionally, the algorithm to solve the k-median problem is recursive: to solve the k-median problem the $(k-1)$-median problem has to be solved first. In other words, the k-median algorithm includes

a search of the parameter space and thereby provides the kind of decision support as it was suggested above. The theory part of Chapter 2 provides an overview of different variants of the k-median problem and states an iterative algorithm to solve the k-median problem.

Unfortunately, this algorithm starts to explore the parameter space from the wrong side. It starts with one facility and increases the number of facilities successively until the desired value is reached. Although the strategy to handle underutilization of facilities waives some of the network quality, a high degree of customer supplement is desirable, i.e. nonetheless, network quality should be high. Therefore, the necessary number of facilities will have to be high. So, it would be better to trail through the parameter space from above, starting with the highest possible number of facilities and reducing their number one by one. The third part of Chapter 2 realizes exactly this approach and states the corresponding algorithm.

In Chapter 3 an extensive empirical study of 106 different local access areas is presented. The main purpose of this demonstration is to give a concrete impression of how decision support can be provided. The methods which are presented and developed in this thesis are used to prepare the planning process by studying strategic questions like the CO circle enforcement or the balance between facility utilization and coverage. The study of the runtime behavior of the algorithms provides information which is useful to set up an appropriate working environment for the future users. Additionally, the two variants of the k-median algorithm — the ascending and the descending method — can be compared.

The thesis concludes with a discussion of topics for further work and cautions against prize-collecting approaches in planning access networks for telecommunications in Chapter 4.

Chapter 1

FTTC and Facility Location

Between January 2006 — the start of project SARU — and December 2009 — the end of the first phase of the project — many things had changed which had their impact on the project to a greater or lesser extend. In 2006 Boris Nemsic - then COO of Mobilkom Austria - was still yet to become Chief Executive Officer of the entire Telekom Austria Group. He actually followed Heinz Sundt in May 2006 who was the last CEO with a strong background in fixed line telecommunication. In March 2009 Boris Nemsic left the company and took over the Chief Executive Office of VimpelCom, a Russian telecommunication provider more than 20 times bigger than Telekom Austria[1]. He was succeeded by Hannes Ametsreiter who was CCO at Mobilkom Austria at that time. This movement of top management from Mobilkom Austria to Telekom Austria was finalized in 2010 by the merger of the two companies to A1 Telekom Austria.

Of course, this change of management was not restricted to the top level. Long term Head of Operations at Telekom Austria Fixed Net Rudolf Fischer left the company in summer 2008 and with him top management personnel who actually initiated this project. This led to critical situations for the project, since its goals and content had to be explained, promoted and justified to new management every time.

Some of the people directly involved in defining the project started their own new projects like the first two project leaders who went into retirement. Due to the reorganizations which followed the new management the Operations Research group of Telekom Austria which was in charge of the project lost some of their members. One colleague was lain off, but became at least the chance to write his thesis on a traveling salesman application for Telekom Austria. Others were transferred to a different department.

But, changes were certainly not restricted to people and personnel. Telecommunication markets changed, too. And the impact of these changes on the course of project SARU was probably bigger.

[1] He left VimpelCom in June 2010 again.

1.1 Telecommunication: Austrian market and technologies

To realize the impact of the change on the Austrian telecommunication market between 2006 and 2009 one has to go back in history a bit further.

1.1.1 The telecommunication market in Austria

The rise of mobile communication since the 1990s started to threaten the business of fixed line telecommunication companies and there especially incumbents like Telekom Austria. In the beginning the threat was identified by the term traffic loss (see Figure 1.1, all numbers are taken from [19] and [20].). One way to measure the market share of a provider is the percentage of total voice traffic (in minutes) which originated at the provider's customers. In Austria for fixed line providers these market shares decreased from 51% down to 24% within 6 years.

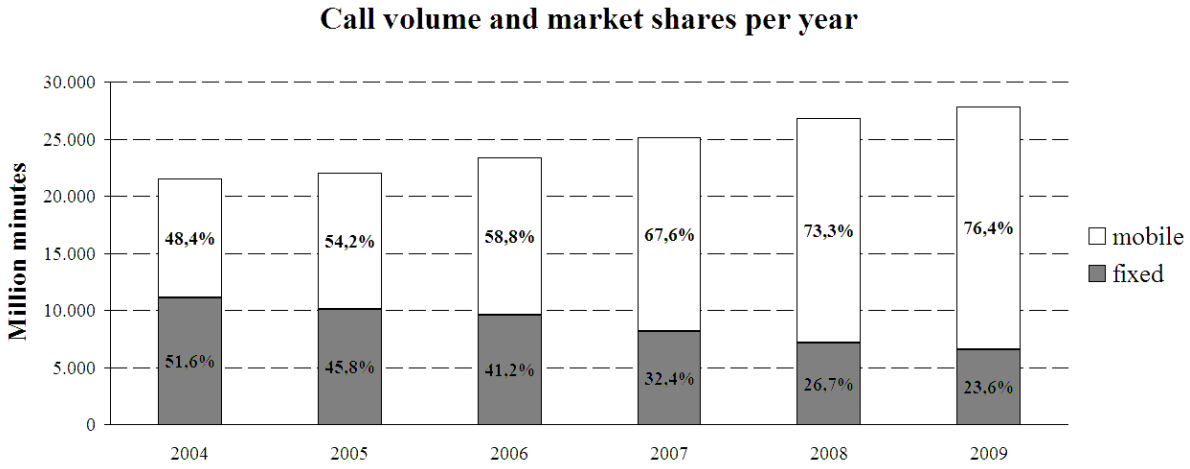


Figure 1.1 Traffic Loss
Source: RTR telekom monitor [19] and [20]

This effect could be increased by mobile telephony companies, because they managed to enlarge the absolute size of the market (from $21,5$ million minutes to $27,8$ million minutes between 2004 and 2009), i.e. they motivated people to make more phone calls.

A similar picture — although with one very important difference — can be drawn from the inspection of the revenues on the telecommunication market in Austria (Figure 1.2). The share of fixed line networks of the revenues declined from 40.6% in 2004 to 35.1% in 2009. The important difference is that the total revenue on the market declined, too, from nearly 6 billion Euros to just over 5 billion Euros per year. This is due to a very competitive mobile telecommunication market in Austria.

A comparison with other European countries gives an indication of how competitive this market is (see Figure 1.3). According to a comparative study of the consulting company Analysys Mason from 2010 prices for mobile telephony are on average around 48% lower compared to fixed line prices in Austria. There are only two countries out of 27 — Poland and Romania —

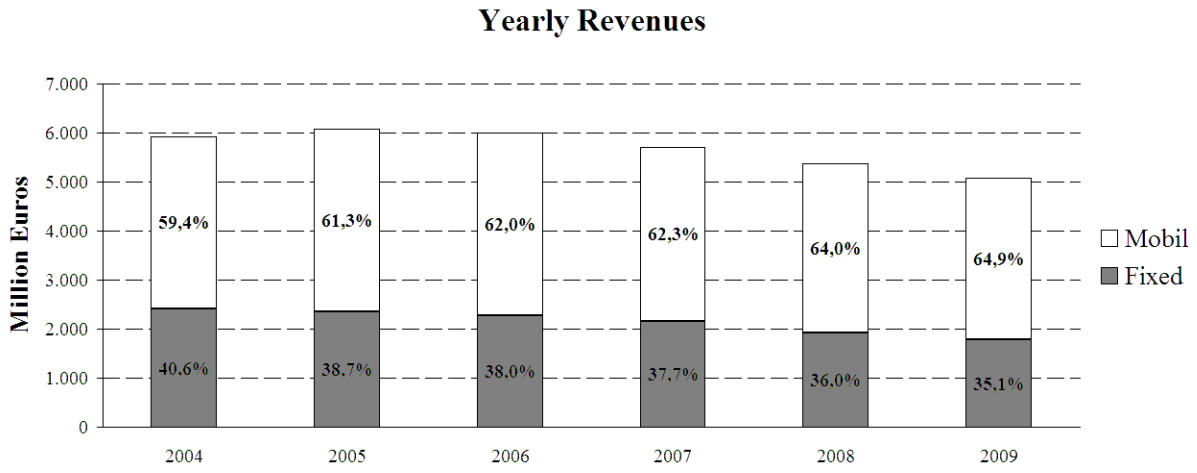


Figure 1.2 Revenue Loss
Source: RTR telekom monitor [19] and [20]

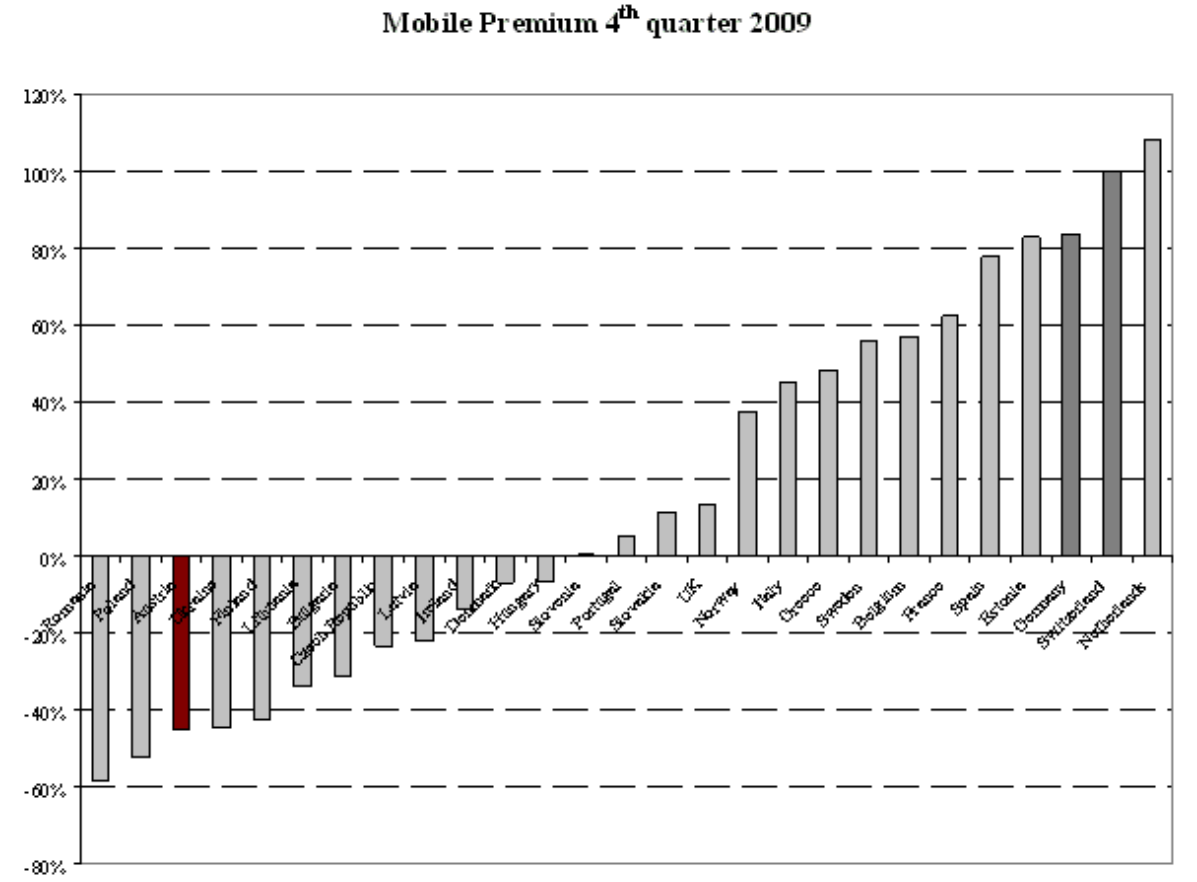


Figure 1.3 International mobile telephony prices compared to fixed line prices
Source: Analysys Mason

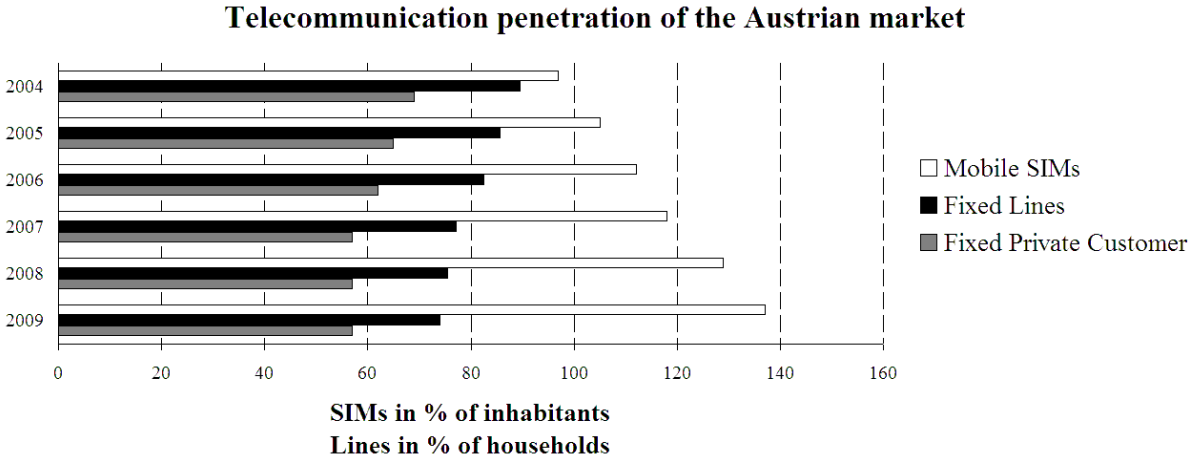


Figure 1.4 Line Loss
Source: RTR telekom monitor [19] and [20]

where the proportion is even more in favor of mobile telephony. On the other side the study shows that there are 15 countries where fixed line telephony is cheaper: in Germany by 80%.

On the Austrian market, this is mainly due to a radical change in the design of mobile tariffs. In the past, fixed and variable costs were charged to the mobile customers, quite like fixed line contractors did. Additionally to a base fee the minutes a customer had telephoned were put on the bill. Later the variable portion of the mobile tariffs became cheaper and cheaper, until finally all-inclusive contracts were offered. Phone calls not only in ones own network but also in foreign networks became free of variable charge. Customers were not billed per minute anymore, but purchased packages which granted them 200 to 1000 minutes of phone calls without any additional costs.

As a consequence a new threat arose for fixed line providers called line loss (see Figure 1.4 numbers are taken from [19] and [20]). With declining prices for mobile telephony compared to fixed line rates customers lost their motivation to pay base fees for two telephony products - mobile and fixed line. Customers started to cancel their fixed line, a phenomenon observed since the beginning of this century.

In 2004 there were nearly as many active SIM cards as people in Austria. Five years later the proportion increased to nearly 140%. Still, not every Austrian owned a cellular phone at the end of the first decade [2]. But, some had two or more in use. Also, the numbers comprise all mobile connections including those used by businesses and machine to machine applications.

At the same time fixed line penetration which is measured in percentage of households dropped. There are two different numbers displayed in Figure 1.4 depending on whether non-private customers are included or not. Number of households is used as the base in both cases. In 2004 70% of private households were equipped with a fixed line access. This number went down by 13 percentage points to 57% by 2009. In absolute numbers this development is depicted in Figure 1.5. It can be observed that the line loss seems to slow down during the last three

[2] E.g. the parents of the author.

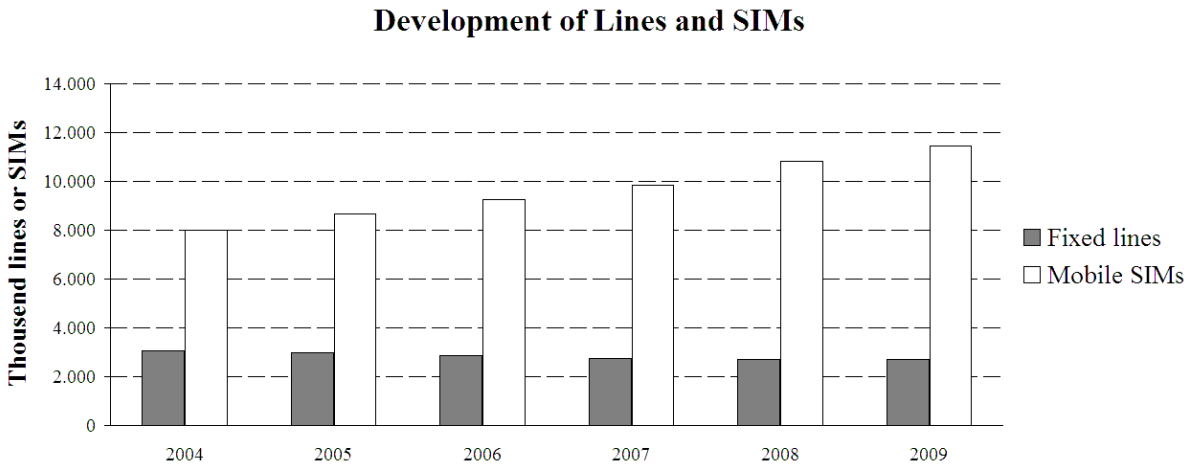


Figure 1.5 Lines and SIMs in absolute numbers
Source: RTR telekom monitor [19] and [20]

years. Certainly, fixed line providers tried to stop line loss and even hoped to accomplish the turn around.

Fixed line services include more than voice telephony. Beside all kinds of special business solutions, TV or security services the major additional service was and is internet access. In this context losing lines means losing the connection to customer's homes and consequently creating an obstacle to sell new internet based services and products later on. Therefore, the main argument of fixed line providers for their customers to keep and pay their lines became their internet access. Instead of selling voice, internet access and TV as separate components individually, fixed line companies started to offer bundles of products at attractive prices.

Telekom Austria started combined products in spring 2008 with aonKombi, an offer comprising of internet access (with transmission rates reaching from 1 Mbps up to a maximum of 8 Mbps[3] and unlimited downloading volume), fixed line voice and also mobile voice telephony (3 SIM cards) for a base fee of € 29.90[4][43]. aonSuperKombi which additionally included digital TV was available for € 34.90. UPC, the main fixed line competitor of Telekom Austria, started the product line FIT (Fernsehen[5], Internet, Telephone) in February 2009 for € 39.90 containing unlimited 2 Mbps internet access, digital TV and fixed line phone[43].

The increasing importance of internet access for fixed line providers can be seen by the growing absolute and relative size of the revenue due to internet access in Figure 1.6. Whereas the relative portion of voice revenues decreased from 73% to 59% between 2004 and 2009, the same indicator increased from 16% to over 27% for internet access. Although, the absolute total revenues decreased on the fixed line market by nearly one billion Euros, the broadband internet access revenues increased by over 100 million Euros from 374 million to 489 million (All numbers cited from [19] and [20]). This happened despite the fact that charges for internet access products were lowered to make them more attractive for customers.

[3]Mbps= mega bit per second. The speed of data transmission is measured in units of data per second.
[4]Installation, activation and variable fees were not included.
[5]German for television

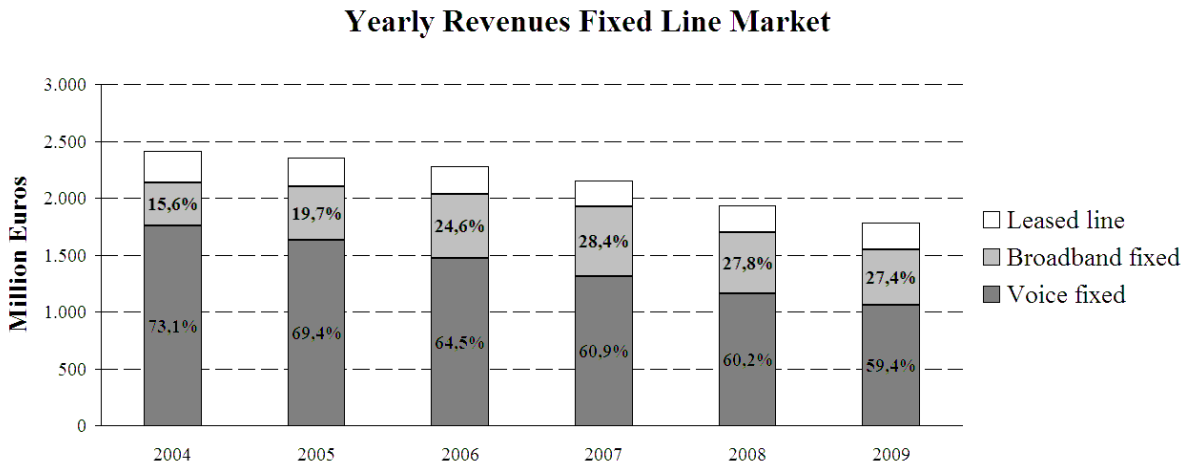


Figure 1.6 Revenue split for fixed line market
Source: RTR telekom monitor [19] and [20]

Line loss could be slowed down. It even seemed possible to re-enter a period of growth. Revenues were still declining. Although, the whole market — fixed and mobile — was affected, a business field for which revenues were growing had been identified. Things seemed to have taken a turn for the better. But, the next threat was on the rise: mobile broadband internet access.

1.1.2 Mobile internet access

Mobile telephony has been known already for a while. The first commercial systems appeared at the end of the fifties of the last century. The so called mobile telecommunication standards of the first generation were all based on analog data transmission for voice telephony. In the beginning they were mainly car based, because of the size of the equipment and the high demand for energy.

In 1982 the Confederation of European Posts and Telecommunications (CEPT) initiated a development process to design an European mobile telephony standard (see [59] for more details). For that purpose an organization called Groupe Speciale Mobile (in short GSM) was founded which defined the standard called Global System for Mobile communication with the same abbreviation. The GSM family of standards represent the second generation of mobile telephony technologies. Data transmission was already digitalized.

The first GSM phone call was made in Finland in 1991. In Austria A1 started with GSM in 1994. At about the same time the world wide web started to spread, too. GSM was mainly designed for voice communication. A short messaging system (SMS) was the only additional data service. Subsequent improvements of the standard allowed higher transmission rates of data. In the beginning GSM transmitted data at up to 14.4 kbps (kilo bit per second)[6]. With the first improvement of the standard — HSCSD, high speed circuit switched data — transmission rates could be increased up to 57.6 kbps.

[6]The speed of data transmission is measured in units of data per second.

Next, GPRS (General Packet Radio Service) improved transmission rates to a range from 56 to 114 kbps. The first commercial GPRS networks were started in 2000, when the first GPRS handsets became available. It enabled new services like multimedia messaging services (MMS) to send pictures and videos. In combination with a computer GPRS provided narrowband internet access as fixed-line modems like the V.90 standard already did at that time. Still, mobile internet access did not present a real competition for fixed line internet providers. New transmission techniques in fixed line communication had carried transmission rates beyond the Mega bit per second (Mbps) boundary bringing the world of data transmission the term broadband.

Both, narrowband and broadband are related to the speed of data transmission. Clearly, the former is slower than the latter. However, the terms actually refer to the fact that the latter uses a broader frequency spectrum to transmit data than the former which as a result makes it faster. Where to put a boundary between the two by means of a certain number of kbps, is hard to tell. It is safe to assume that narrowband speed is better measured in kbps and broadband in Mbps.

The end of the development of the second generation is marked by the introduction of EDGE (Enhanced Data Rates for GSM Evolution) in 2003. In Austria and Germany the EDGE system was introduced two years later. With EDGE data rates between 150 and 200 kbps are typical. EDGE is also considered a transient technology to the third generation of mobile telephony standards.

UMTS (Universal Mobile Telecommunications System) opened the succession of third generation technologies. Defined by the 3rd Generation Partnership Project (3GPP) and first released in 2000 it was commercially rolled out in Austria by A1 in 2002. In the beginning transmission with UMTS was not a lot faster than with EDGE and more costly, since UMTS made new equipment necessary. Despite this fact, the role of EDGE was to fill the space where UMTS had not been rolled out, yet.

During the following years new technologies and standards brought mobile transmission rates from below 1 Mbps up to a theoretical maximum of 56 Mbps and a practical value of 28 Mbps — from narrowband to broadband.

First, there was High Speed Downlink Packet Access (HSDPA) which improved downlink transmission[7] to 3.6 Mbps or even 7.2 Mbps. This was part of the so called release 5 of the 3GPP which was released in 2002. Release 6 brought about, amongst other things, an enhancement of uplink speed in the last quarter of 2004. High Speed Uplink Packet Access (HSUPA) allowed up to 5.8 Mbps uplink. Finally, with release 7 in 2007 Evolved HSUPA (HSPA+) was published which achieves the already quoted 28 Mbps and has got the potential to do even better than this.

The Austrian market saw the launches of these new technologies in the following succession. HSDPA was first launched by T-mobile in 2006 followed by A1 and One (now Orange) in 2007. Hence, there elapsed four years between release and launch. HSUPA was launched in 2007 three years after its release. And two years after the release HSPA+ followed in 2009.

[7] Transmission from the antenna to the mobile device.

On the first of March 2006 T-mobile Austria put the first HSDPA net of Austria into operation. The launch was accompanied by the following tariff: For monthly 45 Euros the customer was allowed to download 1.5GB of data. The data card was for free. At that time the operator promised a download speed of up to 1.8 Mbps.

In October 2006 a comparable tariff of Telekom Austria was AonSpeed 1000. It granted a download volume of 1 GB at a speed of 2 Mbps for 29.90 Euros. For 10 Euros more the customer was allowed to download up to 4 GB[8]. And for 49.90 Euros internet access was unlimited and as fast as 3 Mbps[9] [43].

One year later, in February 2007, One (now Orange) launched three different tariffs for HSDPA internet access called H.U.I.[10], a small, a medium and a large sized variant: 250MB download limit for 10 Euros, 1GB for 20 Euros and 20GB for 50 Euros.

Another year later, in March 2008, no small H.U.I. tariff existed anymore. The medium tariff allowed 3GB for 15 Euros which was still even underbid by Hudchinson (3) where 3GB could be downloaded for 12 Euros. Hudchinson also offered a large sized tariff with 15GB for 20 Euros, A1 with Bob a small version with 9.9 Euros for 0.5GB.

Within 2 years the price for mobile broadband access dropped from 30 Euros to 4 Euros per Gigabyte in the category of medium sized tariffs. Small sized tariffs saw a similar reduction from 40 Euros to 20 Euros. Prices for large sized tariffs dropped from 2.5 Euros per Gigabyte to 1.33 Euros.

At the same time in march 2008 7 out of 9 internet access tariffs of Telekom Austria were flat, i.e. there was no download limit anymore. The product differentiations were based on the transmission rate. That was why they were called "AonSpeed Flat 3MBit"or "AonSpeed Flat 16Mbit". Prices ranged from € 10 to € 50 [43]. The situation was very similar for other fixed line companies.

The events of these years are summarized in Figure 1.7. Before 2007 the fixed line access market grew faster than the mobile counter part. Both lines show a distinct change of behavior in 2007. Growth of volume of data cards increased, whereas fixed line growth was reduced, although still positive. Considering that there are around 3.5 million households in Austria[11] these figures show that by the end of 2009 a household penetration of nearly 90% was reached.

There are two possible explanations for this high penetration rate. First of all, the depicted numbers comprise of private and business customers. Numbers for private customers only were not available. Secondly, there was a tendency at that time to use mobile internet access as an add-on to fixed line access, which was still more stable, faster and allowed more download volume for customers in the high-usage segment.

[8] AonSpeed 4000

[9] AonSpeed flat

[10] H.U.I. stands for "höllenschnelles, ultra-einfaches Internet"which can be translated like "fast as hell, ultra-simple internet". In German the word hui also imitates the sound of a strong blowing wind.

[11] Statistik Austria as of October 2009

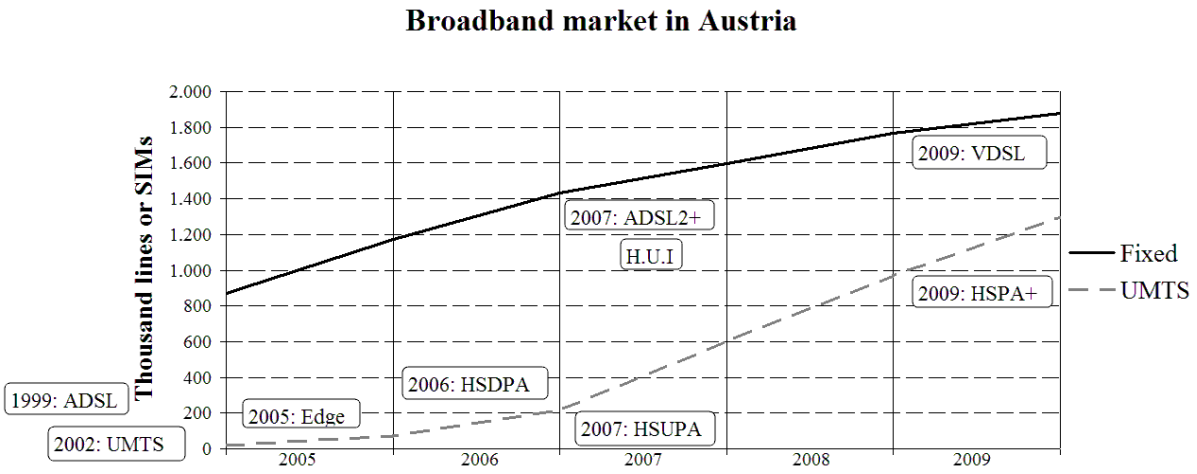


Figure 1.7 Austrian broadband market between 2005 and 2009
Source: RTR telekom monitor [19] and [20]

However, the development does not stop there. The fourth generation of mobile standards is on its way. The key term is Long Term Evolution (LTE) which promises up to 150 Mbps downlink speed. The first test stations went in operation at the beginning of 2010. The next threat for fixed line providers is maybe called "access loss".

1.1.3 Fixed line counterstrategies

There are three principal paths which fixed line providers can take to cope with the situation on the internet access market.

1) Retreat
In the business field of voice telephony mobile companies have shown their dominance already. It is just going to be a question of time before fixed line voice telephony will become extinct or at least be on the fringes.

Furthermore, by differentiating between mobile and fixed technologies it should be observed that only aspects of access to networks are addressed. The internet is a network of computers connected mainly by wires, i.e. fixed lines. The core networks are rarely air borne. Access networks are just the first gate of a user to the internet. In between lies a huge world of wires.

At least the owners of networks would still have enough work to do and money to earn, even if they retreated from retail business and concentrated on wholesale, network support and maintenance.

2) Competition
Mobile technologies have seen enormous advances during the last few years and it is hard to tell, if LTE — the fourth generation — will be the last generation. At least, the notion long term evolution promises the opposite. But, despite this progress it must be observed that fixed line technology is always ahead of mobile access. At any time fixed line could offer higher access speed and will be able to do so during the coming years.

While mobile access was narrowband during its second generation (GSM), fixed line access became broadband with ADSL[12]. Mobile internet providers were closing the gap with third generation high speed packed access. Fixed line providers went off again by implementing VDSL[13] which allows a download speed up to 35 Mbps. Mobile networkers are inventing a long term evolution which promises up to 100 Mbps. For fixed access networks optical fiber technology[14] is already available allowing a maximal download speed of 2.4 Gbps (Giga bit per second).

Nevertheless, it does not look like the fixed line providers could use the technological lead to their advantage (see Figure 1.7). There is a strong argument for mobile access: mobility which actually comprises of two different aspects, the independence of wires (wireless) and being able to move — e.g. in trains or cars — and transmit data simultaneously (nomadic). The convenience of wireless access is obvious: no wires, no additional equipment, connecting to the internet from different rooms with no additional setup. There is no argument for the use of wires, except perhaps that fixed line access is faster.

Although, transmission rates are objectively measurable, speed is a relative quantity in the access market: high speed is what is needed to satisfy the service a customer demands. So, if a customer only wants to check and write e-mails and do research in the internet, a narrowband access will perhaps suffice and a slow broadband access will be luxurious. For services like video streaming, file exchange, or Internet TV (IPTV) high speed will mean something completely different.

But, as soon as the highest available access rates exceed the highest necessary transmission rates of all services available, increasing the speed of access networks even further becomes a more and more expensive investment. Since nobody needs that speed, nobody is willing to pay a price corresponding to the size of the investment. At the time of writing Telekom Austria offered two different IPTV products, a high definition TV settopbox which demands 8 Mbps access power, and a normal TV settopbox with a need for 4 Mbps. Considering now for example an above average household with a desire for one high definition TV and two normal TVs and two internet access lines with a maximal access rate of 6 Mbps each, such a demand can be satisfied in best case with one VDSL line which grants a maximum of 35 Mbps. But, one has to imagine 5 people who watch 3 different channels on TV and perform two high speed services on internet and all of that simultaneously.

The key question whether further investments and improvements of fixed line access networks will pay off is, what are future internet services like? What are their requirements for an access network of the future? Cisco, one of the worlds leading producers of network systems, regularly analyses and forecasts the field of internet services and their future requirements and publishes the findings in a report called "Cisco Visual Network Index: Forecast and Methodology"([12]).

In the report 2009-2014 seven different subsegments of internet services are analyzed and forecasted. The results are shown in Figure 1.8 which shows a detail of Table 10 from page 10 of the report. The numbers displayed are Petabyte (PB) which follows Terrabyte and is equivalent

[12]See the following sections for detailed explanations.
[13]See the following sections for detailed explanations.
[14]See the following sections for detailed explanations.

Consumer Internet Traffic, 2009–2014							
	2009	2010	2011	2012	2013	2014	CAGR 2009–2014
By Sub-Segment (PB per Month)							
File Sharing	4,091	5,075	6,197	7,492	9,125	11,340	23%
Internet Video	2,776	4,725	7,718	11,026	14,838	19,468	48%
Internet Video to TV	107	263	711	1,502	2,686	4,075	107%
Web/Data	1,688	2,273	3,006	3,930	4,933	6,134	29%
Video Calling	83	128	199	284	407	599	48%
Online Gaming	63	86	120	167	226	307	37%
VoIP	122	134	141	144	145	146	4%

Figure 1.8 Future broadband demand
Source: Cisco Visual Networking Index 2009-2014

to 10^{15} Byte. In 2009 the largest segment is file sharing, whereas in 2014 Cisco assumes that it will be internet video which should have grown by then in total by 177% over five years leading to an average growth per year of 48% (CAGR = compound annual growth rate).

However, the segment with the largest CAGR is forecasted to be "Internet Video to TV" with 107% leading to a total growth over 5 years of 3,708%. In total Cisco predicts an increase of customer IP traffic per month from 11,602 PB in 2009 to 55,801 PB in 2014 which corresponds to a CAGR of 37% or a total growth rate of 381%.

As demanding as these numbers look the question remains how much impact the projected development will have on access networks. The numbers do not explicitly show that internet services will ask for higher access rates for individual customers. They may result from an increasing number of customers using high speed services like real time video application which are — according to Figure 1.8 — on the rise. Core networks will certainly be effected by this development. But, will this also be true for access networks? And why shouldn't mobile access providers be able to cope with these demands, too?

3) Cooperation
The Cisco report "Cisco Visual Network Index: Forecast and Methodology, 2009-2014"([12]) also predicts that mobile data will increase from 66 PB per month in 2009 to 2,856 PB per month in 2014. This corresponds to a compound annual growth rate of 112% and a total growth rate over 5 years of 4,227%. The volume of 42 times the current mobile data traffic has to be transported additionally per month in 5 years time. In Western Europe — Cisco predicts — it is not going to be so bad. There it will be around 36 times the current volume which has to be transmitted additionally.

In this case, however, mobile access networks are directly affected. It was already pointed out that mobile networks are mere access networks. Mobile operators use wires from the antenna upward to the internet, i.e. fixed line networks. The gain of 42 or 36 times the current data volume will be met at the antennas and will have to be dealt there making new technologies or denser nets of antennas necessary resulting in the need for significant investments.

Consumer IP Traffic, 2009–2014							
	2009	2010	2011	2012	2013	2014	CAGR 2009–2014
By Type (PB per Month)							
Internet	8,930	12,684	18,092	24,546	32,361	42,070	36%
Managed IP	2,606	3,680	5,248	7,095	9,059	10,875	33%
Mobile Data	66	170	410	904	1,697	2,856	112%

Figure 1.9 Future mobile broadband demand
Source: Cisco Visual Networking Index 2009-2014, Table 9 on p. 9

Furthermore, consumer behavior studies show that a large number of the mobile internet access customers use their data card mainly from home. Some even use it with their desktop computer, in this way doing without the major advantage of mobile access. The Austrian Internet Monitor[15] documents that in 2009 and in 2010 50% of the people who used data cards — which were actually 29% of the respondents — used mobile access "mainly at home"[16]. According to the AIM-C the portion of people who used it "mainly outside"[17] increased from 16% to 21% between 2009 and 2010 [28].

Ericsson, the Swedish mobile phone company, found in their study Ericsson ConsumerLab Mobile Broadband study 2009 - Austria[18] [16] that "the home environment is an important usage situation". Only 25% of the respondents considered the home usage as not very or not important at all. The majority of answers ranged from quite important to extremely important. Around 80% of those customers which were mobile broadband users only, i.e. they did not have and use fixed-line access, used it at home in a fixed place. A little over 30% used it at home in different places. Combined users — those with an additional fixed-line internet access at home — answered both questions at a rate of around 20%.

In the light of Cisco's projection of future data requirements it would be better for mobile operators to motivate all their customers to use fixed line access at home.

This consideration leads straight to the possibility of cooperations between fixed line and mobile access operators and probably to offer products like the following: a single device is offered to access the internet via both mobile networks and fixed line combined with a receiver installed inside the customer's home (much like wireless local area networks - WLAN). At home internet access is granted at high speed and without a download limit. Outside home access speed is lower and download volume is restricted. This hybrid product is available for one price on one bill. The customer is not aware anymore of the actual access technique he uses.

The outlined problem of cell congestion — too much data at one single antenna — is mainly a problem of densely populated areas. In urban outskirts and remote rural areas this is less

[15] AIM Consumer was developed in 1996 and has since established itself as an essential source of information about internet and new communication technologies. The AIM-C allows a representative and comprehensive insight into the views and usage patterns — representative for the Austrian population 14 years and older — based on 12,000 interviews per year.
[16] ge.: Hauptsächlich zu Hause
[17] ge.: Haupsächlich unterwegs
[18] Web based study with 753 Austrian participants

likely to happen. In these areas, however, the return of investment for improved and new fixed line access networks as it will be described in the following sections turns out to be very unattractive and in many cases not economically justifiable. Modern mobile access systems may offer an alternative way of supplying customers in such areas with high speed internet. Of course, mobile operators are not going to undertake this task without any advantage. A solution may be an attractive cooperation between mobile and fixed line operators, or a fusion between two such players.

In any case — reducing the role of fixed line operators to net-providers and wholesalers, tough competition or intelligent collaboration — the existing copper networks have reached their limits. Without incorporation of optical fiber technology into fixed line networks in one or another way it is not possible to increase the data rates in such networks any further. A whole bundle of strategies of how to do so became known in the industry[19] under the abbreviation

FTTx

which stands for Fiber To The x.

1.1.4 What is Fiber To The x? First answer

The first answer to the question what fiber to the x stands for can be found in the English version of Wikipedia which gives the following definition of FTTx [58]:

> *Fiber to the x (FTTx) is a generic term for any broadband network architecture that uses optical fiber to replace all or part of the usual metal local loop used for last mile telecommunications. This generic term originates as the generalization of several configurations of fiber deployment (FTTN, FTTC, FTTB, FTTH...), all starting by FTT but differentiated by the last letter, which is substituted by an x in the generalization.*

Wikipedia also gives a nice schematic comparison of the different strategies.

On the left hand side one finds the central office (CO) of the local area access network. The central office is the gateway of the local area access network to other local area access networks and hence to the entire world. They are connected via the backbone network. On the right hand side the home of the customers is depicted. In the "old"networks the connection between CO and customer is established by metallic cables (usually copper cables, coaxial or twisted pairs).

The key factor of this bundle of strategies is to "shorten"the copper part in the access network — the last mile which connects the customer to the internet, also called the local loop — by bringing optical fiber closer to him and her. In simple terms the effect of this shortening can be described in the following way: the shorter the copper part, the faster the internet access. But also, the shorter the copper part, the longer the fiber line which has to be newly constructed which makes the solution more expensive.

[19] Fixed line telecommunication companies

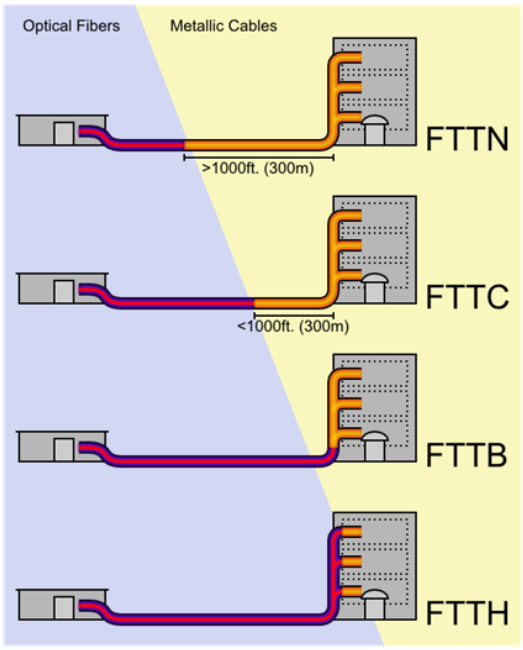


Figure 1.10 FTTx
Source: Wikipedia [58]

The speed of data transmission depends on two main factors, the medium of transmission, e.g. copper cables or optical fiber, electrical or light impulses, and the transmission protocol which defines the way of how data is transmitted. The speed of data transmission is measured by bit rates and the most common unit in use with high speed internet is megabit per second (Mbps).

In networks like the one of Telekom Austria where only twisted pairs of copper wires are in use ADSL 2+ is the standard transmission protocol for data transmission along copper cables between the central office (CO) and the customer. ADSL is short for Asymmetric Digital Subscriber Line and 2+ just denotes that it is an upgrade of previous standards ADSL 2 and ADSL. The term "asymmetric"refers to the property of ADSL that downstream data transmission (data transmission from the CO to the customer) is faster than upstream (from the customer back into the internet).

On short distances — up to 100 meters — and downstream a speed of 24 Mbps is achievable with ADSL 2+, whereas upstream it will be at most 1.4 Mbps. ADSL 2+ operates up to a distance between the CO and the customer of 6 to 7 km — depending on the physical condition of the copper network — and delivers a speed of 1 to 2 Mbps downstream. Beyond a distance of 7 or 8 km there is no signal transmission anymore. Most of the customers ($> 90\%$) are within a distance of 8 km to the central office. So, at the time ADSL 2+ is a strategy to provide most customers with good to decent internet access. But, it is a strategy with no future.

For this reason VDSL was invented. Its name — Very-high-bitrate DSL — says it all. VDSL is designed for a downstream transmission rate of up to 52 Mbps and an upstream rate of

16 Mbps. This is achieved up to a distance of 500 m. From there on transmission rates decrease (downstream $\sim$ 26 Mbps up to 1000 m, $\sim$ 13 Mbps up to 1500 m), until transmission deceases beyond 2000 m to 3000 m. Therefore, only a small to medium percentage between 15% up to — in best case — 60% of the customers can be reached directly from the central office (CO).

The strategy — to offer high speed internet by implementing VDSL at the central office — is sometimes also called VDSL@CO. This is the cheapest way to introduce high speed internet, since equipment has to be changed only at the central office. The CO usually belongs to the network owner and offers enough space and electricity. Furthermore, no alteration of the local area network itself is necessary. Clearly, only a fraction of the customers of the local loop can be reached by this strategy and not all of them will be able to utilize the maximal bandwidth which can be provided by VDSL.

To cover a higher percentage of the customers or even all of them FTTC extends VDSL@CO to a broader area by situating "copies"of the central office all over the local area network and running VDSL from there. Of course, these copies have to be connected to the central office themselves which is done by means of an optical fiber network.

1.1.5 FTTC

An access network can logically be decomposed into two major components: the physical network and the transmission protocol. The physical network is the medium for the signals which transport the data. The transmission protocol defines the way of how data is transmitted. Different networks and different protocols need different equipment and software.

The physical network of access networks which are designed based on the "Fiber To The Curb"strategy can be further decomposed into three components: the optical fiber network, the street cabinet which is missing in the Wikipedia graphic Figure 1.10 and the copper network.

The optical fiber network connects the street cabinet to the central office which itself is the gate to the internet (the backbone network). Usually, there exists no complete or connected fiber network within a certain local loop. It is therefore necessary to plan and build such a network which is one of the cost factors in FTTC.

The copper network connects the street cabinet to the customer. It already exists, and the FTTC strategy implies that the copper networks should usually stay unaltered. Consequently, no additional costs result from including the copper network into the new access network. In fact, this is the very reason why strategies like FTTC are considered at all. They are usually cheaper than connecting all customers directly by optical fiber.

The street cabinet serves as a translator between the optical fiber world and the copper world. Clearly, the two mediums are different and the transported signals need translation. Moreover, the street cabinet houses the equipment and the software for the transmission protocol which is used to transport the signals to the customers via the copper network. And — as its name indicates — the street cabinet is usually situated outdoors next to a street and therefore needs some shelter and electrical connection. In rare cases, if an adequate location is available, the

equipment may also be situated indoors, e.g. in a separate room in the basement. Street cabinets and their installation are expensive. They are the second contribution to the construction cost of a new access network.

Since street cabinets are central to the FTTC strategy, the C in FTTC can also be read as cabinet. This actually gives a motivation to subsume the "Fiber To The Neighborhood"(FTTN) and "Fiber To The Curb"Strategy under FTTC (compare with Figure 1.10). They are essentially the same. The only difference is the maximal admissible distance between the street cabinet and the customers. The damping of the transmission signal is an important factor of the transmission quality, i.e. of the bit rate which can be provided for a customer. Damping increases with increasing distance between source and recipient. As a consequence transmission rate decreases. Hence, the smaller the largest distance is between the street cabinets and the customers they supply, the better is the transmission quality of the newly designed access network for an entire area.

Of course, the closer street cabinets have to be situated to their customers, the more cabinets are necessary, and the wider and longer the optical fiber network becomes. Construction costs of the access network increase. Depending on the size of the area of the local loop and the number of customers which have to be supplied construction costs of several hundred thousand to millions of Euros have to be expected. But, FTTC is a very general and unspecific planning strategy. It does not say anything about how to realize the cheapest possible FTTC access network.

This became one of the main tasks of SARU, a project defined and realized by Telekom Austria. The abbreviation stands for: Situating Access Remote Units. Access remote units are another term for street cabinets. The expression replaces the idea that street cabinets have to be placed next to a street. The goal of SARU is to develop and deploy software which supports the structural planning process of access networks following a FTTC strategy. It also enables the planner to manipulate and change the thereby found solution interactively. The software was named after the project.

1.1.6 FTTH

On very short distances (less than 300 m) and optimal network conditions (e.g. cables are indoors) VDSL2, which is an upgrade of VDSL, may even provide bandwidths of up to 100 or even 200 Mbps. With an end-to-end connection via optical fiber cable transmission rates start at a level of 100 Mbps — for one single customer.

This is due to the transmission medium light, where at least over short distances (a few kilometers) damping is not such an important factor and crosstalk or interference between parallel optical cables is no issue at all.

The protocol in charge with the data transmission in optical networks is called GPON which stands for Gigabit-capable Passive Optical Network. Again, the name says it all. Downstream 2.4 Gbps (Giga bit per second) are achieved on a distance of maximal 5.7 km which will be extended to 17.5 km in future. Upstream 1.4 Gbps are offered. Distances and bit rates also depend on the equipment in use (e.g. quality of the lenses of lasers) and variants of the protocol.

Moreover, bit rates do not continuously decrease as it is the case in the copper network. The same bit rate is available from the central office until that point — 5.7 km or 17.5 km further on — where the signal vanishes.

Of course, in mass markets of private customers and small to medium enterprises there is no need for such a high speed internet at the moment. Only companies of appropriate size and special needs will buy a single line. All other customers utilize the optical network by sharing lines. This is part of the FTTH strategy.

By the use of so called splitters up to 16 customers may share one optical fiber cable and are guaranteed a portion of the maximum available bandwidth. With 2.4 Gbps evenly split between 16 customers this still corresponds to 150 Mbps per customer downstream and 75 Mbps upstream.

The FTTH (Fiber To The Home) strategy avoids the copper network in order to provide the highest possible transmission rates. All customers are connected to the backbone net via optical fiber cables. One positive effect is that access remote units (street cabinets) are then unnecessary, since copper cables are not part of the network. There is no need for translation anymore. But, the savings gained by the absence of street cabinets is outperformed by the need to roll out a completely new optical fiber network over a large area. In this context the use of splitters is another advantage because network size and consequently investment costs are reduced, since customers are not provided with their individual optical fiber line.

So, ideally all customers sharing one splitter live next to each other, for example in an apartment house. Thereby, the number of optical fiber lines in the entire network is reduced by a factor which corresponds to the maximal number of customers served by a splitter. At the moment up to 16 customers my share one splitter[20]. In future this may be possible for 32 or more customers.

This technology reduces cable cost for optical lines. But, it also saves construction cost for the new access network because existing infrastructure like empty ducts are more efficiently used and the need for new trenches decreases. Digging is the most important cost driver in network planning.

Of course, the splitter equipment is not for free and because of the smaller equipment to customer ratio (e.g. 1:16 for splitters and 1:196 for street cabinets) a lot more splitters are needed than street cabinets. But, splitters are also considerably cheaper than street cabinets. They are quite smaller than street cabinets and adequate locations are easier to find, e.g. at a wall in the apartment house's basement. And they do not need power supply which is also a very important cost driver.

[20] A splitter actually does not split the data arriving downstream from the internet — as one might imagine — into separate packages which are then sent to the customer to which they belong - like the postmen do when they deliver the mail. It sends copies of all the incoming data to every customer it is connected to. It is the customer's modems task to select and concatenate its masters data. Eavesdropping is prevented through encryption. Upstream the splitter collects and merges its customers' data. In this sense, a splitter is rather a copier and merger than a splitter.

There are other abbreviations and strategies known in the industry which are similar to FTTH, e.g. FTTT which stands for Fiber To The Toilet. Although, this sounds more like a joke, it has to be taken literally. FTTT expresses the idea to enter the home of the customer along the sewer system through the toilet, one of its broadest gate. The sewer system is an existing system of trenches which already connects and concentrates ducts of widespread customer sites at certain points — an ideal system to minimize digging. The challenges are to insert optical fiber cables into the sewer system and to keep the two worlds of waste water and telecommunication infrastructure clearly separate. Technologies to achieve these goals are available.

FTTB (Fiber To The Building), as the name indicates, aims to bring optical fiber to or into the building but not to the customer. The last yards rather than the last mile are still copper lines. The key idea of this strategy is that the network provider does not have to enter the individual locations (i.e. apartments or offices) of the customers. Its advantage arises if house owners or other authorities in charge of the location are not willing to open doors for the provider. Or if the indoor infrastructure is just not fit to house additional equipment. However, in terms of network quality this strategy can not achieve the same transmission rates as FTTH does. In terms of cost it is nearly as expensive. In this sense FTTB is only an alternative to FTTH, where the latter is not possible and the former finds optical fiber infrastructure near by.

1.1.7 FTTx: Second answer

The summary of the last two sections provides a second answer to the question what fiber to the x stands for: The two main strategies to plan and build new fixed line access networks based on optical fiber technology are FTTC and FTTH. A combination of the two is a third option — a mixed planning and deployment strategy. Densely populated areas could be accessed directly by optical fiber cables, i.e. FTTH or FTTB depending on the local circumstances. Widely spread out areas or neighborhoods with many detached houses may be connected to the internet by FTTC saving costs by using the remains of the copper networks.

FTTx is also used as an abbreviation for this mix strategy. It was studied and realized in a second phase of Telekom Austria project SARU and is not part of this work.

In Berlin a group of researches from several institutions started a cooperation with the working title FTTx in 2010[21]. Andreas Bley (TU Berlin), Axel Werner (ZIB) and Roland Wessäly (atesio) were working together in a cooperation to develop a software which enables the user to cost-optimize a FTTx planning strategy. By FTTx the Berlin group focused on FTTB and FTTH. FTTC was not a topic. Their understanding of FTTx excluded FTTC.

[21]The point of time when they had started the project shows two things. Firstly, it is very unlikely that a commercial planning software for this type of problem was available or at least well known and widespread at that time. Secondly, some German network providers were optimistic to be able to make the fiber business case a success.

1.2 Project: The assignment

At the end of January 2006 the head of infrastructure wireline — a division within Telekom Austria which is in charge of all planning activities with respect to the access nets — invited the Operations Research group of Telekom Austria to assist him and his group in two different planning problems. The first problem was the task to find and plan the cost optimal copper access network based on a greenfield approach. I.e. in the area for which planning is supposed to be performed no telecommunication infrastructure exists at all, e.g. a city or a commune decides to build a new neighborhood. The second problem was the optimal positioning of access nodes — i.e. street cabinets — within already exiting infrastructure and was to become project SARU.

During the following weeks a group of five people met regularly and collected and discussed all specifications and requirements which seemed necessary at that time. Section 1.4 provides an excerpt of those specification which are relevant for the problem as it is discussed in this theses. Later more colleagues were invited to participate in this process as they were needed. Experienced planners joined the group to enter their practical critique and to add additional requests. Data base experts had to be consulted, in order to enable us to utilize different data sources for the project. From today's perspective all these requirements conformed to the FTTC network design strategy. But, at the time I was not familiar with terms like FTTC or facility location at all.

Consequently, I sought help and information. My Ph.D. advisor Prof. Immanuel Bomze from the University of Vienna put me into contact with Ivana Ljubić who had received an award from the Austrian society of Operations Research (ÖGOR) for her thesis [40] about two network design problems one of which being the prize-collecting Steiner tree problem.

Among many other things Ivana Ljubić established the contact to Johannes Hackner from the energy provider EVN (Energie Versorgung Niederösterreich). In his doctoral thesis [24] he dealt with a network design problem for district heating with the final goal to layout the network in such a way to maximize revenues. Although, the problem seemed to be quite like ours, there were a few important differences. The energy problem resembles more a greenfield approach. There is hardly any existing infrastructure which puts certain side constraints on the problem. The FTTC strategy for telecommunication access networks is firmly based on the requirement to use existing infrastructure and add new infrastructure where it is needed. Also, maximizing revenues implies that unprofitable customers are not going to be connected to the heating system. In our case there was no thought about just integrating a subset of customers into the new access net. Furthermore, it was and still is a lot harder to estimate revenues in telecommunications over a longer period of time, since prices continued to fall (see Section 1.1.1) and customer loyalty is not as strong as in the energy sector.

Still, the contact was very helpful because Mr. Hackner developed a planning tool called "ex-Plan"which integrated a Geographical Information System software (GIS) and his optimization algorithms. This was necessary for him, because his work was not only of theoretical interest, but was supposed to be deployed in his company. Later this approach became a motivation for me to realize a prototype of the SARU planning tool with all the interactivity which was demanded by the planners.

Meanwhile, however, Ivana Ljubić and I discussed and found formulations for the problem at hand. Finally, we established a description which was first presented at the European Conference on Operational Research (EURO XXI) in Iceland from July 2-5 in 2006 [57]. The topic was also addressed to an audience of Ph.D. students and their advisors during a workshop for doctorial candidates on topics of network optimization in Lamprecht (Germany)[22]. And finally, on November 28. and 29. of the same year an industrial conference called "Optimizing DSL"took place in Lisbon where the problem was presented to representatives of fixed line operators from all over the world. Throughout these presentations our main questions was, if anybody knew of techniques or an existing tool or a tool in the development stage which dealt and solved the stated FTTC problem. Although, single parts of the problem were already well studied (like facility location problems), we found no hints towards an existing solution of the compound problem nor availability of tools.

To secure a long term collaboration Ivana Ljubić and I agreed upon initiating a cooperation between the University of Vienna and Telekom Austria AG which was successfully established in June 2007. The project was co-funded by the Austrian Forschungsförderungsgesellschaft (FFG) and ended after 3 years in 2010 after the end of SARU Phase 2.

The project plan stipulated two major project parts. To be able to produce first results in a relatively short time the FTTC planning problem was to be solved iteratively by methods already available in the scientific community which needed just a few adaptations to the given problem. Access remote units should be situated by an adequate algorithm which solves a facility location problem. Subsequently, the connection of the previously determined ARUs to the central offices (CO) had to be established by an appropriate algorithm for network design like the Steiner tree problem. After having solved this approach the goal of the second part was to tackle the combined problem of simultaneously opening facilities and connecting them by an optical fiber network to the CO.

At the time of writing SARU consisted of three different phases. Phase one ended sometime during the year 2009. This thesis is mainly concerned with SARU Phase 1. The second phase ended in summer 2010, and since then we have been evaluating whether a third phase was needed. These phases, however, have nothing to do with the structure of the cooperation project just mentioned above. We had to slightly deviate from the project plan due to another topic: data problem.

[22]Workshop on Network Optimization, October 23-24 2006 Lambrecht Germany, Organizing Committee: Horst W. Hamacher, Sven O. Krumke, Dwi Retnani Poetranto, organized by the optimization group of the University of Kaiserslautern

1.3 Data base and its challenges

Parallel to the literature search and theory study we investigated and analyzed data sources. Our first goal was to establish an example of a data representation for one local loop which could be used for testing the algorithms. Further objectives of this part of the project were

2. to investigate and analyze the data quality and eventually identify and solve data problems,
3. to develop a standardized data interface,
4. to produce all necessary data preprocessing algorithms for the optimization, and finally,
5. to set up a fully automated data preparation procedure.

The main problem in this context was clear to us from the very beginning:

1.3.1 Digging and unconnected graphs

For the optimization algorithms the network of the operator (copper, optical fiber network, pipes and so on) is best described as a mathematical graph. To generate a network graph we established a layer model for the different components with which an optical fiber network may be constructed (see Figure 1.11).

The cheapest layer to incorporate in the new network is the first: dark fiber which are existing optical fiber cables owned by the operator and already lying in the ground but not being used

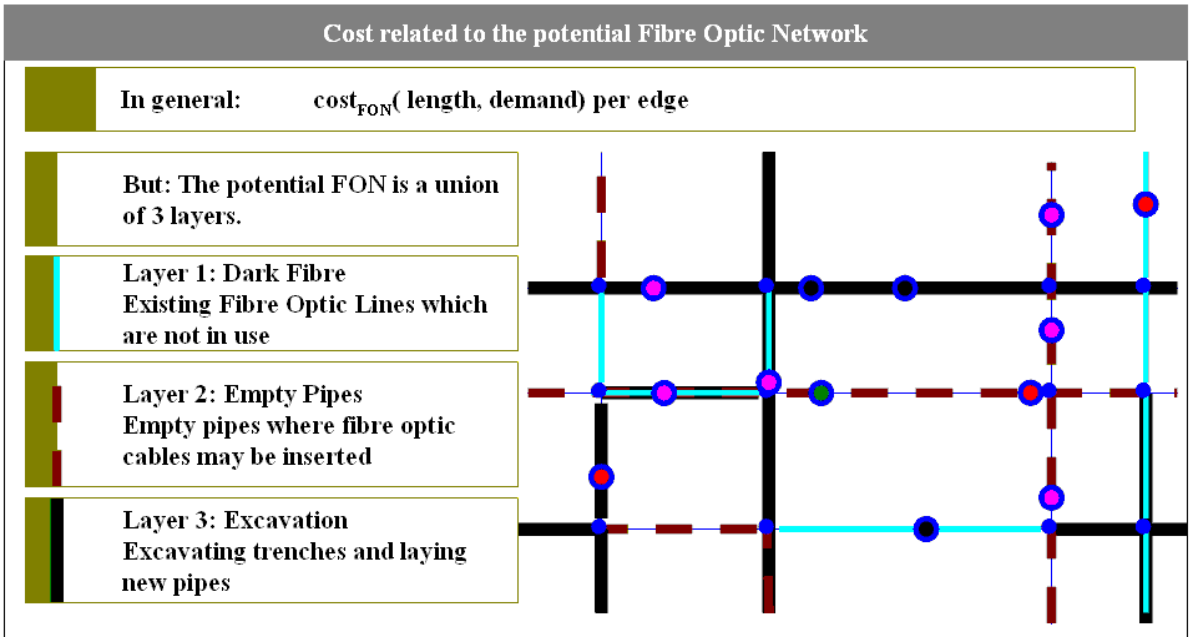


Figure 1.11 Layer Model of SARU

(layer 1: dark fiber). They only have to be activated with the sole cost being labor. The second layer (layer 2: empty pipes) consists of all empty pipes and similar constructions lying in the ground which can be easily accessed and filled with optical fiber cables. To use them for the construction of a new optical fiber network is more expensive, since material cost (optical fiber cables) and significantly higher labor cost (injection of fiber cables) accumulate.

These two layers comprise of all the inventory a network owner can use to plan and construct an optical fiber network. However, to express it in graph theoretical terms, it cannot be assumed — on the contrary — it has to be expected that the graph resulting from the union of these two layers is disconnected. As can be seen from the schematic drawing of Figure 1.11 the fat black lines of the third layer close the gaps between layer one and two (layer 3: excavation). The gaps, in turn, result from the evolution history of the pipe networks. Sometimes the network owner gets to know of some local construction work in the ground (e.g. due to a broken water pipe or a gas leak). Then the network owner may take advantage of the situation and bury some equipment together with what ever else is laid there into the ground. This is a lot cheaper than performing the construction work just by one company.

Layer 1 and 2 are represented in Telecom Austria's inventory database called WebGIS. Since it is an inventory system, layer 3 is certainly not part of it. Several other sources exist which can be utilized to constitute layer 3.

1. *The copper network*: the copper network certainly can be represented by a connected graph. Anything else implies a dysfunctional network. At some point in time the copper wires were buried. Therefore, the copper lines show a path along which a new digging is possible. These paths, however, may not be the shortest. Or, there is already a lot of equipment in the ground such that it is impossible to put more. Or, there may be other reasons.

2. *Tele Atlas*: Tele Atlas is an international company producing and maintaining digital maps especially, road and city maps. These roads, streets and lanes are possible guidance for digging lines and trenches.

3. *Digitale Katastralmappe*: The Austrian DKM is a digital map of all private or public real estate in Austria. Consequently, the DKM also contains all public streets. One potential advantage of the DKM compared to digital street maps is that the available information is more spatial, i.e. streets have also got a width, which is important for calculating street crossings. The disadvantage is that thereby the size of the database grows enormously.

Another disadvantage arises with sources 2 and 3. There is no logical connection between them and WebGIS, the net inventory of Telekom Austria. Of course, there is a geographical connection between them. All three of them use geographical coordinate systems which are compatible. To simply project Telekom Austria's net elements onto a street map or the DKM may cause misleading or even wrong logical connections between different components. So, some thought and work would have to be invested to solve this problem.

1.3.2 Copper net and trees

The study and analysis of the structure of the copper network is of course at least as important as that of the potential optical fiber network. As depicted in Figure 1.12 we liked to envisage the copper network as a tree: the central office being the root right in the middle of the local loop and the leaves being sites where customers live. Additionally, there are switching nodes between the leaves and the root some of which may also be customer sites. This chart is taken from the presentation I gave in Reykjavik in 2006 and all the presentations which would follow. Unfortunately, this image is wrong.

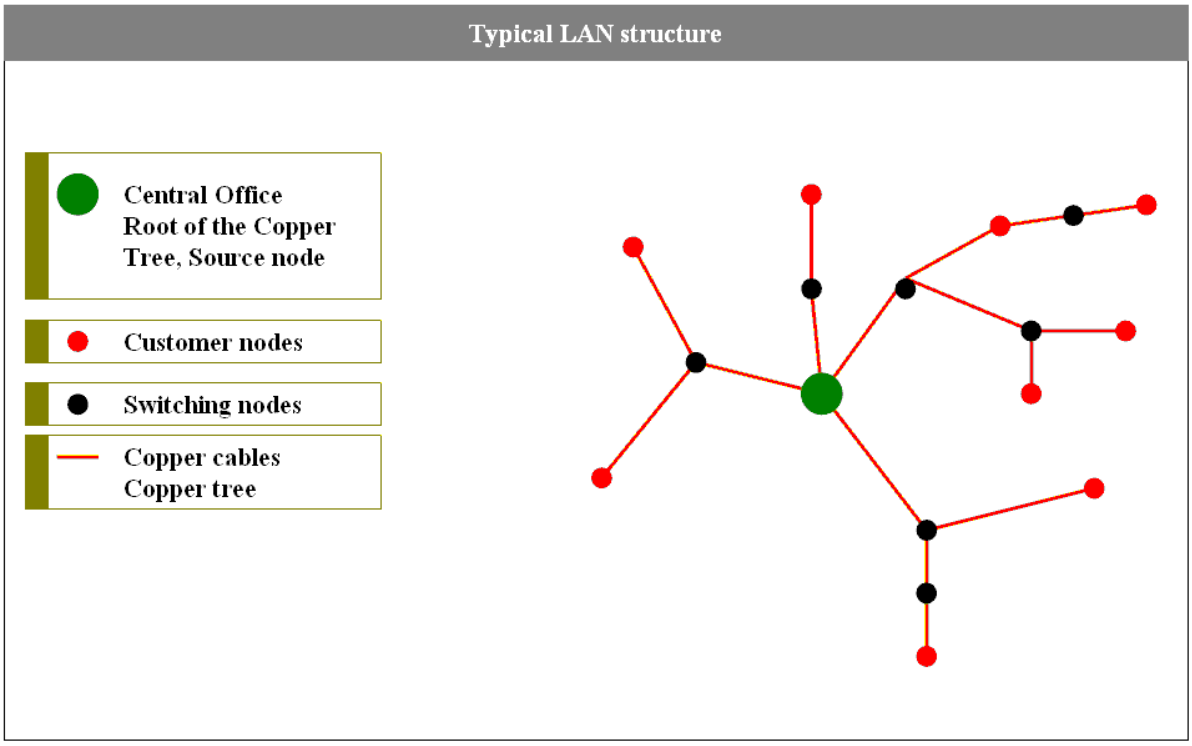


Figure 1.12 Copper Tree, Source: SARU presentation at EURO XXI 2006

Perhaps this sounds a little bit surprising, since every customer is connected to the central office by a unique path — the copper wire — which allows undisturbed and undivided access to the internet at any time. Certainly, when viewed like this the copper net can be represented by a tree graph, to be precise, by a star: the CO node in the middle, customer nodes adjacent to the CO node and no switching nodes in between.

But, the star model is not at all useful to describe the access network for the FTTC problem. The central goal of the FTTC strategy is to identify copper centers, locations where a lot of these individual wires meet and where they can be raised to be connected to an access remote unit. Additionally, such locations are sought relatively close to the customers. Therefore, intermediate nodes and the trails of these wires — a collection of pipes, ditches, trenches and cables — are important. The graph of these trails[23] does not have to be a tree which is

[23] See Section 1.6.3 Definition 6 for a formal statement

illustrated by two examples.

1. Example: Missing parallelism

Figure 1.13 shows the situation of two wires running parallel when leaving the central office. At intermediate node u_1 their paths separate. $wire_1$ follows the main trail, passing intermediate nodes u_2 and u_3 to rejoin again with $wire_2$ at node u_4 in trail t_{main}. The second wire follows a side trail which was probably constructed as a later add on to the already existing main trail, passing node u_5 where probably a customer sits for whom this side trail was constructed.

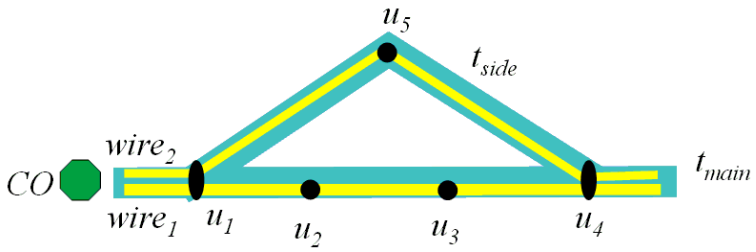


Figure 1.13 Example for wires which do not run parallel through the copper net

The graph derived from these trails, however, contains a loop. A possible solution of this problem is to forbid that $wire_2$ is assigned to an access remote unit at location u_5, in case an ARU is situated there. To achieve this the edge from u_4 to u_5 is removed from the graph and the path of $wire_2$ (and any other wire using edge (u_4, u_5) has to be virtually altered into a path with the segment $u_4 \rightarrow u_3 \rightarrow u_2 \rightarrow u_1$. Of course, $wire_2$ must not be assigned to an ARU which is eventually situated at node u_2 or u_3 because there is no physical connection.

There are more problems connected to this example which will be discussed later.

2. Example: Backward supply (Rückversorgung)

The technique of backward supply (German: Rückversorgung) was adopted for the construction of copper networks to establish points of high wire concentration: cable branching points (German: Kabelverzweiger). The idea of a cable branching point can be illustrated as follows. Uplink, the branching point is connected to the central office by a cable containing 100 wires, for example. Downlink, 200 wires leave the branching point in the direction of the customers. Of course, this implies that only 100 of the 200 customer's wires can actually be connected to the backbone network. Only 100 physical customers can be supplied from that point. But, nobody knows in advance where in the vicinity of the branching point these 100 customers are going to be. The advantage of this strategy is that wire capacities which are set free in one part of the vicinity of the branching point (e.g. customers cancel their contracts) may be easily reused in another part by switching connections at the branching point. That's why such points are also called switching points.

This strategy combined with the attempt to save construction costs (e.g. digging is more expensive than cables) leads to backward supply as illustrated in Figure 1.14.

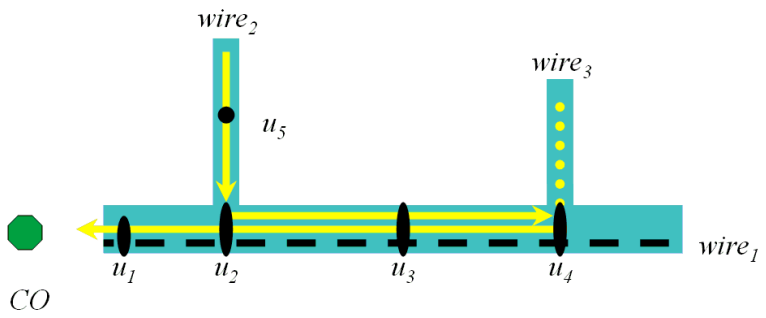


Figure 1.14 Example for a wire connection which runs trough a single trail twice

In this picture node u_4 is the branching point. Both wires, $wire_1$ and $wire_2$, arriving from the central office pass through nodes u_2 and u_3 before arriving at u_4. There, $wire_1$ is continued into the "right"vicinity of branching point u_4. However, $wire_2$ is reversed and continues backward passing again through u_3 and u_2 to finally branch into the "left"vicinity of u_4.

This construction pays off as soon as the customer using $wire_2$, for example, cancels his contract and some time later another customer orders a new connection line which can be fulfilled with $wire_3$. The remaining part of $wire_2$, from u_4 to u_1 and further, which is now unused can be switched, i.e. reconnected to $wire_3$ at branching point u_4.

If the detour of $wire_2$ is avoided, i.e. $wire_2$ runs from u_1 over u_2 directly to u_5, then either $wire_3$ would need a direct connection to the CO or switching activities would become necessary at nodes u_2 and u_4. Since the ratio between uplink and downlink wires at the branching point is much in favor of downlink wires like $wire_3$ both alternatives would produce significantly higher costs, either because the total number of wires in the access net or the number of switching nodes increases. The additional cost for the "backward"wire u_4 to u_2 is relatively low, especially because it uses the same trail as the forward part. No additional digging is necessary.

The graph which can be derived from the trails is — at least in this vicinity — a tree. However, the supply path of $wire_2$ is not tree-like. It contains a loop. But, the challenge of describing the copper network by a network graph is to find a graph which allows to reproduce the supply paths of all wires of the access net one to one. So, either a desired network graph also contains loops, or the supply path of $wire_2$ is altered virtually in such a way that it does not contain a loop anymore.

It should be noted that branching point u_4 is an ideal candidate for the location of an access remote unit (ARU), simply because it is already a concentration point of copper cables. In practice such points are the preferred choice. Rather than situating an ARU at nodes u_2 or u_3 a planner would place it at u_4. This opens up a possibility to resolve the loop.

1.3.3 Digital pipe net and availability

It is mandatory that the source of the data on which the optimization algorithms are supposed to operate is a digitalized database. Digitalization in form of digital pictures of maps of the

inventory is not useful at all. Lists of all elements of the network inventory be it part of the copper network or pipes or optical fiber, together with their properties and their relation to each other are needed.

Telekom Austria is in possession of such a database. It was already mentioned and is called WebGIS, "Web"because it operates on a web based client system, GIS because it is a geographical information system. Telekom Austria was in possession of WebGIS already in 2006. The digitalization process was already in progress but by no means at an end. Not all the copper networks were digitalized at the beginning of 2006. The work on the networks of pipe and dark fiber had not started.

Certainly, all the inventory was documented, just not in a single database. Parts of the information were kept in old and different databases, others were digital pictures of maps. The digitalization made it sometimes necessary to go on-site and check whether all information was correctly collected. So, the process was cumbersome, time consuming and expensive. One local loop may take between 1 to 3 months and cost can be measured in tens of thousands of Euros. Austria has got more than 1.400 local loops.

As a consequence we had to diverge from the plan (see Section 1.2) of the cooperation between University of Vienna and Telekom Austria, because an integral part of this plan was a complete data set for a local loop to test the algorithms which were going to be developed. Instead we decided that the group of the University Vienna would continue their theoretical work without realistic test cases, so that they would be ready to apply their findings to the real world as soon as data was available. It was my task to collect and coordinate all of the old and new requirements and maintain communication between the project partners. Soon, it became an urgent point to deliver practical results despite the fact that not all of the necessary data was available. Therefore, I started to apply facility location techniques to the problem of situating access remote units without considering the supply with optical fiber cables. That is where the name SARU came from: Situating Access Remote Units. The name stayed the same, even after tasks entered the project which had nothing to do with access remote units anymore, e.g. FTTH.

The first phase of SARU, however, was that part of the project where most of the requirements and specifications were negotiated and documented, and a prototype software tool was developed which comprised already of some optimization abilities — lacking the planning of the optical fiber network — and much of the interactivity the planners were asking for. This thesis is about this work.

1.4 List of specifications and rules

The following section provides a list and descriptions of the most relevant specifications and requirements which were developed during the three years of SARU Phase 1. These specifications were the result of a process and discussions which involved several colleagues and experts in different fields of knowledge as it was already mentioned in Section 1.2. The list focuses on rules which are relevant for the cost optimal FTTC planning strategy. It does not contain those rules which deal with features of the software tool SARU, especially rules concerning the interactivity of the planning tool. The specifications — here called rules — are not numerated according to chronological order, nor does the order reflect any kind of importance or hierarchy of the rules. The order is due to a post-phase reflection of the specification process and its results. The order is such that the rules are easily explained and understood.

1.4.1 Rule 1 (R1): Unaltered copper net

This rule simply states that during the automated planning process the existing copper network should not be altered. In terms of project management this is a non-goal. This does not imply that such alterations cannot be useful or cheaper in some cases. In fact, planners reported several instances where they considered changing the copper net at certain spots or under special circumstances. The desire to manipulate the existing network is best documented in the software tool itself.

One of the requirements for the interactive part of the tool is the possibility to assign certain customers to certain ARUs at the will of the planner, even if there is no physical connection, i.e. a copper wire, between the ARU and the customer node. In subsequent optimization runs these customers have to stay assigned to their ARUs. This consequently implies that these ARUs — being optimally chosen or not — had to stay in the solution, too. Of course, the optimization algorithm has to respect these fixed assignments. For the planner these fixed assignments without physical connections have to be specially reported, when he moves on to the detailed planning process. That way he can ensure all necessary means to resolve the missing connections.

Altering the copper net is the responsibility of the human planner.

1.4.2 Rule 2 (R2): Customers and network graph

In the SARU data base of the copper access net customers are represented by special nodes called Kabelausmündungen (KA - cable outlet point). These customers will also be called logical customers. From the Kabelausmündungen individual wires (TNAK: Teilnehmer Anschlusskabel - subscriber's access cable) run to the physical customers: households, business offices, stores and so on. In most cases these cables are relatively short (a few meters). In some cases (rural areas in mountainous regions) the length may be as long as a few kilometers. They are not documented in WebGIS.

The last point of documentation for every physical customer is the KA. Therefore, the KA becomes the logical customer. Consequently, all leaves of the network graph have to be KAs.

There are cases where a KA is not a leave. This does not happen very often and the subtree rooted at such a node is very small. Most of the time it contains a SubKA, a sub cable outlet point. The situation may be simplified by assigning the demand of the SubKA to the corresponding KA and removing it afterwards.

1.4.3 Rule 3 (R3): Customers and their demands

In general one KA represents several physical customers. The KA will be called the logical customer or even shorter just the customer. The demand of such a logical customer will comprise a certain number of physical customers which are somehow related to the switching point — the KA — in question. Alternative approaches of defining this relation are described in what follows.

Three different ways of determining the size of customer's demands were developed over the years. The KA — the cable outlet point — has got a certain maximal capacity which ranges form 10 to 200, 500 or even more. Small ones are more frequent. This maximal capacity can be the base for the definition of the demand. However, this is a very unrealistic approach. Line loss as described in Section 1.1.1 by Figure 1.4 certainly had the effect that not the entire capacity of a KA is actually active.

Hence, an alternative base of the definition of customer demand at a KA is the number of active lines at the KA. Now, it can happen — again because of line loss — that there are no active lines at a certain KA at all. All customers have canceled their contracts. In this case the cable outlet point in question will not be considered during optimization. To prevent this happening zero demands may be replaced by some positive default value.

The third and final version of the count of demands is an assignment of postal addresses to KAs, cable outlet points. It is known approximately how many private and business addresses are in the vicinity of a cable outlet point to which they are connected.

To increase the ways of defining customer's demands even further it is also possible to take a certain percentage of the base number of ones choice. Customer demand may, for example, be defined as

- 50% of the maximal capacity of the KA, because it is too optimistic to assume more than that, or as
- 115% of the number of active lines, since it may be assumed that all the existing customers are going to use ARU services eventually and some 15% of new customers may be found, or as
- 80% of all accessible postal addresses, because not all households and businesses in the vicinity will subscribe to new services, but some — especially business — will need more than one line.

However, there is at least one method of defining customer demands in the context of access net improvements which was never considered during project SARU. The FTTx strategies aim

to increase the transmission rates for customers. Therefore, customer demand could also be defined on the bases of bandwidth demands individual physical customers have. The demand of a logical customer would then be the sum of these individual bandwidth needs and measured in Mbps rather than in the number of subscriber lines as it is defined in SARU.

1.4.4 Rule 4 (R4): Undivided customer demand

The demand of one customer, i.e. of a KA, a cable outlet point, has to be assigned to and supplied by one access remote unit only.

As pointed out in Section 1.4.3 the demand of one customer site will in general be larger than 1. It may be an advantage in certain cases to split the demand of one customer site and satisfy it from different access remote units. This may happen, if capacities of ARUs are limited, for example, or some ARUs are charged with a lot of assignments whereas another ARU close by is nearly empty. But, dividing customer demands is not allowed due to technical restrictions.

1.4.5 Rule 5 (R5): Full coverage of customer demands

According to Section 1.4.3 the demands of customers may be defined as a percentage of one of three base numbers: maximal number of active subscriber lines, number of active subscriber lines and number of postal addresses. In case it turns out that some of the logical customers have zero demand, these demands may be replaced by a positive default value. But, after this definition step every customer has got a fixed (positive) demand.

Rule 5 states that this demand has to be completely covered and satisfied by the solution for the FTTC problem. Especially, every logical customer has to be assigned to an access remote unit which — as a consequence of rule R4 Section 1.4.4— attributes to 100% of the demand which is generated at the customer site.[24]

1.4.6 Rule 6 (R6): Admissible locations for ARUs

The ARU, to which a customer is assigned, has to be situated somewhere along the copper wire which connects the customer to the central office (CO). This copper wire defines a unique path on the network graph which will be called supply path[25].

Theoretically, ARUs can be erected anywhere along the supply path. Because of Rule 1, unaltered copper net, they must not be placed off the supply path. Otherwise, reconstruction and therefore alteration of the copper network becomes necessary.

The copper network, or better, the description of the copper network in the data base provides the supply path with different nodes. For example, the two end-points of the supply path are the central office and the logical customer (see Section 1.4.2), the switching node called KA

[24]The author cannot recall from memory or find a written document which specifies this rule. However, the rule describes what we actually believed and implemented. And, at least initially, the approach was not questioned. Though, for the second part of SARU Phase 1 this rule became the obstacle which forced developments into a different direction. This is a second reason to recognize and document the implicit rule "Full Coverage"with its own number.

[25]See Section 1.6.3 Definition 5 for a formal formulation.

or cable outlet point. In between there are more nodes which function as switching nodes like KV[26] and LV[27] or simply as muffles. A typical supply path is shown in Figure 1.15.

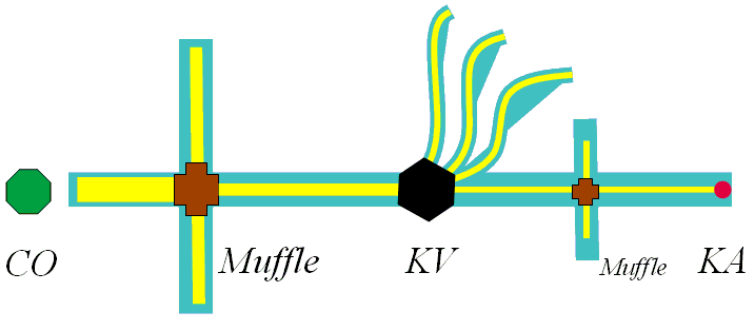


Figure 1.15 Typical supply path

Admissible locations for ARUs are all nodes of the copper network along the supply path including the customer node and the central office. ARUs must not be situated between these nodes, on or off the supply path.

One reason for the decision to use only nodes of the supply path was that these nodes allow easy and therefore relatively cheap access to the copper network. Cutting lines in the middle of the trails would also make a rebuilding of the copper network necessary which is prohibited by Rule 1. And finally, it is the task of the planner during the detailed planning phase to determine the exact location of an ARU. Too many aspects have to be considered to find the appropriate position. It is impossible to determine all of them before the structural planning phase to have them available as input for an optimization algorithm. It would therefore be to no purpose to first solve a rather complex continues optimization problem which subsequently has to be changed locally for every ARU by hand.

Like in Section 1.4.1 it should be added that cutting lines on the trail could be useful, lead to a better solution and be even cheaper. However, alteration of the copper net are no objective for the automated planning process. It is jurisdiction of the human planner.

1.4.7 Rule 7 (R7): Disjoint supply areas of ARUs

Since access remote units are links between the worlds of copper and fiber (see Section 1.1.5), every single ARU acts like a central office within its local loop. An ARU rules over a small area of the original copper network supplying thereby all customers within this area. It is the task of the optimization to completely cover the original local loop with such areas. However, care has to be taken that such areas do not overlap. Figure 1.16 gives an example of overlapping supply areas.

$customer_{1a}$ and $customer_{1b}$ are assigned to ARU_1, whereas the other two customers are assigned to the second street cabinet. Services are provided along the solid yellow lines for the

[26] Gr. Kabelverzweiger, cable switching point
[27] Gr. Linienverzweiger, line switching point, similar to KV, but higher in the hierarchy of switching points.

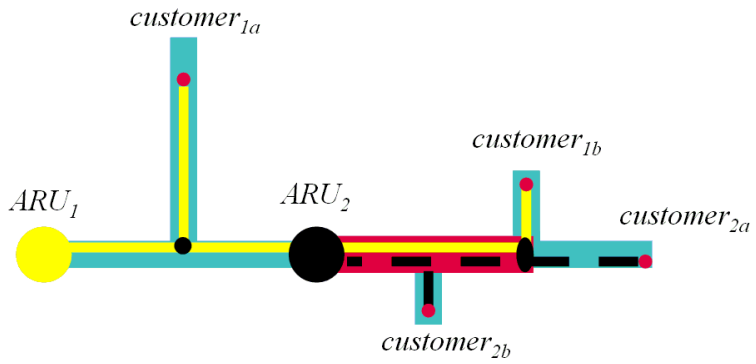


Figure 1.16 Overlapping supply areas of two different ARUs

customers of ARU_1, and along the dotted black lines for customers of ARU_2. This assignment of customers causes the two access remote units to share infrastructure in the highlighted red area. Rule 5 forbids that something like that happens. Electromagnetic interferences between the signals which are transmitted along the two copper wires within the shared area may cause bad service quality or even complete malfunction.

A similar problem occurs in case of backward supply (compare Section 1.3.2). It is illustrated in Figure 1.17.

An access remote unit at node u_1 supplies $customer_{1a}$ along the solid yellow line which depicts $wire_2$. This copper cable passes through the highlighted area twice because of backward supply which leads to a similar problem as before now involving only one wire. This problem may be resolved either by switching the connection of $wire_2$ at node u_2 which is anyway forbidden by Rule 1, or by situating the access remote unit at nodes u_4, u_3 or u_2.

In other words, $customer_{1a}$ should not be assigned to an ARU at a location which is closer to the central office than node u_2.

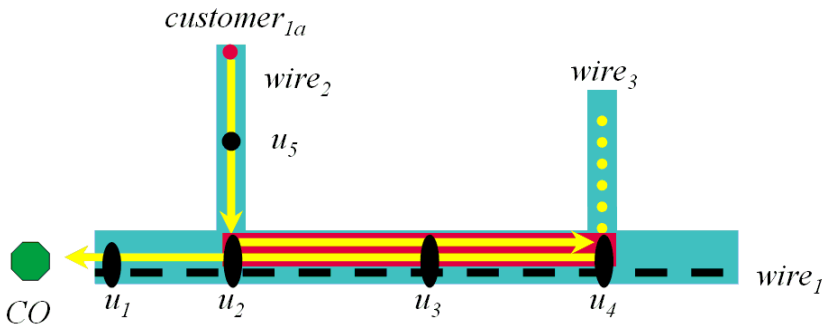


Figure 1.17 Overlapping supply areas because of backward supply

1.4.8 Rule 8 (R8): Limitation of ARU capacities

The capacity of access remote units is limited in regard of two aspects.

- The number of copper wires an ARU may hold is bounded.
- The space for the equipment which is necessary for signal transmission and consequently for providing services is finite.

Typically, an ARU can house less equipment for transmission than it can take copper wires. This is intentional and correct, since in future not every wire will carry services, but it is not known in advance which wires will. An ARU also functions as a switching point in the future network.

Therefore, the capacity of an ARU is usually defined by the number of pieces of equipment it may hold which corresponds to the number of customers an ARU may simultaneously provide services for.

The capacity of ARUs varies.

- There exist different types of street cabinets which allocate more or less space. However, they should not be too large, since otherwise they need air conditioning which makes solutions more expensive and may lead to conflicts with the people in the vicinity of the street cabinets. Cabinets are not very aesthetic, and air conditioning may be noisy.
- More than one street cabinet may be established at one location. This is not a very popular solution, since again problems with the neighborhood are possible consequences.
- An ARU may be set up indoors in case an adequate room — e.g. in a basement — can be provided. This is hardly known in advance of the planning process, because such adaptations make negotiations with house owners necessary.
- The central office is wrongly addressed as a remote unit. The remoteness of an access unit expresses exactly that it is set up in a certain distance to the CO. Nonetheless, the CO may house equipment and provide services like any access remote unit and it can be done there a lot easier and cheaper than anywhere else. The central office is already established by the network provider and offers empty space. Or at least it is easier to manage space and equipment in ones own estates. Therefore, the central office is usually considered as a location of an access remote unit with zero adaptation cost and unlimited capacity.

Despite the practical impact of this rule planners never insisted on its implementation during the first phase of the project. They argued that it is important to learn where copper wires concentrate and how many they are. This information directs the search for adequate locations.

1.4.9 Rule 9 (R9): ARU setup cost

Capacities of ARUs may vary according to Rule 8. Consequently, setup costs for ARUs will vary, too. Setup costs are composed of three main components:

- equipment (cabinet, DSLAM[28], air-conditioning, ...)
- installation (splicing, digging, construction, power supply, ...)
- operating expenses (electricity, rental, one-off payments, support, ...).

Economy of scale applies to these cost factors. For example, two ARUs of the same type on the same spot will cost less than twice the price for one ARU of this type, since power supply has to be provided only once.

An important driver for setup costs is the question of indoor or outdoor location. Usually, indoor will be cheaper than outdoor, at least if equipment for a larger number of customers has to be installed. Street cabinets (outdoor) have limited space compared to basement rooms. A high concentration of copper lines is best handled in a large indoor location. Otherwise, they have to be served by several outdoor ARUs which have to be spread over a wider area, since they cannot be simply installed in one location. Otherwise, complaints from the neighborhood would be the consequence. The problem with indoor rooms is not so much the price, but the availability and accessability of adequate location and the approval of house owners and renters.

Equipment cost are relatively easy to estimate, because the base cost (cost per unit of equipment depending on the type of equipment) is known beforehand. The final equipment cost is then a function of the calculated demand which has to be satisfied by the ARU under consideration.

Installation costs are already harder to assess. Splicing depends again on the demand and is therefore straightforwardly calculated. Digging and power supply depends very much on the constitution of the location which is barely known in advance, i.e. before a site inspection.

The case is worst for the factors listed under operational costs. Site inspections are necessary to be able to decide whether an indoor location is accessible at all. Then, during negotiations availability and price of the location have to be asserted.

Summing up, realistic total setup costs cannot be provided for all potential ARU locations, since it is completely impractical and extremely expensive to perform site inspections and prenegotiations for all possible locations beforehand. Therefore, it was suggested to focus on the number of ARU locations. Very much like in Rule 8 planners argued that it is important to learn where copper wires concentrated and how many they are. This information will guide the search for adequate locations during the detailed planning phase.

[28] Digital subscriber line access multiplexer

1.4.10 Rule 10 (R10): Distance rule

The quality of a solution of the FTTC planning strategy is controlled by means of the distance between customers and supplying access remote units. In the context of FTTC quality stands for transmission rates — in contrast to network reliability, for example: High transmission rates - high quality network, low transmission rates - low network quality.

During the first four years of project SARU three different ways of measuring the distance between customers and the ARU they are assigned to were established.

At first, distances were measured by the length of the cables and wires which run between potential facility locations and customer nodes[29]. Because of damping which occurs along transmission paths the bandwidth of signals decreases and so do transmission rates. This is also the reason why broadband internet access is a synonym for fast internet access: transmission starts with broad bandwidths. Consequently, there exists a high negative correlation between the length of cable paths and transmission rates.

Cable lengths are a far from exact measure, but good enough to serve as an approximation for transmission rates. The approximation was accepted, simply because at that time no valid models for determining transmission rates solely based on network information like cable length, cable diameter and other similar database information were available.

This case changed later. By the beginning of 2008 the Telekom Austria group responsible for transmission technology made such models available. The models were implemented in the SARU planning tool in summer 2008. From then on distances between potential ARU locations and customers could be estimated more directly by a model for transmission rates which was based on parameters which are very important for damping like the diameter of the copper cables. Other parameters reflected the influence of concurrent transmission system like ADSL2+ which had to be respected at least as long as there were customers paying for them.

The model uses families of piecewise linear functions which reveals a nonlinear relation between the cable distance and the transmission rate. In SARU these piecewise linear functions are substituted by quadratic functions to make calculations and data handling a bit easier. The data measurements on which the piecewise linear functions are based show a perfect fit to the quadratic approximations.

Finally, in the beginning of 2009 damping itself became the measure of distance and quality. Damping, like cable length, is negatively correlated to transmission rates. As signals progress through the wire damping increases reducing speed of transmission. Damping depends on the length of the wire, its diameter, how often diameters of the wires change and on the quality of the switching connection where the wires change.

Why exactly damping was sometimes favored over the transmission rate model did not become clear completely. But, the model for transmission rates obeys some parameters which do not depend on the physical network. They rather reflect the usage of the network, for example the parameters which are concerned with concurrent systems or the maximal number of customers

[29] This is the reason, why the rule is called distance rule and not transmission rate rule instead.

operating simultaneously on one access remote unit and transmitting data through the same cable[30]. In this sense damping is a measure which reflects more the physical condition of a network and does not mix with the usage of a network which may change over time independent of the physical network.

In summary, the ultimate quality measure of the solution of a FTTC planning strategy is the factual transmission rate. There are three different ways of measuring this rate in SARU which are

- cable length
- the transmission rate model or in short transmission rate[31] and
- damping.

The distance rule R10 is a threshold value which distinguishes good network or transmission quality from bad quality. Of course, the value of the threshold has to be chosen depending on the distance measure in use. Typical values and therefore default values are

- 600 m with cable length (up to 600 m is good quality)
- 20 Mbps transmission rate (at least 20 Mbps is good) and
- 7.5 dB for damping (up to 7.5 dB is good) .

Damping was never implemented in the SARU planning tool during Phase 1 simply because of lack of request. It would not be hard to realize, since the transmission rate model contains a model for calculating damping.

1.4.11 Rule 11 (R11): Individual distance rules

The distance rule R10 as defined in Section 1.4.10 applies globally. First of all, it determines which measure is used to control transmission quality. This measure is used throughout the entire optimization process everywhere in the local loop under consideration. Secondly, it establishes a global default value for the threshold which distinguishes between good and bad transmission quality for every customer included in the problem.

Despite of Rule 10 the second part of the distance rule may be individually altered, i.e. for individual customers (see Section 1.4.2 for the discussion of the term customer) the global threshold value of Rule 10 can be changed. In other words, Rule 11 says, the quality threshold does not have to be the same for all customers.

One motivation for Rule 11 is rather obvious. Although, a global threshold of, for example, 600 meters seems desirable, there may be certain customers like police stations, hospitals,

[30] Cable and not wire. The relation between customers and wires is 1:1. The relation between customers and cables is n:1.

[31] Which should not be confused with the factual transmission rate. But, the factual transmission rate will be known only after the network is build and users are online.

research facilities or similar institutions within a local loop which have a broader desire for bandwidth than is granted to the average user. For such customers threshold values can be defined individually by the planner, for example.

Furthermore, Rule 11 may also serve as a work around for Rule 5, full coverage of customer demands (Section 1.4.5). In the outskirts of a local loop customer density decreases dramatically depending on the topography of the area. However, all rules — especially Rule 5 and Rule 10, the distance rule — have to be obeyed. This may lead to underutilized access remote units. To avoid or at least dampen this problem affected customers may be assigned a weaker distance rule than others leading to a reduced number of access remote units with a higher utilization factor.

1.4.12 Rule 12 (R12): CO circle

In Section 1.4.8 it was already discussed in detail that the central office (CO) is a special location for situating an ARU. It can be done with comparatively low costs and capacities can been viewed as unlimited.

The set of customers who can be assigned to and supplied by an ARU situated at the CO without violating the distance rule R8 is called the CO circle[32].

Two things should be noted. Firstly, different distance measures will produce different CO circles. The CO circle depends on the distance measure. Secondly, the definition states a "can"and not a "must"condition. I.e. the customers from the CO circle may be assigned to the CO, but they don't have to be assigned to it. Rule 12, the CO circle rule, deals with the issue of enforcing the CO circle. In the case where the rule is activated, the CO circle has to be enforced in the solution of the FTTC problem, i.e. the customers from the CO circle have to be assigned to the CO.

In the case where Rule 12 is not activated this definition considers the maximal number of customers who can be assigned to the CO while still laying within the given distance rule. It might happen that in the final solution of the FTTC problem not all customers of the CO circle are actually assigned to the CO. It may be more beneficial to place an ARU "inside"the CO circle — i.e. on a copper cable which connects a customer from the CO circle to the CO. In this case the concerned customer must not be assigned to the CO because of Rule 7, disjoint supply areas of ARUs.

For some technicians it seemed to be the easiest approach to begin with the CO circle when searching for a solution of the FTTC problem. They determine all the customers in the vicinity of the CO who do not violate the distance rule R8 with respect to the provisioning by an ARU situated at the CO. So, for example, either all customers who are within a cable distance of at most 600 meters to the CO, or all customers who may be granted at least 20 Mbps transmission rate, or all customers whose wire does not produce more than 7.5 dB damping are assigned to the CO. Then they are removed from the problem, and the remaining customers have to be assigned to real access remote units in some other way. But, this strategy does not necessarily lead to the cheapest solution.

[32]The original definition was of course in German. It is called HV Kreis (Hauptverteiler Kreis).

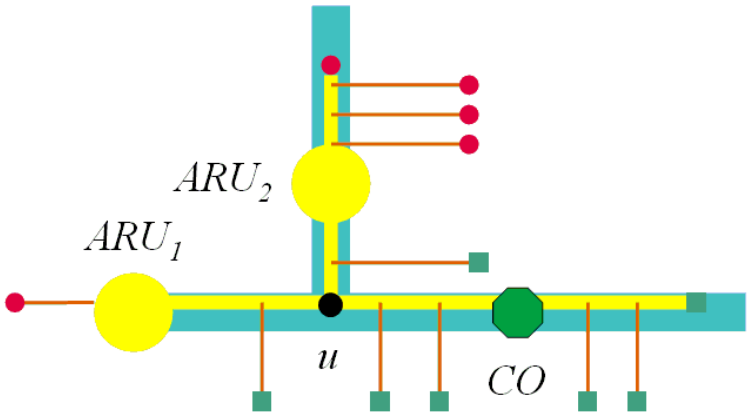


Figure 1.18 CO Circle and its disadvantages

In Figure 1.18 the squares symbolize members of the CO circle. The small circles are definitely not part of the CO circle and have to be assigned to ARU_1 and ARU_2. In this picture the round customers may also be assigned to a facility which can be situated at node u. That way one ARU could be spared and set up and equipment costs would be saved.

There are two members of the CO circle whose supply path contains node u. So, if the two ARUs are replaced by one ARU at node u, these two customers will have to be reassigned to this ARU. This is due to Rule 7 (Section 1.4.7), disjoint supply areas of ARUs. If these two customers are not reassigned and the new ARU is erected anyway, then the new ARU and the CO share a supply area which is forbidden by Rule 7.

If, on the other hand, the CO circle is enforced, i.e. it is mandatory to assign the CO circle to the CO, then this will eventually lead to solutions which install more ARUs than necessary and consequently to more expensive solutions to provide the same quality standard.

1.4.13 Rule 13 (R13): TNAK length

As described in Section 1.4.2 the documented part of the wire which connects a customer to the central office ends at a switching node of the copper network called KA (Kabelausmündung, cable outlet point). This point is also called the logical customer in this collection of specifications. A final copper cable leads from there to the actual and physical customer. This cable is called in German TeilNehmer AnschlussKabel — in short TNAK (subscriber's access cable). Since this last part of the access network is not documented, there is no information about the length and the diameter of individual TNAKs available.

Both, length and diameter of the TNAK, are relevant parameters for the factual transmission rate which can be achieved for a physical customer. Since it was clear that documentation would not be provided, it was decided to resolve this lack of data by estimating both values globally for all logical customers of a local loop. As a result, this procedure effects all physical customers in the same way.

Additionally and similarly to the distance rule (compare 1.4.11), these global estimates for the TNAK are allowed to be altered by the planner individually for logical customers in case more

detailed information is available. Here, it has to be observed that all physical customers who are related to one logical customer (i.e. KA or cable outlet point) whose parameter is altered are effected by the alteration in the same way. There is no more detailed differentiation beyond the level of the logical customer.

The values of the TNAK parameters have to be respected and applied during the solution of the FTTC problem. Cable lengths or estimated transmission rates have to be adapted accordingly.

1.4.14 Rule 14 (R14): One local loop at a time

The optimization process of the FTTC strategy is focused on one single local loop (or access area) at a time. This implies especially that only one central office and all customers connected to it together with the underlying copper network are considered in the problem formulation. The possibility to simultaneously solve the FTTC problem for several local loops and their central offices is thereby explicitly excluded.

The issue of simultaneous solution was discussed. A potential for cost saving through simultaneous planning of adjacent local loops was identified especially because of the shared usage of pipe infrastructure. But, the counterargument was that the area under consideration would become too large for one planner to handle during the detailed planning processes. Actually, there was a tendency of some planning groups to even break one local loop into a few subsegments and work on them separately.

1.5 Theory: FTTC and facility location

The cost optimal planning of an access network according to the FTTC strategy consists of two main problems: a facility location problem and a network design problem.

ARUs are the facilities. The number of necessary ARUs is not known in advance, nor is their location. The task is to determine as many ARUs and their locations as necessary and as cheaply as possible, in order to assign all customers of the local loop to exactly one ARU and observe all other rules which are listed in Section 1.4. This is the facility location part of the problem.

Additionally, installed facilities have to be connected to the central office via an optical fiber network. The facilities, i.e. ARUs, become the customers which have to be connected to a source, the CO. Usually, the network has to be determined based on a given graph which defines a potential network comprising of existing infrastructure like pipes or cable ducts and trails along which the construction of new ducts is possible (Compare Section 1.3.1). Certainly, a list of specifications and rules similar to those found in Section 1.4 has to be observed in this case, too. This list is not stated here, since the topic of network design exceeds the scope of the thesis.

Both topics — facility location and network design — are well known and studied in the literature. The focus of this thesis lies on facility location. However, since the realization of the FTTC planning strategy involves network design, a brief overview of strategies and methods for the network design problem as it was studied during the cooperation between University of Vienna and Telekom Austria is given in the following section.

1.5.1 Network design: network loading and ConFL

One attempt to handle the FTTC planning problem is to first solve a facility location problem and then apply a network design algorithm like a Steiner tree algorithm (see for example [27] for an extensive collection and discussion of the solution of the Steiner tree problem in networks) to the previously found ARUs which connects them to the source, the central office. This iterative method was adopted for the first part of the cooperation between University of Vienna and Telekom Austria (Section 1.2).

The iterative approach, however, may lead to suboptimal solutions. For example, it may be more beneficial to open up one more facility, if this allows the ARUs to be placed closer to already exiting infrastructure like pipes which will save money, since digging is very expensive. Or – the optimal solution of the facility location problem may not be unique. A different set of ARUs may be cheaper to connect to the central office.

Another way of solving the FTTC planning problem is the challenge to tackle the two problems not iteratively but in one step. The study, solution and implementation of this approach was planned for the second part of the cooperation of University of Vienna and Telekom Austria.

In any case, Ivana Ljubić and her group worked over three years on the topic of network design, implemented their algorithms and tested them in the Telekom Austria environment. Some of their findings have already been published.

The iterative method to solve the FTTC planning problem is addressed in "Exact approaches to the single-source network loading problem"[41] by I. Ljubić, P. Putz and J.J. Salazar-González. The network loading approach deals with the design of the optical fiber network connecting the facilities to the CO. The facilities (ARUs) are given, determined, for example, by the algorithms which are presented in Section 1.6 and 2.5. The major challenge of the approach lies in capacity constraints which are imposed on the edges of the potential network.

As the title indicates the authors seek exact instead of heuristic ways to solve the problem. Moreover, the single-source network loading problem is a generalized formulation of the Steiner tree problem, the classical network design problem, which reflects the specific requirements of Telekom Austria for their network design.

The data model of the network graph (see Figure 1.11) allows different types of resources on every edge of the network. These resources are dark fiber, pipe infrastructure and — where there is no infrastructure or where infrastructure does not provide enough free resources — excavation together with construction of new trenches. These layers provide free resources which are equivalent to optical fiber lines in different quantities for different prices. One important problem of these prices is that they lack economy of scale. Dark fiber, where it is available, is basically for free. Consequently, the price per unit (fiber) is nearly zero. Using empty pipes is a bit more expensive, since equipment has to be paid for and work to be done. Even if digging and building new ducts guarantees the greatest possible increase of free glass fiber, it is the most expensive layer in absolute value as well as in price per unit. The reason for the lack of economy of scale is simply that existing infrastructure is already paid for, an investment which is not recognized during the optimization. Therefore, existing infrastructure is not only cheaper in absolute numbers, but also in price per unit. This is probably the specific challenge of the Telekom Austria network design problem.

The method of solving the facility location and network design problem simultaneously is called connected facility location problem (ConFL). In the paper "A GRASP Algorithm for the Connected Facility Location Problem"[55] the authors A. Tomazic and I. Ljubić applied a heuristic method called a greedy randomized adaptive search procedure to solve the connected facility location problem (ConFL). They compensated for the increased difficulty of solving facility location and network design problem simultaneously with simplifying the data model of the network graph. The cost step-function is omitted and replaced by one value which may of course vary over edges. Furthermore, they reported a relation between ConFL and the minimum Steiner arborescence problem and gave computational results based on 135 randomly generated graphs.

The attempts to solve the connected facility location problem to optimality are strengthened in [21], "MIP Models for connected facility location: a theoretical and computational study". The authors S. Gollowitzer and I. Ljubić analyzed mixed integer programming models to solve the connected facility location problem. The edge cost model, however, stays the same as in the previously cited paper. They also gave a literature review which mainly comprises of approximation algorithms.

Additionally to the network design problem multi period planning is considered in [2].

1.5.2 Facility location (FL)

The first phase of SARU saw the collection of all requirements and technical specifications and the implementation of a prototype of the planning software. The software contains an optimization module which uses a facility location approach to situate access remote units. The network design problem is omitted. There were three reasons which led to this modus operandi for the optimization part.

1. According to the cooperation plan between University of Vienna and Telekom Austria the connected facility location problem was first to be tackled by the iterative approach: solving a facility location problem, then applying a network design algorithm to the determined facilities. It was the part of Telekom Austria to implement a solution for the facility location problem.
2. The pressure to deliver practical results increased constantly.
3. Two problems with the database which was necessary to work on the network design problem in practice were not resolved before the end of 2007 and 2008 respectively (see Section 1.3.1 and 1.3.3 for details).

Therefore, initially all the concentration was put on the implementation of a solution for the facility location problem.

Problem definition

The paper "Telecommunication and location"[22] by E. Gourdin[33], M. Labbé and H. Yaman[34] is a thorough review of the research done for models in telecommunication network design up to 2001 where a location problem is involved. The authors acknowledged a telephone network as a multi layer network and identified three main layers:

> A generic telecommunication network consists of access networks which connect the terminals (user nodes) to concentrators (switches or multiplexers) and a backbone network which interconnects these concentrators or connects them to a central root[35].

The authors stated a list of six questions, which have to be answered to design such a network. They concluded the introduction by suggesting a general strategy to answer these questions:

> It is hard to develop tractable models that can come up with answers to all these questions above simultaneously. So, usually the design is done in an iterative manner as follows:
>
> - Decide about the number of locations of concentrators and the assignment of the terminals to these concentrators
> - Design the access network

[33] France Telecom
[34] At that time both Université Libre de Bruxelles.
[35] Page 3

- Design the backbone network.

> The idea of this iterative method is that once the location of the concentrators and the assignment of the terminals to these concentrators are done, the design of the access networks and the design of the backbone network become independent and can be handled separately[36].

The topologies of the graphs on which these networks are based (stars, rings, trees) suggest that the existence of a physical network is a requirement. The design problem is one to identify the roles of the individual nodes and route the interconnection between all nodes in such a way that restrictions on capacities are met.

In what follows the authors focused on the first part of the iteration, i.e. the facility or concentrator location problem. They classified the analyzed papers, problems and suggested solutions into four groups:

- Uncapacitated models
- Capacitated models
- Capacitated models with multitype concentrators and
- Dynamic models.

The question of capacity is mainly associated with facilities. Capacity is understood as the answer to demands which are associated with customers[37]. Capacities of facilities are in general limited and have to be respected when customers are assigned to facilities. If it is possible to chose concentrators from a set of different types of facilities, then different quantities of demand may be served by concentrators which vary with respect to capacities and prices. If networks have to be considered as they change and grow over time, models are applied which situate facilities and expand networks in a dynamic manner over time. The associated problems are also called multiperiod facility location problems.

Definition UFLP

The definition of the uncapacitated facility location problem as it is given in [22] is repeated here:

> The UFLP can be stated as follows: given a set of terminals N and the set of possible locations for the concentrators M, determine the number and the location of concentrators and assign the terminals to these concentrators. The goal is to minimize the sum of the cost of installing concentrators and the cost of serving terminals via the installed concentrators.

In [22] customers are called terminals. A formal definition is presented by an integer linear programming formulation.

[36] Page 6.
[37] Which are called terminals in [22]

Definition 1 (UFLP) *Let N be the set of customers, and M be the set of possible locations for facilities. For $c \in N$ and $f \in M$ AC_{cf} denotes the cost of assigning customer c to a facility at node f (assignment cost) and OC_f the cost to open a facility at location f (opening cost). The uncapacitated facility location problem is to solve*

$$\min \sum_{c \in N} \sum_{f \in M} AC_{cf} * x_{cf} + \sum_{f \in M} OC_f * y_f$$

subject to

$$\sum_{f \in M} x_{cf} \quad = 1 \quad \forall c \in N \tag{1.1}$$

$$x_{cf} \quad \leq y_f \quad \forall c \in N \text{ and } \forall f \in M \tag{1.2}$$

$$x_{cf} \quad \geq 0 \quad \forall c \in N \quad \forall f \in M \tag{1.3}$$

$$y_f \quad \in \{0,1\} \quad \forall f \in M. \tag{1.4}$$

The assignment variables x_{cf} state that customer x is assigned to facility f, if the value of the variable is positive. The opening variables y_f decode by a value of 1 that a facility has been opened at location f. Constraint (1.1) together with (1.3) demands that every customer is assigned to some facility. Constraint (1.2) assures that if a customer is assigned to a facility that this facility has to be opened, or if a facility stays closed, no customers are assigned to it.

It can be noted that the term customer demand is not mentioned, although customers certainly have demands. However, if capacities of facilities are unlimited, then there is no need to differentiate demands for different customers. It has no impact on the solution.

In Chapter 2 of [22] and also in Section 2.1.1 of [60] — "Concentrator Location in Telecommunications Networks"a book by H. Yaman — a long list of papers is cited which deal with this problem and variants of it. Some of them also consider the associated network design problem based on special assumptions of the underlaying network topology. The solution approaches mainly use different linear programming formulations, introduce special families of cuts and apply Lagrangian relaxation.

In 1978 Erlenkotter published a dual-based procedure in [17] to solve the UFLP.

In [18] Filho and Galvão suggested a tabu search algorithm to solve the UFLP. The heuristic uses two moves, an add - move to open new facilities and a close - move to close facilities. After every move customers have to be reassigned to the now available set of facilities. The moves are compared by the amount of savings they gain. The best one is chosen. To avoid cycling the reverse move of a successful move is set on a tabu list for some time. Lower bounds are determined by Lagrangian relaxation: Equation 1.1 is relaxed and moved to the objective function.

In [23] Guha and Khuller combined a greedy heuristic with a constant factor approximation algorithm given by Shmoys, Tardos and Aardal in [49] to improve the latter. The 4-approximation of Shmoys et al. is improved to a 2.408 approximation. However, the approximation algorithm of [49] is amongst others also applied to the capacitated facility location problem (CFLP) and the capacitated concentrator location problem (CCLP) which are discussed in the following subsections.

The results presented by Tamir in [53] allow a cross reference to Chapter 2 of this thesis where the k-median problem is studied. The k-median problem can be understood as an uncapacitated facility location problem with an upper bound k on the number of facilities which are allowed to be opened. Usually, the k-median problem is solved without considering opening costs for facilities. Tamir, however, gave a solution algorithm which accounts for opening costs. Now, if k is chosen equal to or larger than the number of potential locations for facilities, then Tamir's algorithm can be used to solve the UFLP in trees. It does so in maximal $\mathcal{O}(n^2)$ time. A similar result for trees without referring to the k-median problem can be found in [13].

R. Shah and M. Farach-Colton were able to improve this result on trees to $\mathcal{O}(n\ log\ n)$ in [47]. They applied a technique which is called undiscretized dynamic programming. The previously mentioned algorithms on trees compute certain cost functions explicitly between every node and all their "neighboring"nodes. The resulting values depend on the distance between node and neighbor. Since the space of "neighbors"is discrete, the set of values of these cost functions is discrete, too. Or in other words, the cost functions are discrete with respect to the distance to neighboring nodes. The trick of undiscretized dynamic programming is to construct all necessary cost functions in such a way that they depend on a continuous distance variable. A cost function can be written by one single expression for all "neighbors". (For an exemplary illustration see also Chapter 2 Section 2.3.12 depth based algorithms)

Definition CFLP

The capacitated facility location problems associate a demand — a positive number — with each customer. It is denoted by d_c. The capacities of facilities are limited, i.e. only a limited number of customers may be assigned to a facility and the sum of their demands must not exceed a capacity limit Q_f which may be different for each facility f. The linear programm reads:

Definition 2 (CFLP) *Let N be the set of customers, and M be the set of possible locations for facilities. For $c \in N$ and $f \in M$ AC_{cf} denotes the cost of assigning customer c to a facility at node f (assignment cost) and OC_f the cost to open a facility at location f (opening cost). The demand of a customer is denoted by d_c and the maximal capacity of a facility by Q_f. The capacitated facility location problem is to find a solution for*

$$min \sum_{c \in N} \sum_{f \in M} AC_{cf} * x_{cf} + \sum_{f \in M} OC_f * y_f$$

subject to

constraints (1.1), (1.3), (1.4) *and*

$$\sum_{c \in N} d_c * x_{cf} \quad \leq Q_f * y_f \quad \forall f \in M. \tag{1.5}$$

Three out of four constraints are unaltered. Constraint (1.2) is replaced by (1.5) which takes care of two tasks now. On the one hand, it keeps solutions consistent as Constraint (1.2) does: customers cannot be assigned to closed facilities. On the other hand, it restricts the demands which are assigned to a facility by its capacity. Some approaches to the solution leave Condition (1.2) in the set of side constraints.

The formulation of CFLP is in conflict with the specifications for the telecommunications requirements, especially with Rule 4, undivided customer demands (Section 1.4.4). Constraint (1.3) implies that a customer may be assigned to more than one facility which especially for the CFLP formulation means that fractional parts of the demand of a customer may be supplied by different concentrators.

This conflict can be ignored for uncapacitated formulations, since a given solution of UFLP can always be remodeled into a solution where a customer is assigned to a single facility only. There is enough capacity available at a facility to take care of all fractions of one customer.

This cannot be guaranteed anymore in the capacitated case. But, it is imperative for the given telecommunications problem to assign customers to a single source only.

Sridharan gave an overview of solution methods of the CFLP which he called the capacitated plant location problem in [51]. There the description of two greedy heuristics can be found which are similar to the add and close moves of the tabu search to solve the UFLP presented previously. The add heuristic starts with all facilities closed and opens one facility after another always reassigning customers and choosing to open the facility which gives the greatest amount of saving. Cycling is avoided by not closing facilities once they are opened. The procedure stops as soon as no savings are gained anymore. The drop heuristic works similarly with all facilities opened and closing down one by one.

More recently Levi et. al presented a LP-based approximation algorithm for CFLP in [36].

Definition CFLPS or CCLP

The previously identified conflict with Rule 4 is resolved by the class of capacitated facility location problems with single sourcing (CFLPS) which are also called capacitated concentrator location problems (CCLP).

Definition 3 (CFLPS, CCLP) *Let N be the set of customers, and M be the set of possible locations for facilities. For $c \in N$ and $f \in M$ AC_{cf} denotes the cost of assigning customer c to a facility at node f (assignment cost) and OC_f the cost to open a facility at location f*

(opening cost). The demand of a customer is denoted by $d_c \geq 0$ *and the maximal capacity of a facility by* $Q_f \geq 0$*. The capacitated facility location problem with single sourcing is to find a solution for*

$$min \sum_{c \in N} \sum_{f \in M} AC_{cf} * x_{cf} + \sum_{f \in M} OC_f * y_f$$

subject to

$$constraints\ (1.1),\ (1.4),\ (1.5)\ and$$

$$x_{cf} \quad \in \{0,1\} \quad \forall c \in N, \forall f \in M. \tag{1.6}$$

Now, the assignment variables can take only one out of two values, 0 for not being assigned to a facility at f and 1 for being assigned to it.

Lower bounds for the optimal value can be calculated by means of Lagrangian relaxation. There are two types of relaxations considered in the literature. In [42] and [32] the Capacity Constraint (1.5) is relaxed and moved to the object function which actually leads to an uncapacitated formulation of the problem. An alternative way of relaxing the problem is by dualizing the Assignment Constraint (1.1) which for example is studied in [50].

To solve CCLP in best case to optimality Holmberg et al. suggested a complex procedure in [26]. By means of Lagrangian relaxations lower bounds are calculated which are improved by subgradient optimization. Then a heuristic based on the repeated matching algorithm delivers feasible solutions and hence upper bounds. For any pair of facilities (open or not) the best possible local improvement involving only the two facilities and all the customers assigned to them is calculated. The results are stored in a matrix $D = (d_{ij})$, where d_{ij} is the partial assignment and opening cost involving the facility pair (i, j) as calculated before. Then the matching algorithm is applied. This procedure is repeated until no improvement is found anymore. By combing these two approaches under a branch and bound setting they try to determine the optimal solution.

Rönnqvist et al. adapted the CCLP in [45] in such a way that they can apply the repeated matching algorithm to solve the problem heuristically. They defined six different types of matches which result from all possible pairs of the three entities unsupplied customers, closed facilities and open facilities.

Definition CCLP with multitype concentrators

Finally, one last concept of those collected in [22] shall be introduced to deal with facility location for the FTTC planning strategy: several types of concentrators.

Definition 4 (CCLP with multitype concentratros) *Additionally, to the customer set* N*, the facility set* M*, customer demands* $d_c \geq 0$ *and assignment costs* AC_{cf} *a set of different facility types* K *is defined. Each pair of potential facility location* f *and concentrator type* k *of* K *is associated with a maximal capacity* $Q_{fk} \geq 0$ *and an individual opening cost*

OC_{fk}. The capacitated concentrator location problem with multitype concentrators seeks the answer to

$$min \sum_{c \in N} \sum_{f \in M} AC_{cf} * x_{cf} + \sum_{f \in M} \sum_{k \in K} OC_{fk} * y_{fk}$$

subject to

$$\sum_{f \in M} x_{cf} = 1 \qquad \forall c \in N \tag{1.7}$$

$$\sum_{k \in k} y_{fk} \leq 1 \qquad \forall f \in M \tag{1.8}$$

$$\sum_{c \in N} d_c * x_{cf} \leq \sum_{k \in K} Q_{fk} * y_{fk} \qquad \forall f \in M \tag{1.9}$$

$$x_{cf} \in \{0,1\} \qquad \forall c \in N, \forall f \in M \tag{1.10}$$

$$y_{fk} \in \{0,1\} \qquad \forall f \in M, \forall k \in K \tag{1.11}$$

Opening variables depend now on the type of facility, too. They can take one out of two values, and they have to add up to at most 1 per location, i.e. per location only one type and only one facility are allowed to be situated.

There is basically no restriction to the set of concentrator types. Opening cost and capacity limits both not only depend on the facility type, but may vary with locations. So, concentrators may be designed individually for every location.

A solution algorithm was proposed by Lee in [35].

The FTTC planning strategy and facility location problems

Based on this collection of definitions the facility location part of the FTTC planning strategy for telecommunication networks as it was roughly defined in Section 1.1.5 can be identified as a capacitated concentrator location problem with multitype concentrators. Of course, it should also conform with the list of specifications and rules collected in Section 1.4.

The concept of CCLP with multitype concentrators does not explicitly ask for the alteration of the underlaying network. In fact, many of the cited applications in [22] require an existing network. However, the linear program in Definition 4 provides assignment variables for any pair of customer and facility location. This either contradicts Rule 6 (Admissible locations for ARUs, 1.4.6) or, in case this discrepancy is resolved through establishing new copper connections, it contradicts Rule 1 (Unaltered copper net, 1.4.1).

There are several ways to map the admissible assignment structure (R6, R10, distance rule 1.4.10, R11, individual distance rules 1.4.11 and R13, TNAK length 1.4.13) of the given FTTC problem into the linear program of Definition 4. Since in practice there are no assignment costs (AC_{cf}) associated with the problem, these cost factors can be (ab-)used to model admissible

assignments. Whenever a customer must not be assigned to a certain location — may it be because of the network structure (R6) or because of distance rules (R10, R11, R13) — the corresponding assignment cost is set to such a high value that the optimal solution is guaranteed to abandon such an assignment.

A consequence of this attempt is that the space of feasible solutions of the linear program of Definition 4 is larger than the set of admissible solutions which are in accordance with the rules of Section 1.4. So, either the linear program has to be solved to optimality or, as it is common practice to stop the program before reaching optimality, and to content oneself with a suboptimal solution in shorter timer, a suboptimal solution has to be checked whether it confirms with the cited rules.

Another way to deal with this conflict is to force the affected assignment variables explicitly to be zero by introducing additional side constraints, or by changing the design of the linear programm of Definition 4 such that these variables are simply omitted. In any case, it is clear that some work has to be done to bring the binary linear program of the CCLP with multitype concentrators in accordance with the stated FTTC planning strategy.

All rules concerning customer demands (R2 to R5) are covered by the CCLP concept. Especially R4, undivided customers demands (1.4.4), is met by the capacitated concentrator location problem in contrast to the capacitated facility location problem as was already discussed. Rule 5, full coverage of customer demands 1.4.5, is fulfilled by Assignment Constraint (1.7).

The issue of restricted capacities (R8, 1.4.8) is reflected by the capacitated part of the CCLP concept. Questions of facility setup cost which vary depending on locations and different types of access remote units as they are addressed in Rule 9 (ARU setup cost, 1.4.9) are conceptualized in Definition 4 by multiple concentrator types with capacity restrictions and setup costs depending on type of concentrator and location.

The facility location concepts presented in [22] and [60] are not restricted to certain areas or just a single local loop. On the contrary, together with the network design problem the planning of supra-regional telecommunication networks is considered which connects several access areas by backbone networks. Rule 14, one local loop at a time 1.4.14, is a self-imposed restriction which is certainly in the scope of the CCLP formulation.

Rule 12, CO circle 1.4.12, can be adhered to by preprocessing. All customers inside the CO circle may be determined in advance, and together with all information about their supply paths they can be excluded from the database on which the linear program is then defined. It is important to also remove this information from the supply paths of all other customers. These nodes belong to the access area of the ARU situated at the central office. Because of Rule 7, disjoint supply areas of ARUs 1.4.7, they must not be part of the access area or even a location for other ARUs.

The last point to check is whether the CCLP concept is in compliance with rule number 7, disjoint supply areas of ARUs 1.4.7. It turns out that Definition 4 cannot guarantee that different ARUs use disjoint access or supply areas of the copper network. On the one hand, the ILP formulation of CCLP or CFLP lacks the concept of a network, and consequently individual

access areas cannot be modeled. On the other hand, the capacity constraint may cause access areas of different ARUs to overlap, since capacity restriction can force a customer who actually lies in the access area of a certain ARU to be assigned to another ARU with free capacities.

The uncapacitated formulations are not so affected by this problem. Even if it occurs, solutions are easily repaired, since enough capacities are available. Customers may always be assigned to the "nearest" access remote unit. This gives the cue for a potential adaptation of the capacitated problems. In the beginning of this section it was already stated that assignment costs have no realistic correspondent and can therefore be omitted. In this context, however, they can be used to express closeness between potential facility locations and customers which should lead to solutions where customers are assigned to the nearest available facility.

Again, the problem of differing sets of feasible solutions as it was described in the beginning of this section for a similar treatment of violated specifications occurs. Moreover, even the optimal solution of the ILP with assignment costs which correlate to distances between locations and customers does not guarantee compliance with R7. Figure 1.19 shows a counterexample.

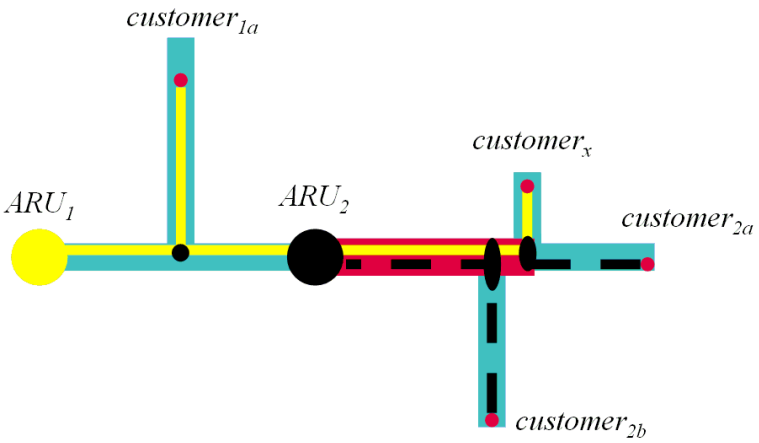


Figure 1.19 Example for ILP formulation of CCLP in conflict with R7

There is just one ARU type. The capacity of ARUs equals 2. Opening costs are the same everywhere. There are four customers in the picture. The demand of all customers is 1. Customer $2a$ and especially Customer $2b$ are too far away from ARU_1 to be assigned to it. The furthest location to which they might be assigned is the location of ARU_2. Customer $1a$ is served by ARU_1. Customer x may be assigned to ARU_2 or ARU_1.

At least two access remote units are needed to provide full service because of the stated demands and the capacity constraints. Due to the capacity restriction and the network situation the CCLP will give the depicted solution or a similar one. Especially, Customer x will be assigned to ARU_1. There is basically no other location for ARU_2 than the one where it sits. The solution is in full compliance with side constraints (1.7) to (1.11) of the ILP formulation of the CCLP, but it is in contradiction to R7. To resolve the conflict ARU_2 has to be closed and two ARUs have to be opened one at each location of customers $2a$ and $2b$. Customer x is assigned to ARU_1. Clearly, this solutions is more expensive.

One way to ensure that R7 is fulfilled is to introduce additional side constraints which reflect the structure of the copper network. For every quartet of two costumers and two potential facility locations it has to be checked whether an eventual assignment would be in conflict with R7. In case no conflict arises, no additional side constraint is needed. Otherwise, these clashing assignments have to be forbidden by adequate constraints.

The customers are denoted by c_1 and c_2 and the two conflicting facility locations by f_1 and f_2, then the set of constraints of the ILP in Definition 4 has to be extended by

$$x_{c_1 f_1} + x_{c_2 f_2} \leq 1. \tag{1.12}$$

One of the two assignment variables has to be 0. This only works for concentrator location problems, but not for facility location problems. There the assignment variables are real valued and the side constraint (1.12) does not force one of them to be zero. But, the case of a fractional assignment of a customer's demand to more than one facility is a violation of R7 in itself.

Another resolution of the conflicts with R7 may be achieved after the ILP has delivered a solution by fixing overlapping supply areas during a postprocessing step. Certainly, an eventual optimality of the solution may be lost that way.

1.5.3 Structural planning process

The theoretical analysis of the previous section arrives at the conclusion that the FTTC planning strategy together with the specifications from Section 1.4 can be realized by an adapted version of the capacitated concentrator location problem with multitype concentrators. However, from a practical perspective there exist some arguments which suggest choosing a different approach — at least to begin with.

In Section 1.4.9 (R9, ARU setup cost) the different cost factors for ARU constructions were presented. It was argued that site inspections and prenegotiations for all potential sites before the structural planning process would be far too expensive. But, such site inspections are necessary to determine exactly where which types of ARUs may be situated (e.g. indoor versus outdoor) and at how much cost. That was the motivation why planners — as mentioned at the end of Section 1.4.8 (R8, ARU capacities) — declared that they would prefer an uncapacitated approach, since this would guide them in the search for spots of high copper concentration.

The structural planning process could roughly take the following steps based on the concepts just presented.

1. Solve an uncapacitated facility location problem as stated in Definition 1 with assignment cost AC_{cf} set to zero and the same opening cost OC_f for all locations. I.e. opening costs can be set to 1 and the number of facilities is minimized.

2. A planner works through the solution to identify locations for ARUs which need closer inspection.

3. Either the planner has already got information about the locations in question, because he knows the local loop very well, or the planner gains additional information by an inspection of the sites.

4. The database for the optimization tool is updated based on the collected information. The update may comprise of:
   (a) Allow one or more types of capacitated facilities for a certain location, i.e. especially, encode where indoor ARUs are possible and where they are certainly forbidden.
   (b) Forbid locations completely where a situation of an ARU would be impossible.
   (c) Enforce a facility and the assignment of certain customers to it where such a step seems to be very beneficial.
   (d) Capacity limits and costs have to be updated not only for the inspected locations, but for all other locations, too.
5. Next, the capacitated concentrator location problem with multitype concentrators is applied.
6. The planner may enter an iteration process by starting at Step 2 again and collecting information, updating the database and optimizing the FTTC problem until a satisfactory solution is found.

1.6 Application: Situating ARUs by dynamic programming

The development of optimization algorithms as they are described in Section 1.5.3 for the structural planning process was started in SARU Phase 1, but not finished. Solutions for the capacitated and multitype problems were never developed nor implemented in Phase 1. In the course of the project a different problem assumed a higher importance. This problem can be paraphrased by the term underutilization which means there are facilities in a solution whose degree of capacity utilization is considered as being too low. Underutilization and its solution is the topic of Chapter 2.

If underutilization is the problem of not enough customers being assigned to a facility, then overutilization is the problem that there are too many such assignments. Overutilization is addressed by the capacitated concentrator location problem with multitype concentrators. Underutilization prevailed over overutilization during the project. Therefore, CCLP with multitype concentrators was not realized in SARU Phase 1[38].

The base of optimization in SARU Phase 1 became a dynamic program dealing with the uncapacitated facility location problem as it is required in Step 1 of the structural planning process which is described in Section 1.5.3.

1.6.1 Basic idea of the dynamic program

Telekom Austria planner Andreas Kriesel was aware of the basic idea of the algorithm which became the base optimization during SARU Phase 1 when the specification process started [34]. To get a quick and easy understanding of the algorithm it is best to imagine the copper network as a tree where the supply paths of the customers are identical with the shortest paths between customers and CO. For the moment, the distance rule (R10, Section 1.4.10) is based on cable distances. I.e. edge and path lengths are given in cable lengths. The value of the threshold is the same for all customers. Then the algorithm basically runs like this:

1. Pick the customer c_{max} furthest away from the central office CO.
2. Determine the potential facility location $f_{c_{max}}$ of greatest distance to customer c_{max} which is in agreement with the distance rule.
3. Assign all customers contained in the subtree $T_{f_{c_{max}}}$ of the copper tree which is rooted at node $f_{c_{max}}$ to a facility opened at this node.
4. Remove subtree $T_{f_{c_{max}}}$ — all its nodes and edges — from the tree describing the copper network.
5. Apply steps 1 through 4 to the remaining tree until all customers are assigned to a facility.

This is a bottom-up strategy. Starting at the periphery of the local loop the algorithm works its way towards its center, the central office. It will always be the CO where the last facility is opened. Frequently, there were discussions with some technicians who were convinced that a

[38] It was, however, part of Phase 2.

top-down strategy — starting with the CO circle and then working ones way somehow to the periphery — would definitely achieve cheaper results which is wrong as already demonstrated in Figure 1.18.

The central idea of the algorithm is, since all customers of the local loop have to be assigned to a facility (R5, full coverage of customer demands 1.4.5) that a facility has to be situated somewhere along the supply path of every customer. So, a customer is picked and assigned to a facility at a location of the greatest distance (Step 1 and 2). There are two reasons why the customer himself is also chosen at the greatest distance to the CO.

First of all, this implies that customer c_{max} is also at the greatest distance to node $f_{c_{max}}$ in the subtree $T_{f_{c_{max}}}$. This is true because of the assumption that the copper network forms a tree. The path from the central office CO to node $f_{c_{max}}$ is contained in the supply paths of all the customers inside the tree $T_{f_{c_{max}}}$. The length of this path is the same for all affected customers. Therefore, the residual length of the supply paths of all customers is less or equal to the residual length of the chosen customer c_{max} which confirms with the distance rule which by assumption is the same for all customers. Consequently, the assignment of all these customers to a facility at $f_{c_{max}}$ in step 3 is in perfect compliance with the distance rule.

Secondly, since all customers of $T_{f_{c_{max}}}$ can be assigned to the root of the subtree, the entire tree can be removed from the problem in step 4 and the remains of the graph are a tree again. If alternatively a customer c_{alt} is chosen in Step 1 who is not on maximal distance to the CO, then it might happen that there are some customers contained in the tree $T_{f_{c_{alt}}}$ which cannot be assigned to a facility at its root. Such customers cannot be removed from the graph. Otherwise, they are not assigned to a facility (violation of R5), or they are assigned to a facility which is too far away (violation of R10).

The combination of these points guarantees the start of a beautiful recursion in step 5.

Moreover, this dynamic program ensures the compliance of the solution with all specifications of Section 1.4 including R7 (Disjoint supply areas of ARUs 1.4.7). R7 is observed, since the subtree which is removed from the graph cannot be used anymore by subsequent facilities and it is identical with the supply area of the newly installed ARU. Previously opened ARUs and their supply areas are already removed from the graph. So, they will not have an overlapping supply area with the newly situated facility.

1.6.2 Problems and adjustments to reality

The validity of the basic algorithm is founded on the assumption that the copper network can be described as a tree. In Section 1.3.2 there were two incidents reported where the copper network diverges from a tree graph (see Figure 1.13 and 1.14). Especially, the case depicted in Figure 1.13 may result in that the customer furthest away from the CO is not the customer furthest away from the designated facility location ($f_{c_{max}}$). The customer at the end of $wire_2$ may be furthest away from the CO. But however, $wire_2$ takes a detour via node u_5. This may be the reason why this customer is further away from the CO than the customer at the end of $wire_1$. Node u_4 — after the detour of $wire_2$ — is probably closer to the end - customer of $wire_2$ than the subscriber of $wire_1$. This causes conflicts during step 3 of the base algorithm.

Moreover, R11 (Individual distance rules 1.4.11) allows variable distance rules for different customers. So, there is no guarantee anymore that step 3 produces an admissible assignment of customers inside the tree rooted at $T_{f_{c_{max}}}$ to a facility. Even if the distance between all customers in $T_{f_{c_{max}}}$ and the root $f_{c_{max}}$ is bounded above by the distance between $f_{c_{max}}$ and c_{max}, there might be a customer inside of this tree with such a restrictive individual distance rule, such that he must not be assigned to the root.

Finally, according to R10 (Distance rule 1.4.10) distance rules may be alternatively defined based on transmission rates or damping. It is very likely that there are many customers of the same greatest distance to the CO with regard to the transmission rate. From a certain relatively low cable distance transmission rates are zero[39]. Therefore, it is advisable to always use cable distance in Step 1 independent of the selected distance rule. But, if in Step 1 a customer is chosen according to cable length and in Step 2 a potential facility is located based on transmission rates or damping, there is again no guarantee anymore that all the customers in the subtree are assignable to the root. Transmission rates and damping depend on more than cable length (see 1.4.10 for details).

Therefore, the base algorithm has to be expanded by a distance rule test. After identifying the new candidate facility all the customers inside the tree which is rooted at this location have to be checked whether their distance rule is in conflict with their actual distance to the facility. In case the check is positive and no conflicts arise the algorithm continues with step 3. In the negative case, if at least one conflict can be identified the procedure is repeated again beginning at Step 1 with one of the conflicting customers instead of the previously chosen customer c_{max}.

Another problem results from the network situation described in Figure 1.13. If node u_3 is chosen as the candidate location $f_{c_{max}}$ then the subscriber of $wire_2$ cannot be assigned to this location. Node u_3 does not lie on its supply path because of the detour of $wire_2$. The picture shows that the concept of a tree rooted at u_3 becomes invalid. There is a path ($wire_2$) leading into the graph "below"u_3 which does not pass through u_3.

This problem, however, is easily resolved. Instead of investigating those customers inside a certain subgraph "below"u_3, all customers whose supply path contains node u_3 have to be checked before and during step 3. After a positive result of the distance rule test these customers together with their supply paths have to be removed from the copper graph as described in step 4.

By switching from the tree representation of the copper network to a representation by a set of supply paths the underlying data structure is changed. Removing the affected supply paths from this set does not ensure that all edges and nodes of the supply area of the newly installed facility are removed from the underlying graph. As a consequence R7 (Disjoint supply areas of ARUs 1.4.7) may be violated. To prevent this conflict the edges and nodes of the new supply area have to be removed from all supply paths even from those customers who remain in the problem.

[39]Zero is the greatest possible distance between a customer and a "supplying"facility in terms of transmission rate.

However, should this procedure be successful and find and remove edges and nodes from supply paths of customers which are still part of the problem and not assigned to a facility yet, then the resulting graph will not be connected anymore. Again, Figure 1.13 gives an example. If a facility is opened at node u_3, then one consequence is that node u_4 is removed from all supply paths. The subscriber of $wire_2$ is certainly not assigned to the facility at u_3. But, its supply path contains node u_4 which was removed. Consequently, the supply path is intercepted and the resulting graph must be disconnected. This situation has to be avoided. A disconnected graph is an indicator for the production of a suboptimal solution. In Figure 1.13 the problem can be solved by opening a facility at node u_4 instead of u_3.

Again, a test is necessary. If its result is negative, i.e. the copper net which is used by the supply area of the newly opened facility is also utilized by another customer who is not assigned to the new ARU, then the preceding assignment and opening step is revoked and the facility location in question is temporarily suspended. For the customer chosen in Step 1 or after the distance rule test an alternative facility location has to be found. The algorithm restarts with identifying all customers who can be assigned to this new candidate facility.

1.6.3 The CU Net algorithm

The considerations of the previous two sections are now formally stated starting with the definition of supply paths.

Definition 5 (Supply path) *The **supply path** of a customer is a directed path derived from the route of the copper wire which connects the customer to the central office. The nodes of the path represent switching nodes, muffles and similar components of a copper net which allow easy access to the copper wires. The edges define how nodes are connected by the copper wire of the customer and the sequence of the nodes — the direction of the path — determines the physical direction of data transmission from the customer to the CO along the wire.*

For example, the sequence

$$c_1 \rightarrow u_5 \rightarrow u_2 \rightarrow u_3 \rightarrow u_4 \rightarrow u_3 \rightarrow u_2 \rightarrow u_1 \rightarrow CO$$

which is taken from Figure 1.14 is an example for such a supply path. It illustrates that nodes may appear several times on a supply path.

Definition 6 (Copper network graph) *The **Copper network graph** or **trail graph** is the rooted undirected graph derived from the union of the supply paths of a given set of customers from a local loop. Multiple edges between the same pair of nodes are replaced by a single edge. Edges represent trails containing copper connections between nodes where they are provided by supply paths.*

An edge of the copper network graph may be used only in one direction like edges which enter the central office or leave the customer site. But, they also may be used in both directions like in the case of backward supply (Rückversorgung) (see Figure 1.14).

Definition 7 (Admissibility) *A node of the copper network graph is called an **admissible** facility location or **admissible node** for a customer, if it lies on the supply path of the customer*

and an assignment of the customer to a facility which is opened at this node is in accordance with the distance rule (D10). The ***admissible path*** *of a customer is the directed subgraph of all admissible nodes of the customer's supply path plus every edge between two admissible nodes.*

Definition 8 (Admissible node of greatest distance)
The ***admissible node of greatest distance*** *to a customer is the occurrence of an admissible node of a customer with the longest cable distance between the occurrence of the node and the customer site which is still in accordance with the distance rule R10.*

Customer c_1 from Figure 1.14 with supply path

$$c_1 \rightarrow u_5 \rightarrow u_2 \rightarrow u_3 \rightarrow u_4 \rightarrow u_3 \rightarrow u_2 \rightarrow u_1 \rightarrow CO$$

gives a few examples to illustrate the previous definitions. If the admissible path of customer c_1 is given by

$$c_1 \rightarrow u_5 \rightarrow u_2 \rightarrow u_3$$

then the admissible node of greatest distance is node u_3. If the admissible path of customer c_1 is

$$c_1 \rightarrow u_5 \rightarrow u_2 \rightarrow u_3 \rightarrow u_4 \rightarrow u_3 \rightarrow u_2,$$

then the admissible supply path contains more than one occurrence of nodes u_2 and u_3, and the admissible node of greatest distance is node u_2. But, if the admissible path looks like

$$c_1 \rightarrow u_5 \rightarrow u_2 \rightarrow u_3 \rightarrow u_4 \rightarrow u_3,$$

then the admissible node of greatest distance is node u_3 and not u_2. The latter is an admissible facility location for customer c_1 and the second occurrence of u_2 in the supply path is further away from customer c_1 than node u_3, but this second occurrence is not an admissible location for this customer.

Finally, if u_2 is not an admissible facility location for c_1, but u_3 is, then the admissible supply path could be

$$c_1 \rightarrow u_5 \quad u_3 \rightarrow u_4 \rightarrow u_3,$$

i.e. it is a disconnected graph containing two components.

The following two properties should be associated with admissibility to allow a constructive solution of the facility location problem at hand.

Definition 9 (Properties of admissibility)
(A1) There exists at least one admissible location for every customer in the graph.
(A2) If f is a node on the supply path of a customer c which is an admissible facility location for that customer then any node descendant to f is also admissible, i.e. any node on the path between c to f.

(A1) is a necessary assumption, in order to make the facility location problem feasible. (A2) is a more involved assumption. It states a sort of correlation between admissibility and the distance between a customer and its ancestors. The second assumption implies that for any node f on the supply path which is not an admissible facility location for a customer c, all ancestors are not admissible locations, too. (A2) further implies that the admissible path is a connected graph and therefore really a path.

Both assumptions imply that the customer's location must be an admissible facility location.

To complete this sequence of definitions the supply area of an access remote unit is described.

Definition 10 (Supply area of a facility) *Given the assignment of a certain set of customers to a common facility location, then the* ***supply area*** *or* ***access area*** *of the facility is the smallest subgraph of the trail graph which contains the shortest directed subpaths of the supply paths which connect each customer to the facility.*

With this definition Rule 7 (disjoint supply areas of ARUs 1.4.7) can be stated as the requirement that the intersection of the node sets of the supply area of two different facilities has to be empty. In other words, those parts of the supply paths which connect customers to their particular facility have to be disjoint for different facilities. This rule could also be formulated based on edge disjointness of supply paths which would give a weaker restriction.

Next the formulation of the complete algorithm follows.

Algorithm 1 (CU Net algorithm)

Preproc.	Fixate the distance rule and prepare the database. If applicable, remove CO circle.	
Initializing	Step 0.1	Determine the vector of cable distances DOC between customers and central office in descending order of the distance.
	Step 0.2	Determine the set of admissible paths SAP for all customers.
	Step 0.3	Set the iteration counter i to 1.
	Step 0.4	For every customer all admissible facility locations are marked as unsuspended.
Main	Loop	As long as there are customers listed in DOC:
	Step 1.1	Choose the first customer in DOC as the active customer c_a.
Part 1	Step 1.2	Determine the admissible and unsuspended location f_{c_a} for customer c_a of greatest distance according to Definition 8 using the distance as given by the distance rule.
	Step 1.3	Determine all customers $C_{f_{c_a}}$ who are not supplied yet and whose supply path contains f_{c_a}.
	Check 1.4	Is f_{c_a} an admissible node for all customers in $C_{f_{c_a}}$? No. Continue with Step 1.5. Yes. Continue with Step 1.8.

Continue on next page

	Check 1.5	Was the pair c_a and f_{c_a} already chosen during the actual — the i^{th} — iteration. No. Continue with Step 1.6. Yes. Continue with Step 1.7.
	Step 1.6	Determine the customer from $C_{f_{c_a}}$ of greatest cable distance to node f_{c_a} for whom this node is not admissible. Replace c_a by this customer and return to Step 1.2.
	Step 1.7	Remove all customers from set $C_{f_{c_a}}$ for whom f_{c_a} is not an admissible node. Document and report incident. Continue with Step 1.8.
Part 2	Step 1.8	Extract all nodes $N_{f_{c_a}}$ from the supply area of node f_{c_a} and the customers in $C_{f_{c_a}}$.
	Check 1.9	Does the admissible supply path of any unsupplied customer except those in $C_{f_{c_a}}$ contain a node in $N_{f_{c_a}}$? Yes. Continue with Step 1.10. No. Continue with Step 1.11.
	Step 1.10	Suspend the admissibility of node f_{c_a} for customer c_a. I.e. the node is marked as suspended for c_a. Document and report incident. Return to Step 1.2.
Part 3	Step 1.11	Open the i^{th} facility at f_{c_a} and assign all customers in $C_{f_{c_a}}$ to it.
	Step 1.12	Remove the admissible supply paths of all customers in $C_{f_{c_a}}$ from SAP and all their entries from DOC.
	Step 1.13	Increase the value of iteration counter i by 1.
	Step 1.14	If applicable, revoke the suspension of any node which was blocked during Step 1.10.

The graph which is derived from the set SAP in step 0.2 may not be connected. In practice it will have several components, each of which can be optimized on its own which can be utilized to reduce the size of the problem. This idea was followed for some time, and the components were called segments. Also, the number of segments is a lower bound for the number of facilities which are needed. Unfortunately, as soon as the problem of underutilization is tackled the concept of admissibility is weakened and the boundaries of segments become fuzzy. Then they cannot be worked on separately anymore.

The reason for Check 1.4 was already pointed out in the previous section and is mainly due to Rule 11 (individual distance rules 1.4.11).

Check 1.5 is necessary to guarantee the termination of the algorithm. Special network topologies may cause that the same customers and facility nodes are repeatedly selected. Figure 1.14 provides an example. It demonstrates the situation of backward supply (Rückversorgung) in copper networks, a frequently met phenomenon.

If the subscriber of $wire_1$ (c_1) is selected during Step 1.1, and if the furthest admissible node for this customer is u_4 and if this node is not admissible for the subscriber of $wire_2$ (c_2) and this subscriber is not assigned to a facility yet, then c_2 is selected in Step 1.3. The algorithm

detects a conflict during Check 1.4, since node u_4 lies on the supply paths of both customers. If in Step 1.6 customer c_2 becomes the next active customer and if its furthest admissible facility location is node u_2, customer c_1 is now selected during Step 1.3 and produces a conflict during Check 1.4. If c_1 is the only customer with a conflict, then it replaces the active customer which is c_2 in Step 1.6. From there on the algorithm is trapped and will not terminate anymore. That makes Check 1.5 necessary.

In Step 1.7 the resolution of the discrepancy which arises from a positive Check 1.5 is postponed to Check 1.9. The first facility location which appears a second time and causes Check 1.5 to be positive is suggested as a candidate. All customers who find no admissible location in the candidate are neglected. If Check 1.9 yields a negative answer, the assignment of Step 1.7 is not in conflict with R7 (Disjoint supply areas of ARUs) and is kept. Otherwise, in Step 1.10 the next attempt to resolve the problem is undertaken.

Whenever the algorithm enters Step 1.10 during an iteration of the main loop the admissibility of the current potential facility candidate is suspended. Since the problem is finite, i.e. there is only a finite number of potential facility nodes, the algorithm will run out of nodes to suspend during Step 1.10 sooner or later. As pointed out in Rule 2 (Customers and network graph 1.4.2) it can be assumed that all customers are leaves of the graph. This implies that the customer site is always an admissible node for the customer which will never produce a conflict in Check 1.9, since it is not contained in any customer's supply path except of that customer himself. Consequently, the main loop will terminate successfully after finitely many steps.

Check 1.9 is supposed to make sure that R7 (Disjoint supply areas of ARUs 1.4.7) is not violated. A guarantee is obtained, if admissibility is provided with property (A2) from Definition 9. Check 1.9 makes sure that only nodes from supply paths of customers who are previously assigned to a facility during Step 1.11 are removed from the problem during Step 1.12.[40], such that all admissible nodes of customers who still remain in the problem stay available for further iterations. Since the one set can be removed without affecting the other set, the two sets are disjoint.

However, this is not enough to guarantee R7. In Figure 1.20 admissible supply paths for two customers are depicted. The nodes ARU_1, u_1, u_3 and c_1 are admissible facility locations for customer c_1, and ARU_2, u_4, u_5 and c_2 for customer c_2. Node u_2 is not admissible for both customers. There is a "whole"in both supply paths. Of course, u_2 belongs to and is necessary for the supply area of both facilities. It is just not allowed to open a facility there.

If, for example, during the first iteration of the CU Net algorithm customer c_1 is active and assigned to ARU_1, during Step 1.8 all the nodes — even u_2 — of the supply path of customer c_1 up to the facility node will be identified and removed later in Step 1.12. Since u_2 is part of the supply path of the second customer, but not an admissible node for him, Check 1.9 will not trigger an alarm. The algorithm can move on to Step 1.12 and start the second iteration during which c_2 is probably assigned to ARU_2 which would cause a conflict with R7. Of course, at the same time property (A2) for admissible paths is violated, too. A formal proof that (A2) and Check 1.9 really guarantee R7 is given after the next paragraph.

[40]Note, these nodes need not to be admissible for these customers.

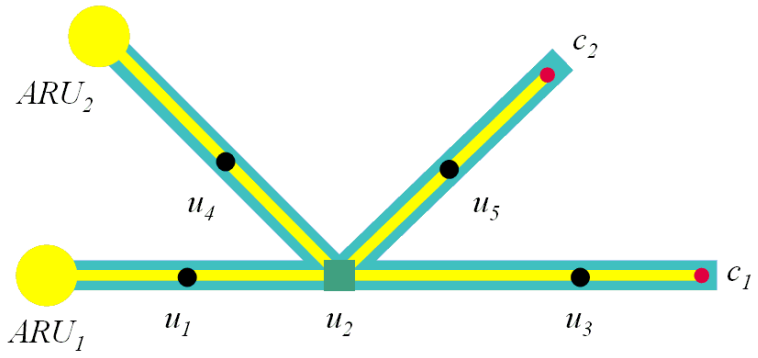


Figure 1.20 Admissible supply paths with wholes

Since there is always at least one of the finitely many customers selected and assigned to a facility during an iteration of the main loop, the size of the set of customers decreases because of the elimination during Step 1.12 and the algorithm will consequently terminate.

Proposition 1 (CU Net algorithm and disjoint supply areas of ARUs) *If the supply paths of all customers of the stated problem are in accordance with properties (A1) and (A2) for admissibility, then the CU Net algorithm yields a solution which complies with R7, the rule of disjoint supply areas of ARUs (1.4.7).*

Proof:
For any two facilities f_1 and f_2 which are contained in the solution provided by CU Net algorithm the following considerations apply. Without loss of generality f_1 is opened during an iteration prior to f_2. Customer c_2 is one of the customers assigned to f_2. Since c_2 is assigned to f_2, the location of the facility is admissible for c_2. Because of (A2), the supply path connecting c_2 with f_2 consists only of admissible nodes. I.e. the supply path between c_2 and f_2 is identical with the admissible supply path connecting them.

Since customer c_2 is assigned to its facility during a later iteration than the opening of f_1, c_2 is an unsupplied customer during this earlier iteration. Consequently, the entire admissible supply path of c_2 is cross-checked with the node set of the supply area of f_1 during Check 1.9. Obviously, the check was positive in the sense that no conflict was detected, since the algorithm moved on to Step 1.11, opened f_1 and assigned all customers to the facility to which they are actually assigned.

Therefore, the intersection of that part of the supply path which connects c_2 to f_2 and the supply area of f_1 is empty. Since this holds for all customers assigned to f_2 and any pair of facilities, the proposition is proven.
□

The manner in which the problems detected in Check 1.5 and Check 1.9 are treated in steps 1.7 or 1.10 respectively may cause suboptimal solutions, i.e. more facilities are opened than necessary. But at least, the problematic nodes of potential facilities and customers — the affected part of the copper network — can be documented and reported to the planner. This enables him to eventually resolve the problems by hand.

1.6.4 Lower bound for the number of facilities

How many facilities are necessary to supply all customers in the stated problem?

Given two customers who lack a common admissible location, then at least two facilities are necessary. If a third customer is found who cannot be supplied simultaneously from the same location with any one of the other two customers, three facilities will be needed. One way to determine at least a lower bound for the minimal number of necessary facilities is to find a maximal set of customers, such that no admissible assignment to a common facility exists for any pair of these customers.

This reasoning delivers "only"a lower bound for the number of facilities, since three customers may exist who share admissible facility locations in pairs. But, there is no admissible location for all three of them. Therefore, at least two facilities will be needed in this case.

The CU Net algorithm finds a set of facilities which enables to supply all customers. Clearly, there is at least one customer assigned to every ARU. If one customer is chosen from the assignment to every facility, a set of customers is determined which can be checked to see whether this maximality is fulfilled. If it is not fulfilled, the check itself produces candidates of customers who are superfluous in this set. Removing such customers will eventually produce a set of customers who do not share admissible nodes. Whether this set is maximal in the above sense, depends on how intelligently the customers are removed from the set of candidates.

Algorithm 2 (Lower bound for the number of facilities)

Preproc.	Run the CU Net algorithm. From every facility choose one assigned customer.	
Main	Step 1	Determine for every selected customer the number and a list of other selected customers with whom he shares at least one common admissible location.
	Check 2	Is there any customer for whom a positive number and a non-empty list was found? No. End of algorithm. Yes. Continue with step 3.
	Step 3	Solve the following linear programm: For every selected customer who shares an admissible node with one of the other selected customers introduce a binary variable $x_{c_i} \in \{0, 1\}$. $$\max \sum_{c_i} x_{c_i}$$ subject to $$x_{c_i} + x_{c_j} \leq 1$$ for every pair of customers (x_{c_i}, x_{c_j}) who share a common admissible node.

Clearly, if Algorithm 2 terminates in Check 2, this proves that the CU Net algorithm has found the optimal solution of the stated uncapacitated facility location problem with zero assignment costs and the same opening costs for all potential facility locations.

It may turn out that this is not the solution which is going to be built in practice. It may contain too many big sized facilities or locations where no consent to situate ARUs has been obtained. It may not even be the cheapest possible solution, because the suggested locations are too expensive and a lot cheaper alternatives are available. But, if the result of the CU Net algorithm is optimal, then it states the least number of access remote units which are necessary to fully supply all customers of a given local loop.

1.6.5 Optimality conditions for the CU Net algorithm

Theorem 1 (Sufficient condition for optimality of the CU Net algorithm) *For the UFLP with assignment costs equal to zero and identical opening costs where admissible facility locations are provided for all customers in accordance with properties (A1) and (A2) the result of the CU Net algorithm*

a) gives the minimal number of facilities which are needed to fully supply all customers of the problem,

b) provides a set of customers C of the same size as there are facilities with the property that no two customers from this set share a common admissible facility location,

if the CU Net algorithm never enters Step 1.10 during runtime, i.e. the admissibility of no facility location has to be temporarily suspended.

Proof:
Property (A1) together with (A2) simply guarantees that all customer sites are admissible facility locations and the CU Net algorithm will eventually terminate. Property (A2) ensures a solution which adheres to R7 according to Proposition 1. That all customers are assigned to some facility results from the termination of the algorithm. To prove optimality a set of customers will be provided such that no two of them share a common admissible location.

The algorithm does not only produce a set of facilities F, but also a set of Customers C, since every facility $f_i \in F$ is chosen with respect to a customer $c_i \in C$ during Step 1.2. According to this step the facility f_i is selected from all admissible and unsuspended facility locations of customer c_i of greatest cable distance. Now, since by assumption the algorithm never enters Step 1.10 during runtime, there are no suspended facility locations present at any time for any customer. Consequently, facility f_i is chosen from all admissible nodes for customer c_i of greatest distance according to Definition 8.

If $c_j \in C$ is a customer from an iteration subsequent to iteration i where customer c_i is assigned to facility f_i, then customer c_j is an unsupplied customer during the i^{th} iteration. Consequently, his entire admissible supply path is compared with the node set of the supply area of f_i during Check 1.9. Since f_i is on maximal admissible distance to c_i, the entire admissible supply path

of c_i is contained in the node set of the supply area of f_i. Since by assumption the algorithm moves directly from Check 1.9 to Step 1.11 skipping Step 1.10 the test proves that the set of admissible nodes of customer c_i and c_j are disjoint.
□

The algorithm itself produces means to decide whether the delivered solution is optimal. Of course, if Step 1.10 is entered, this does not imply that optimality is lost. A report about the affected nodes, locations and customers may enable the human planner to clarify the actual circumstances.

Concluding this section and Chapter 1 two results about the optimality of the CU Net algorithm are presented which reflect the structure of the trail graph of the given problem.

Corollary 1 (Rooted directed tree)
If

a) the trail graph of the copper network of the local loop in question can be described as a rooted directed tree with the central office CO as its root and directed from the customers to the CO,

b) such that the supply path of every customer coincides with the shortest directed path from the customer node to the root node in the given tree,

c) and admissible facility locations are provided for all customers in accordance with properties (A1) and (A2),

then the CU Net algorithm delivers the minimal number of facilities which are needed to fully supply all customers involved in the problem.

Proof:
During Step 1.3 of the CU Net algorithm the set C_{f_a} of all customers whose supply path contains the candidate facility location f_a and who are not yet supplied and therefore still contained in the problem, is determined.

Furthermore, f_a is a node of the trail graph which is a rooted and directed tree. Consequently, there exists a directed subtree T_{f_a} with root f_a which consists of all nodes whose directed shortest path between the node and the global root CO contains f_a. All directed paths of nodes outside the tree T_{f_a} do not contain f_a.

Since the customers in C_{f_a} are also represented by nodes in the trail graph and by assumption b) their supply paths coincide with the directed paths which connect the customer nodes to the root of the tree, all nodes from customers in C_{f_a} are contained in the subtree T_{f_a}. Moreover, every customer with a node in T_{f_a} has to be contained in C_{f_a} for the same reason.

To summarize, the set of customers C_{f_a} is equivalent to the set of all customer nodes in T_{f_a}.

In case a conflict arises during Check 1.4 — one of the customers in T_{f_a} does not have f_a as a admissible facility location — a sequence of customers c_i and corresponding candidate facility

locations f_i is produced. For each of the facility locations the argument from before applies again and the tree T_{f_i} contains the node c_i. The subsequent customer c_{i+1} is contained in T_{f_i}, but does not have f_i as an admissible location. This is the very reason why the sequence exists.

The candidate facility f_{i+1} which is chosen in respect to c_{i+1} lies inside of the tree T_{f_i}, because otherwise the node f_i is contained on the path from c_{i+1} to f_{i+1} which contradicts property (A2). Consequently, the tree $T_{f_{i+1}}$ is contained inside the tree T_{f_i} and the two trees are different, since their roots are different. Since the trail graph is finite, the sequence of subtrees must terminate and none of the candidate locations is picked twice. Therefore, Step 1.7 will never be entered.

Finally, in Step 1.8 the tree $T_{f_{c_a}}$ is presented. The set $N_{f_{c_a}}$ is identical with the nodes of subtree $T_{f_{c_a}}$. All the customer nodes in $T_{f_{c_a}}$ can be admissibly assigned to a facility situated at its root. Every unsupplied customer whose supply path contains a node from $N_{f_{c_a}}$ must be contained in $T_{f_{c_a}}$ because of assumption b).

Therefore, Check 1.9 yields a negative answer, the algorithm moves on to Step 1.11 and Step 1.10 is not entered. Since this is true for all iterations, the algorithm delivers the stated solution because of Theorem 1.
□

Unfortunately, this result can rarely be applied, since in practice the copper net of local loops will suffer from the problems described in Section 1.3.2 which contradict the required condition for the supply paths. The attempt to prove a similar result with the graph derived from the admissible supply paths of all customers with a similar condition for the embedding of the admissible supply paths in the derived graph, fails[41]. It is not enough to embed the admissible supply path. An alternative concept is more successful.

Definition 11 (Admissible trail graph) *The smallest subgraph of the trail graph of a given copper network into which the admissible paths of a set of customers can be embedded is called the* ***admissible trail graph*** *of this set of customers.*

In practice the admissible trail graph will not be connected. The tighter the distance rule R10 is, the more components the admissible trail graph will have[42]. If it is possible to find a node in one of the components, such that

(C1) this node is part of the supply path of every customer who is contained in the component,

(C2) the component of the admissible trail graph can be described as a directed tree with this node as a root and

(C3) the subpaths of the supply paths from all customers of this component to this node can be embedded[43] in this tree, such that the subpath of every customer coincides with the shortest directed path from the customer node to the root node in the given component,

[41] A counterexample can be constructed based on Figure 1.14 (backward supply).

[42] In the worst case let the distance rule be 0. Then only the customer sites are admissible locations and there are no edges in the admissible trail graph.

[43] Note, subpaths of the the supply paths are embedded and not the admissible supply paths.

then corollary 1 can be applied at least to this component. This way for every component which can be described by a rooted, directed tree which fulfills conditions (C1) through (C3) the minimal number of necessary access remote units can be determined. For the other components upper and lower bounds can be calculated and the problematic regions can be determined, reported and eventually treated by hand.

This result is summarized by the following corollary.

Corollary 2 (Forest of rooted directed trees) *If the admissible trail graph can be described as a wood of rooted directed trees and fulfills admissibility properties (A1) and (A2) and the component properties (C1), (C2) and (C3) for every component, then the result of the CU Net algorithm gives the minimal number of facilities which are needed to fully supply all customers involved in the problem.*

Proof:
The arguments following Definition 11 are applied to all components of the admissible trail graph.
□

Chapter 2

Network Quality and K-Median Problem

2.1 Project: Major challenge

On first of October 2009 the FTTC network of Villach (Carinthia) was put into operation under the name of Giga Netz. The planning of this network and another one for the city of Klagenfurt was done by the Telekom Austria planning group in Klagenfurt (also Carinthia). The project was partially funded by the EU commission and the regional government of Carinthia. The structural planning process had already started in the beginning of 2007 long before a FTTx strategy for all of Austria was worked out by the Telekom Austria Group.

In 2007 SARU was only known to people at TA headquarters in Vienna. I hadn't been informed about the Giga Netz project or been invited to be part of the project group in Klagenfurt. So, the project started without SARU. When I learned about the project, I suggested having at least one meeting with the project group, to introduce SARU and to compare results, which I would have calculated by that time. The proposal was accepted. The meeting took place on the 29^{th} of June 2007.

My goals for the meeting were twofold: on the one hand I wanted to learn more about the planning work, the strategies of the planners and the abilities and weaknesses of SARU to support the structural planning process, on the other hand I intended to use the meeting to promote SARU. I hoped to motivate the planners in Klagenfurt to use SARU in the project.

The results of the meeting were that I learned a lot about the weaknesses of SARU and I was not able to motivate the project team to work with SARU, because the tool had essential weaknesses at that time[1].

[1]Maybe there were additional reasons, why SARU was not welcomed with hurray. 1) Personnel reductions in Telekom Austria were still on the minds of all employees and deservedly so. During the following two years another 2.000 people left the company. And a tool that speeds up the planning process may also speed up the reduction process. Sometimes this issue was openly addressed in team meetings by suggesting a "careful"communication of the abilities of SARU to management personnel especially of a higher level. They might "misunderstand". 2) Some of the planners were of the old school. Certainly, all the planners were able to use computers (CAT) for planning. But, some preferred the good old paper and pencil approach. At a later date

To understand better what happened, we have to go back in time again.

2.1.1 A GIS tool

Soon after SARU kick off we started to discuss how and where to implement and maintain the software which necessarily would be developed in the course of the project. It was also clear that visualization-software was needed to make the results of an optimization tool accessible to users. At that time Telekom Austria already owned and used geographical information software called WebGIS[2] to keep and maintain data of all their facilities and networks. Moreover, WebGIS was under construction to be advanced to a planning tool for all detailed planning tasks. It almost suggested itself to integrate SARU in WebGIS. But actually, it was Andreas Ludwig — a Telekom Austria employee in charge of WebGIS — who suggested it. Additional functionality would enhance the value of WebGIS not only for this specific purpose but certainly also for other planning problems (e.g. green field planning). That was one of his main arguments.

A project team was put together to compile all necessary specifications for the tool which was to be developed[3]. The team comprised people from rmDATA who were in charge of the specification process, planners, technicians for outside plant and the members of Operations Research group. The part of the specifications which is concerned with optimization is documented in Chapter 1 Section 1.4.

Finally, rmDATA delivered a quotation based on this specification list. But, their quote for the integration of the optimization software in WebGIS and the visualization of the optimization result was so high that nobody was willing to open up sources for this amount of money.

Shortly before the meeting with the planner group in Carinthia I was informed about this decision. The importance of the meeting in Klagenfurt increased. There was a necessity, but also a chance of finding allies to put pressure on managers who had access to funding.

2.1.2 Improvised visualization

However, I was not prepared to communicate results of my optimization tool directly to planners. I had no reason to. For several months I was part of a project team whose definite goal was to specify and develop a visualization tool for exactly that purpose.

Certainly, I had developed some means to access and visualize results for myself. After all it was necessary to check the software I implemented by inspecting solutions. Since all the vertices of the network were furnished with geographical coordinates it was easy to produce a primitive map of the solution with the help of a MATLAB scatter plot. Even identification information could be easily attached to each vertex. This ID information was used to search and track vertices of interest in the database which contained all network information. To go the reverse way — picking a vertex from the solution database and trying to find it on the scatter plot — was more difficult.

---

the Carinthian group decided that half of the planners would use SARU and the other half would stick to paper and pencil.

[2] WebGIS was developed by rmDATA.

[3] See also Chapter 1 Section 1.2

2.1.3 Different IDs

To make matters worse Telekom Austria planners use a different identification key for switching nodes than we did in SARU. The ID is called Schaltstellenbezeichnung (switching node labeling). A local loop itself is organized in subdistricts. This Schaltstellenbezeichnung reflects this structure of subdistricts. By using this ID a planner can identify two things:

1. Approximately where a switching node with a certain ID is located. (He knows that because he knows his local loop.)

2. Which switching nodes have a logical connection to each other, i.e. whether they lie in the same subdistrict and therefore may be served by the same ARU.

For example, the Schaltstellenbezeichnung "2766-02-1- -G-9-"identifies the local loop by "2766-02". The letter G tells the planner that the switching node in question, i.e. the switching node with number 9, lies in the subdistrict which is supplied via cable branching point G (German: Kabelverzweiger). So, switching node "2766-02-1- -G-6-"lies in the same local loop and has a close common "ancestor"in common with the previous example, namely branching point G.

The Schaltstellenbezeichnung is an alpha numeric ID which makes it a bit uncomfortable to use within MATLAB. Numeric IDs are easier to handle. This was not the only reason why I introduced a different and numeric ID for all the vertices. The Schaltstellenbezeichnung is only defined for switching nodes. Many nodes of the local loop are not switching nodes (e.g. muffles) but important for finding a good solution to the facility location problem. So, at least at a certain point of the implementation of the optimization software it seemed the best solution to use a newly created numeric ID for all vertices.

As a consequence, during the feedback meeting in Klagenfurt I had to translate the Schaltstellenbezeichnung into the SARU internal ID of a node by inspecting a database when ever a planner wanted to know something about a certain switching node.

2.1.4 The meeting

An information retrieval system of this kind was confronted during the meeting in Klagenfurt with questions like the following:

1. Which ARU supplies customer K or his switching node respectively? Clearly, the switching node was addressed by its Schaltstellenbezeichnung.

2. What is the cable distance between the two?

3. Who are the customers supplied by ARU X? Can we see this set?

4. Which ARUs are situated on switching nodes and what is their Schaltstellenbezeichnung?

And since the facility location solver does not situate all ARUs at switching nodes,

5. What is the closest switching node to such an ARU and how far away is it?

The colleagues in Klagenfurt spent an entire day with me discussing solutions by comparing their paper plans with the MATLAB scatter plots, asking their questions and watching me tracing through the database. Not surprisingly, they found this process very tiresome and exhausting. It became evident that they could not utilize this tool without my help and presence.

2.1.5 Prototype

As a consequence, it became a must to make a proper visualization tool, or even better, an adequate information retrieval tool[4] available. This was the starting point for the development of the SARU prototype - a tool to solve the facility location problem for a given local loop, to visualize and analyze the solution and subsequently to facilitate some interactive functionality which allows the planner to alter the suggested solution. It was implemented in MATLAB[5].

During the following two years this prototype formed the center of the further specification process. A key user group of planners from all over Austria was assembled. They were trained to use the prototype and applied it to their local loops. This consequently advanced the prototype again.

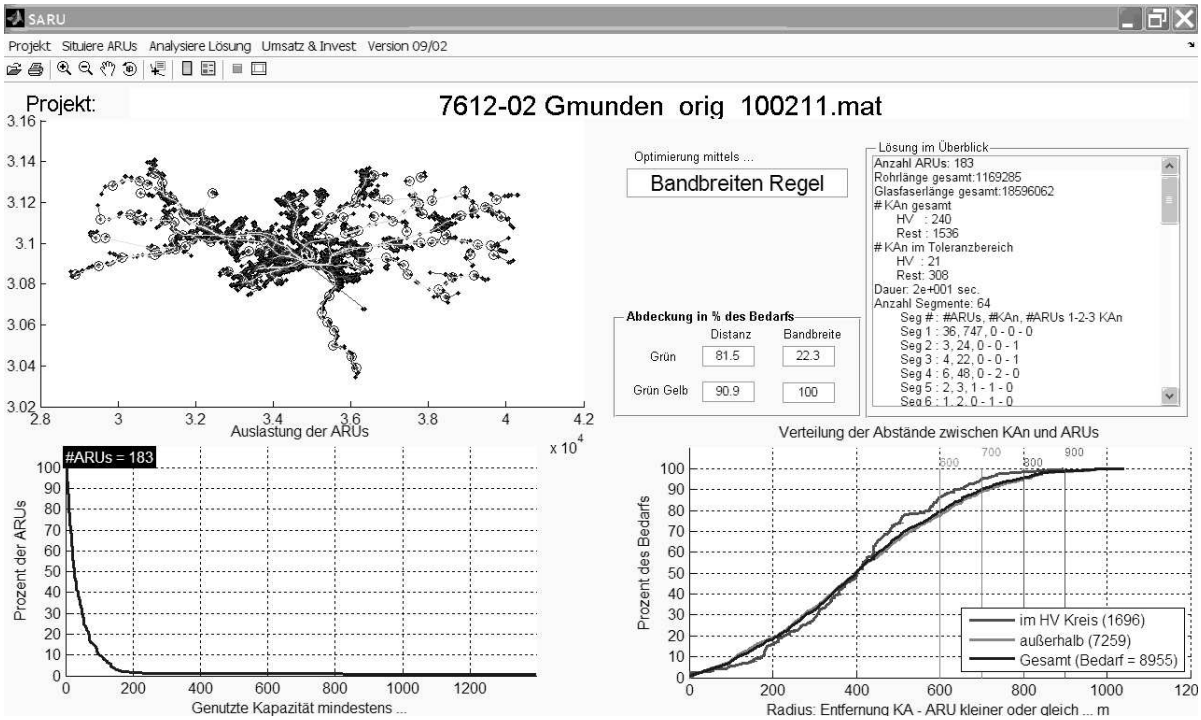


Figure 2.1 Screenshot from first prototype, main window. It provides an overview of the solution (number of ARUs, facility utilization, distribution of transmission rates, ...).

[4]Information retrieval is more than visualization. Visual representation does not easily answer questions about cable distances for example.

[5]Back from Klagenfurt in Vienna I immediately started with the implementation. It was not a moment too soon as I was told later. The sponsors of project SARU started to consider alternative approaches to tackle their planning problems and to release OR from the task.

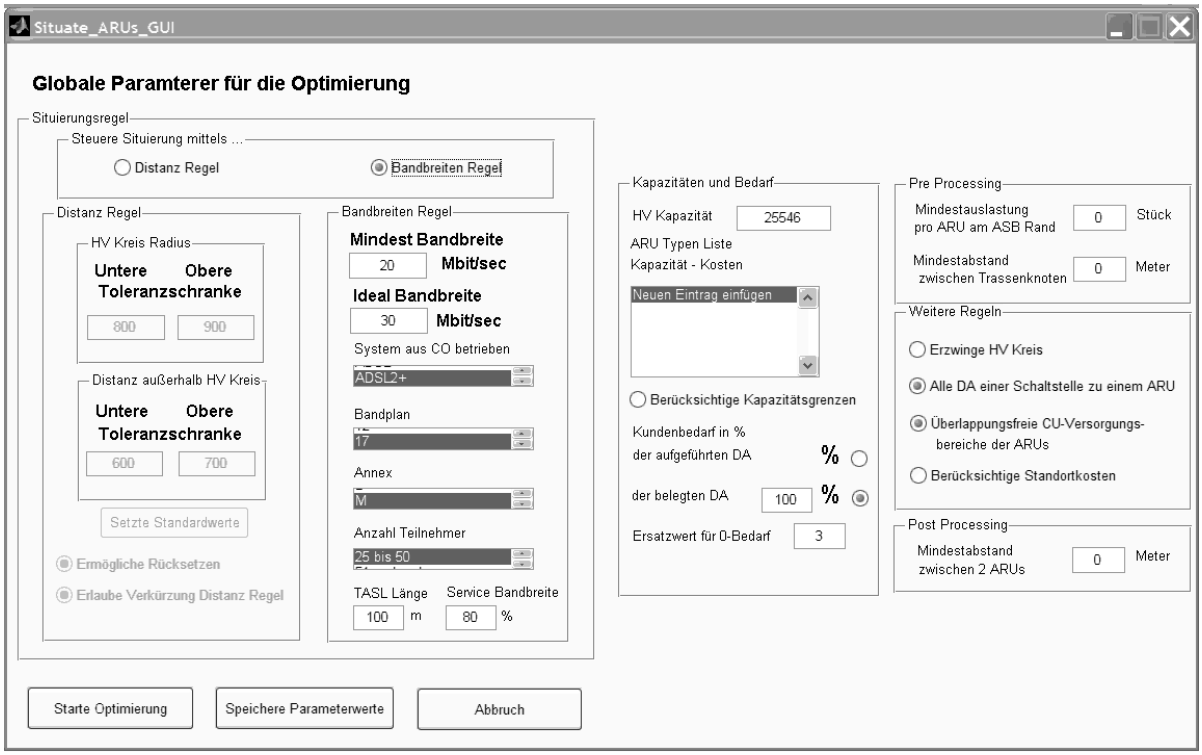


Figure 2.2 Screenshot from first prototype, parameter GUI. It allows the user to enter all strategic and global parameters (damping is not included)

All in all the prototype fulfilled four purposes:

1. It supported and advanced the specification process for the optimization problem and the visualization tool.

2. On its own it was an optimizer and visualizer; it was used to study and solve facility location problems for several local loops.

3. It allowed communicating the advantages to support the structural planning process through optimization. Subsequently, this convinced the leading management to fund the originally planed integration of SARU in WebGIS. The project which originated was called SARU P2 (SARU Phase 2).

4. The prototype was the materialized specification list for SARU P2.

2.1.6 Capacity utilization

The second lesson from the meeting with the planners in Klagenfurt concerned the optimization algorithm itself.

The first solution I presented in Klagenfurt was immediately dismissed. One of the planners literally called the solution non optimal. Others described it as unrealistic. They asserted that solutions like that would under no circumstances ever be planned and realized, except if they

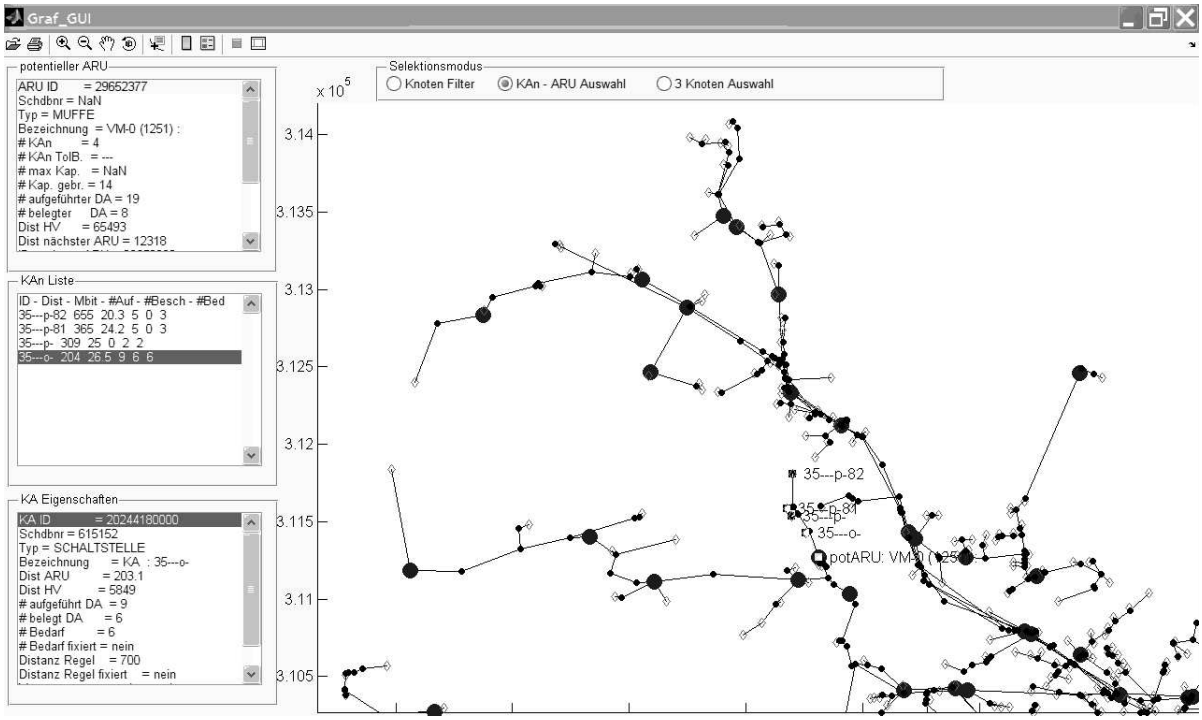


Figure 2.3 Screenshot from first prototype, interactive window. So called because it allows the user to investigate details of the solution (e.g. individual ARU and assigned customers) and even change locations of facilities and assignments of customers.

were forced to by their superiors. And even then they would want the order to be put down in writing. It was a very emotional situation on both sides. I was astonished and asked for the reasons.

The reason was easily identified in the degree of capacity utilization of some of the suggested ARUs. Some of the ARUs in the solution had got low capacity utilization. Only three to four customers were assigned to one ARU. In other local loops some ARUs served only one customer. Considering that at that time ARUs were designed for servicing up to 170 customers and the cost only for the equipment of an ARU[6] was around 30.000 €, it was understandable that the planners rejected the solution.

Certainly, we could have anticipated the underutilization problem during the specification process. In retrospect this was especially true. Then, all of a sudden, it seemed such an obvious problem that it was impossible to explain why nobody had thought of it. However, we simply had not. Nobody had thought of it through the specification process, and of course nobody had talked about it. We had discussed costs and distance rules. Even overutilization[7] had been a topic. But, we had not envisioned low utilization problems.

There they were and every body could see them projected on a wall: a list of ARUs together with their capacity utilization sorted in decreasing order. This list represented the suggested

[6]This number does not include the necessary fibre infrastructure to connect the ARU with the central office.

[7]Too many customers are assigned to one facility. Overutilization is solved by capacity constraints for facilities. See also the discussion in Chapter 1 Section 1.5.3.

solution for a local loop the planners knew. They could see where the low utilized ARUs were located and judged quickly.

2.1.7 Incomplete specifications or new insights

These events had clearly a high impact on my mission in Klagenfurt. I was there to find allies for fund-raising to realize SARU. Instead I found fundamental critique. Not only was a proper visualization tool missing, but also the optimization concept was incomplete and needed adapting.

I was tempted to seek a scape goat and found a candidate in an incomplete specification list. I argued that based on the list of specifications[8] that I had been given before, the presented solution was optimal. The solution did not make an optimal impression, because a criterion was applied which was not part of the original specifications. I was not very convincing. The planners looked at the solution and saw how they could improve it by hand. So why should my solution be optimal? Later I found a different way of looking at it.

What is the major difference between the specification process and what happened during the meeting in Klagenfurt?

It is the same difference as between studying a cake recipe and eating the cake. Approaching the cake by studying its recipe may very well give insights which stay hidden by only tasting it (e.g. necessary cooking time and temperature). The same is true the other way round. Tasting a cake whose recipe asks for black pepper, will allow a final judgment on the amount of black pepper used.

Or it is like renovating a bathroom. One will spend a lot of time carefully planning it, choosing the tiles, placing the lights and discussing the furniture. But, once the bath is ready and in use for a while, chances are high that the bath owner would want to change the one thing or another, if he only had the chance to.

Another way to see the difference is on the level of signs and representations. The specification list lives in the world of words and descriptions. Charts may illustrate the sentences. But, the solution of the optimization, once properly visualized (!), lives in a different world, a world which planners are very familiar with - the network plan.

Specifying all rules, goals and constraints of an optimization project and looking at the result of the optimization based on these specifications is a change of perspective on the optimization problem.

This change of perspective may help to see possible misunderstandings and miss-specifications. It may also produce new insights to the given problem and consequently lead to new specifications which could not have been foreseen during the first specification feedback loops.

Going back to the example of the bathroom, to misplace power outlets, to plan too many of them or simply to forget them, rank probably among the most typical miss-specifications

[8] Compare Chapter 1 Section 1.4.

during renovation work. To change the mind and to want a shower separate from the bathtub instead of the shower the way it has been realized inside the bathtub, is more like a new insight, a re-evaluation of ones own preferences after using the newly renovated bath for some time.

Typical misunderstandings and miss-specifications in the context of cooking are easy to imagine. And they will be detected by tasting the result. However, with recipes from a cookbook we presume that they have been proven and tested. So, if we just stick to the recipe, the result should be fine. But, any recipe was created at some point in time. It is hard to imagine that this was done without trial and error.

2.1.8 Prototyping

The specification process for an optimization project-formulating and defining the problem, collecting all side constraints, restrictions and rules which are to be obeyed by the problem solution - may be enhanced by discussing the results of this process based on a prototype which should be as close to the desired final result as possible. This second stream of feedback which closes the feedback loop could be called Prototyping.

2.1.9 Capacity utilization problem

Prototyping - on a very primitive level - helped us during project SARU to detect the underutilization problem. The planners immediately identified a weakness of the suggested solution and hence of the underlying optimization model by inspecting the solution. They could even improve the presented solution by hand. However, they were not able to supply us with a general strategy of how to conceptualize and finally tackle the underutilization problem. It became a task for Operations Research to suggest a concept and find a solution for the stated problem. The term supply rate best describes the suggested concept and the theory of the k-median problem was used to solve the resulting facility location problem.

Another capacity utilization problem is overutilization of facilities: too many customers are assigned to one facility. This problem is addressed and discussed in Chapter 1, especially in Section 1.5.3.

2.2 Strategy: Network quality and Coverage constraint

2.2.1 Pruning and strategic parameters

The first attempt to solve the underutilization problem was more like an imitation of what the planners' first reaction to SARU solutions was: a pruning strategy. If utilization of a facility is too low, it is removed from the solution and its customers - when possible - are served by the next closest facility. Since the CU Net algorithm, Chapter 1 Section 1.6.3, runs rather quickly, it can be used to automate this process the following way:

1. The CU Net algorithm is applied to the network.
2. A facility of low utilization is identified.
3. The individual distance rules[9] for the customers it supplies are relaxed in a way that new locations for supplying facilities are possible.

These three steps are repeated until all facilities are of a minimal degree of utilization.

Some care has to be taken in identifying the facility in Step 2. Preferable the facility should be the only one in the subtree rooted at its location. In other words, the facility should supply all customers contained in the tree which is rooted at the location of the facility.

This approach calls for a strategic parameter: Minimal Degree of Capacity Utilization. In contrast to a constant, it should be a parameter because its value might vary over space and time. Different values may be used in different local loops, and nobody knows what kinds of facilities will be available tomorrow or what their best minimal utilization might be.

Furthermore, the parameter is called strategic until somebody has decided what its best value is. Some doubt may exist about its best value, because a slightly lower value might produce a much more qualitative solution with reasonably increased cost. Or a slightly higher value might save a lot of cost with a comparable low decrease of quality.

2.2.2 Underutilization and quality

Of course, such a pruning strategy effects the resulting solution. Especially, it changes transmission rates for customers whose facility is removed. Relaxing individual distance rules implies increasing the distance between customers and supplying facilities which increases the damping which in turn leads to lower transmission rates. Figure 2.4 depicts the worst case scenario.

The left most facility suffers from underutilization (ARU_1). The pruning strategy removes it from the solution. The effected customers (red round nodes) have to be assigned to a different facility. But, the next best location inside a more densely populated area is too far away (ARU_2). The transmission rates for the red round customers will be zero. The same level of quality of service is not granted to all customers anymore to the extreme that some customer will be provided with no service at all.

[9] Compare Chapter 1 Rule 11 individual distance rules, 1.4.11.

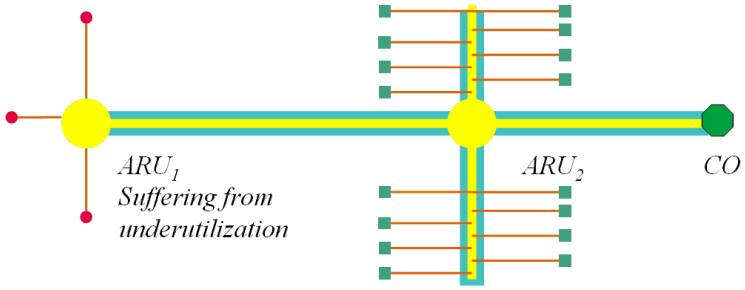


Figure 2.4: A detail from a local loop with an underutilized facility

To summarize, improving the solution of the FTTC planning problem in the sense of making it more efficient by removing underutilized facilities will worsen the solution by lowering the overall service quality.

As obvious as this may seem to be, it still remains to consider one important question. How can quality be conceptualized for the task of FTTC network design?

In SARU the concept of network-quality was discussed the first time after we became aware of the underutilization problem. Until then all considerations of quality existed outside of the projects' boundaries. They were not part of the project. They only entered, or better touched the project through the strategic parameter distance rule R10 (Chapter 1, Section 1.4.10). Of course, quality was supposed to be controlled by the values of this parameter. But, the values were to be given from outside. Optimization had no task in defining them, finding the "best"values, or controlling any aspect of quality. Quality became an issue for optimization, as soon as the same quality standard should no longer be guaranteed for all customers, i.e. as soon as quality itself began to vary.

So, how can quality be conceptualized? It can be done by the demand-weighted average distance of the customers to their supplying facility. A low average value indicates high quality. Since averages are very sensitive towards outliers, quartiles can be used alternatively. A low value of the third quartile may stand for good quality, or if distance is measured in terms of the transmission rate, a high value for the average or the first quartile will indicate good quality.

In sales and marketing however, average quality is a problematic idea. Average quality hardly sells. For a specific product, either a certain quality is guaranteed, or displeased customers will be the result. This connects the quality of the facility location with the requirements of a future product. Interesting side effect - this discussion thereby links network planning strategies with marketing strategies[10]. So, if the design for the mainstream product asks for a minimum of 16 Mbps transmission rate[11], a measure of quality for the solution of the facility location is the percentage of customer demand which is satisfied beyond that limit.

[10]This may produce new and necessary communications, but certainly will also produce new problems because more people are involved.

[11]In August 2010 such a setting was analyzed for two products — Giga Speed 16 and Giga Speed 30 with distance rules 10 Mbps and 16 Mbps respectively — for all of Vienna.

2.2.3 Rule 15 (R15): Coverage constraint

The percentage of customers who are guaranteed a certain quality of service through a given solution will be called supply rate. To aim for a solution which delivers a certain supply rate constitutes a coverage constraint.

This coverage constraint has to be added to the list of specification and rules listed in Chapter 1 Section 1.4 as Rule 15 (R15) coverage constraint. Additionally, Rule 5 (Full coverage of customer demands, 1.4.5) has to be suspended, since it obviously contradicts the coverage constraint R15.

In this context, it is of secondary importance if quality of service is measured in terms of cable distance between customers and supplying facility or transmission rate between them. Although, for all customers it should always be the same criterion through out an optimization run.

This concept was finally utilized in SARU.

2.2.4 Maximal quality for minimal cost

Maximizing quality and minimizing cost simultaneously leads to a bi-criteria optimization problem. Several strategies are known to cope with two target values at the same time:

1. Constructing one target function by combining the two targets[12], e.g. by forming weighted differences of the two original target values.
2. Removing one of the target values from the objective function and transferring it to the set of constraints.
3. Studying the bi-criteria target function.

The key to the first approach is to know how to weight cost compared to quality. Since cost and quality are usually measured on different scales[13], this becomes a challenging and fanciful task. How much quality is one Euro? If possible, this question can only be seriously answered after thoroughly studying cost and quality relation of the given problem. This path was never pursued in SARU.

The second approach produces two variations. Either the budget for the problem solution is limited and the quality is maximized (budget constraint) or a minimal quality standard is established and cost is minimized (quality constraint or better coverage constraint). In both cases a parameter is introduced - budget ceiling and minimal supply rate respectively. The problem of weighting cost against quality is now disguised in picking values for one of the parameters. Again, since it is a priory not obvious, what their best values are, the parameters become strategic. However, there is a difference between the two variants, which was important

[12] More general, the concepts of quality and cost may be combined in a concept for efficiency.

[13] In regard to SARU cost is measured first in terms of number of facilities and later in Euros. Quality which is defined as supply rate measures the percentage of satisfied demand, i.e. number of supplied customers.

for SARU at least until a certain point in time. Cost was defined by the number of facilities and not by Euros for the facility location problem. So, cost did not correspond to real cost, i.e. money. But, even if the number of ARUs had been converted into a Euro price, the estimated cost would not have become a lot more realistic. (See the discussion in Chapter 1 Rule 9 ARU setup cost 1.4.9 and Section 1.5.3 structural planning process.)

According to a Telekom Austria internal analysis of the cost structure for building FTTC it was estimated[14] that between 40% and 45% of the total cost would be allotted to street cabinets, i.e. facilities. So, facility costs are rather important. The left side of the budget constraint would therefore be quite unrealistic. The right side of the constraint — the budget ceiling — stayed unrealistic, too. As long as no final plan for all of Austria including an annual budget was provided and no strategy as to how this budget was to be split between different local loops, there was no hope of defining a proper budget ceiling.

So, the task of defining the budget constraint became one of limiting unrealistic cost by a budget which was not known (yet). It is not advisable to enter a feedback loop with vague concepts like that. A spiral of questions and discussions about how unrealistic the costs and the budget are will be the result.

The coverage constraint stayed an unknown for a while, too. The products for which the network was to be planned were not designed yet. A basic long term strategy for evolving the local loops was not delivered yet. But, quality in terms of supply rate was much easier to work with. It was easy to visualize as the number of customers who would be provided with a certain standard[15]. It stayed the same even if underlying prices changed, and it could be interpreted the same way even if the standard it was based on changed.

This is not a principal argument against budget constraints. It gives the reasoning for using a coverage constraint in SARU. At least for the first phase a quality constraint seemed more reasonable than a budget constraint.

2.2.5 Decision support

The third of the three stated resolutions of a bi-criteria optimization problem leads to the concept of decision support. Instead of awaiting or guessing a proper value for a strategic parameter a systematic study of the value space of the parameter in question is performed and the effects on the solutions are described. This may help to draw a picture which supports to decide which the best values or ranges for the parameter finally will be. It may at least provide material for further discussions. And it will certainly release some of the pressure which results from the uncertainty of the strategic parameters.

Both goals, implementing the coverage constraint in SARU and supporting the exploration of the parameter minimal supply rate, were realized by mathematical theory based on the k-median problem.

[14] Dipl. Ing. Gerald Clerckx (2007)

[15] A different problem is that the strategic parameter minimal supply rate is based on another strategic parameter which is customer demand. See Chapter 1 Rule 3 customer demand 1.4.3.

The importance which was ascribed to the resolution of the underutilization problem by the study of capacity constraints lead to the prevalence of the underutilization problem over the overutilization problem as already discussed in the beginning of Section 1.6 of Chapter 1.

2.2.6 Coverage constraint and k-median problem

The k-median problem - at least the way it is applied in SARU - is the task to open k facilities in such a way that network quality is maximized, i.e. as many customers as possible are guaranteed a certain service level. At this modeling stage of SARU cost is expressed by the number of open facilities. Hence, the k-median problem actually solves the facility location problem with a budget constraint, a fixed number of facilities.

To solve the facility location problem with a coverage constraint the k-median problem has to be solved for different values of k until the lowest value of k is found for which the minimal supply rate holds. This procedure harmonizes with the standard strategy for solving the k-median problem for directed trees which is recursive: to solve the k-median problem one solves first the $(k-1)$-median problem. The facility location problem with a coverage constraint is therefore solved by finding the k medians for increasing values of k, starting with k equal to 1 until the minimal supply rate is hit.

Moreover, by means of this procedure parts of the parameter spaces are explored[16] resulting in a list of the number of opened facilities, the corresponding supply rates and the minimum facility utilization. So, this approach additionally provides an instrument for decision support.

Unfortunately however, the parameter spaces are explored from the wrong direction. Starting the exploration at k is equal to 1 the optimal value for k is approached from below. Although, the underutilization problem forces us to waive some of the network quality, the aspired quality level will still be high. In turn high network quality corresponds to high cost, i.e. a high number of open facilities. So, for both parameters the really interesting regions of the parameter spaces are those of high values.

This asks for approaching the optimal value for k from a higher value. Starting with the largest possible value for k, the value which corresponds to a supply rate of 100% coverage solutions are explored and supply rates are determined by successively decreasing the value of k until the coverage constraint (R15) is violated. In fact, to stretch the possibilities for decision support the search may be continued below the optimal value to give an impression of the solutions in the neighborhood of the optimal solution.

In the literature about k-median for directed trees such an approach could not be found. On the one hand, it seems more natural to start iterations from zero — to start from scratch so to say — than to start with a solution which supplies all the customers' demands[17]. On the other hand, the applications of the k-median algorithms for directed trees described in the literature

[16] Since the parameter space of the number of open facilities is discrete and finite unlike the parameter space of the minimal supply rate which is continuous and infinite, it makes more sense to conduct the exploration from the perspective of the number of facilities.

[17] Nonetheless, such a solution is easily determined. It will suffice to open one facility per customer. However, the CU Net algorithm will quickly give a more reasonable solution.

suggest that their main interest lies in low values for k. Special and faster algorithms were developed for the cases k equal to 1, 2 or 3.

The foregoing considerations of applying the k-median concept to the FTTC planning problem are made with respect to network graphs which can be represented by rooted directed trees. However, in Chapter 1 Section 1.3.2 it was thoroughly discussed that the underlying graph of the copper network is not necessarily a directed tree. Of course, the concept of k-median and algorithms for solving the k-median problem also exist for general networks and graphs. The main arguments to apply the k-median problem for directed trees to the design problem in SARU were:

1. The graphs of copper networks are very similar to directed trees.

2. Those parts of the network graphs which diverge from the tree structure (see Chapter 1 Section 1.3.2 missing parallelism and backward supply) must not be part of the supply area of an access remote unit because of Rule 7, Disjoint supply areas of ARUs, (Chapter 1 Section 1.4.7). The basic argument simplifies to the requirement that the supply area of facility has to be a directed tree.

3. As already mentioned the recursive nature of the algorithms solving the k-median problem is beneficial for providing decision support.

Consequently, it is necessary to extract a directed tree or a forest of directed trees from the underlying network graph before applying a k-median algorithm for directed trees to solve the design problem.

2.3 Theory: FTTC and k-median

The k-median problem is well studied as can be seen from two bibliographies on the subject. First, there is [44] an annotated bibliography of Reese published in 2006. Reese started his paper with a brief history of the problem which he dated back to Fermat. Then he quoted, analyzed and categorized 132 papers. Categories are derived from the different variants of the problem and solution methods. Second is Shindler who focused on the special case of the metric k-median problem. In [48] which he published in the internet in 2008 he analyzed 44 papers 3 of which date after 2006.

2.3.1 The k-median problem as a facility location problem without opening costs

In essence the k-median problem is a facility location problem with a constraint imposed on the number of open facilities which has to be equal to k. Hence, one equation has to be added to the set of constraints of the LP formulations of the FLP as they are stated in Chapter 1 Section 1.5.2:

$$\sum_{f \in M} y_f = k.$$

This condition resembles a budget constraint, although it is none. Facility costs of feasible solutions range between the total sum of opening costs for the k cheapest and the k most expensive locations. But, assignment costs are still unrestricted, and the sum of facility and assignment costs is minimized.

However, the versions of the k-median problem which are found and solved in the literature differ from the definitions of facility location problems in one important point. They do not consider facility opening cost. This fact and the additional side constraint combined with — for example — Definition 2 of the capacitated facility location problem (Chapter 1 Section 1.5.2) results in the following formulation:

$$\min \sum_{c \in N} \sum_{f \in M} AC_{cf} * x_{cf}$$

subject to

$$\sum_{f \in M} x_{cf} = 1 \qquad \forall c \in N \tag{2.1}$$

$$\sum_{c \in N} d_{cf} * x_{cf} \leq Q_f * y_f \quad \forall f \in M \tag{2.2}$$

$$\sum_{f \in M} y_f = k \tag{2.3}$$

$$x_{cf} \in [0,1] \qquad \forall c \in N \quad \forall f \in M \tag{2.4}$$

$$y_f \in \{0,1\} \qquad \forall f \in M \tag{2.5}$$

where c is a member of a set of customers N which has to be assigned to one or more facilities f from a set M. Assignment costs are denoted by AC_{cf} and depend on customers and facilities; whereas x_{cf} are the assignment variables whose values characterize what percentage of a customers' demand $d_{cf} \geq 0$ is assigned to which facility. Assignment constraint (2.1) and the constraint (2.4) guarantee that in total exactly 100% of every customers' demand is assigned to some facilities. Capacity constraint (2.2) ensures that the capacity limit $Q_f > 0$ of a facility is obeyed for every facility. Opening variables y_f are either zero or one because opening a facility knows only two states: it is opened or stays closed.

In this formulation the problem constitutes a mixed-binary linear program.

Opening costs may be ignored, in case they are irrelevant for the application which is studied. Or they are irrelevant with respect to total costs, since total costs are dominated by assignment costs. Another reason may be that opening costs are (nearly) identical and therefore independent of locations. In this case the fixed number of facilities suffices to control opening cost[18].

But, for solving a FTTC problem opening costs are known to be substantial and are expected to vary depending on the location (See Chapter 1 R9 1.4.9). Opening costs depend on the equipment for the facility, the size of the facility and the location where the facility is installed. But, to properly estimate the price for a certain location it has to be inspected and evaluated. The final price is then determined in negotiations with the owner of the location. To survey all of this information in advance for all potential locations spread over a specific local loop would not only be very costly and time-consuming, but also a waste of resources. After all, only a small percentage of all these locations are going to shelter facilities. Furthermore, according to the discussion in Chapter 1 Section 1.5.3, the goal of the structural planning process as it was stated by the planners is to determine locations which are worthwhile to investigate.

Consequently, it was decided for SARU that opening costs are controlled by the number of facilities.

2.3.2 First variations

Variations of the k-median problem are concerned with the demand of the customers and the capacity of the facilities.

The demand of a customer is said to be splittable[19] (divisible), if the customers' demand is allowed to be satisfied by more than one facility. Accordingly the k-median problem is called one with splittable demands. For the formulation of the k-median problem as a linear program splittable demands are realized by real valued assignment variables with values ranging from 0 to 1. Opening variables remain integer valued.

In case of SARU only unsplittable demands are admissible[20]. Assignment and opening variables are dichotomous taking values 0 or 1. Hence, the k-median problem with unsplittable demands is a binary integer linear program.

[18] In this sense there exists a budget constraint on the opening cost.

[19] Terminology taken from [33].

[20] This is stated in Chapter 1 by specification R4 (Undivided customer demand, 1.4.4) and also a consequence of specification R7 (Disjoint supply areas of ARUs, 1.4.7).

Considering facilities whose capacities are bounded, i.e. only a limited sum of customer demands may be satisfied by a single facility, leads to the capacitated formulation of the k-median problem[21]. Actually, SARU should be modeled as capacitated KMP. However, during the specification process planners made clear that they preferred solutions which reveal those locations where demands of customers are concentrated[22]. Hence, the modeling for SARU is based on the uncapacitated k-median problem.

Summing up, SARU was modeled as an uncapacitated k-median problem with unsplittable demands. Actually, in this context the term unsplittable demands becomes dispensable, since the need to divide demand between several facilities ceases to exist, once all facilities grant infinite capacity. Below follows the binary linear programm formulation of the problem:

$$\min \sum_{c \in N} \sum_{f \in M} AC_{cf} * x_{cf}$$

subject to

$$\sum_{f \in M} x_{cf} \quad = 1 \quad \forall c \in N \tag{2.6}$$

$$x_{cf} \quad \leq y_f \quad \forall c \in N \quad \forall f \in M \tag{2.7}$$

$$\sum_{f \in M} y_f \quad = k \tag{2.8}$$

$$x_{cf} \quad \in \{0,1\} \quad \forall c \in N \quad \forall f \in M \tag{2.9}$$

$$y_f \quad \in \{0,1\} \quad \forall f \in M \tag{2.10}$$

This definition corresponds to the one given by Reese in his bibliography [44].

Similar, to the facility location problem in Chapter 1 at the end of Section 1.5.2 the compliance of the results of this linear programm with all the specifications listed in Chapter 1 Section 1.4 — except of course R5, full coverage — has to be assessed.

However, side constraint (2.6) seems to enforce Rule 5 (full coverage, 1.4.5) again. This is the case insofar every customer has to be assigned to some facility. But, unlike the ILP formulation for the facility location of FTTC where assignment costs AC_{cf} are meaningless and therefore set to zero[23], assignment cost are now used to express a violation of full coverage which consequently has to be minimized.

To guarantee that customers are assigned only to facility locations where a physical connection exists, the design of the ILP has to be slightly changed. Every assignment variable x_{cf} which

[21] A capacity constraint for the facilities quickly leads to a situation where problems become unfeasible, e.g. if k is chosen too small.

[22] For details see Chapter 1 Section 1.5.3.

[23] See the discussion at the end of Section 1.5.2 in Chapter 1

represents a physically impossible assignment has to be forced to be equal to zero or removed from the design.

An eventual violation of Rule 7 (Disjoint supply areas of ARUs, 1.4.7) is easily repaired, because the uncapacitated k-median problem is considered. Every customer should be assigned to the closest facility which poses no problem since capacities of facilities are unrestricted. The desired closeness can be controlled by the assignment cost parameters AC_{cf}. The choice of these parameters is a central issue of the application and the corresponding discussion is postponed to Section 2.5.1 and following.

2.3.3 Metric k-median problem

Assignment costs are a further and very interesting source of variation for the k-median problem. In the literature the cost of assigning a customer to a facility is usually modeled as the weighted distance between two locations, the customers weight $w(c)$ times the distance between the customer site and the supplying facility site $d(c, f)$. The following definition is taken from Shindler's survey of the metric k-median problem [48].

> *Given N, a set of points in some metric space, and some integer k, our goal is to select k points, $K \subseteq N$. Once the points are selected, the solution quality is the sum of all N's points' distances to their nearest element of K; a higher quality solution corresponds to a lower cost. The set of data points that have a given $m \in K$ as their closest is known as m's cluster. Point m is the cluster's center, or median.*

The medians are sometimes also called facilities or locations for facilities, and the points assigned to the medians are then addressed as customers or customer sites. The metaphor of customers and facilities suggests that the weighted distances are transportation costs where the weight gives the price for transport per unit of distance. Or the weight is a measure of demand associated with customers and the distance becomes some sort of penalty or an indicator of lost quality for satisfying the demand too far from the home base.

In the metric k-median problem distances are drawn from the Euclidian plane or space. Customers are situated at locations which are furnished with plane or spatial coordinates. Distances are calculated by means of the Euclidian metric, which implies nicer properties for the assignment costs. Weights are associated with customers. Usually customers and facilities are not differentiated and any customer location may become one for a facility ($N = M$). The metric KMP is more like a clustering algorithm: there are k clusters formed and a representative — a median — is chosen from each.

An airfreight example provides a demonstration of the metric k-median problem. An airfreight company seeks to reorganize their distribution system. From all the airports which the company is servicing three are chosen to serve as central distribution bases. In future all national traffic has to be conducted between the three central distribution bases. Local traffic is restricted to local airports and the central base which they are assigned to. Distances are defined by the flying distance between airports. The weights assigned to the airports may be defined by the average freight volume per year which comes in or goes out.

Lin and Vitter gave a polynomial time approximation algorithm in [39] with the additional restriction that the total sum of distances between customers and facilities does not exceed a

given upper bound D. The algorithm approximates the cost value of the optimal solution by $2(1 + 1/\epsilon)D$ and the desired value k of medians by $(1 + \epsilon)k$. In [38] the same authors gave a polynomial time approximation algorithm for the metric k-median problem (without upper bound on the total cost) with an approximation factor of $(1+\epsilon)$ for the cost and $(1+1/\epsilon)(\ln n+1)$ for the number of facilities.

A constant factor approximation algorithm which is exact on the number of opened facilities was given by Charikar et al. in [9]. Their method yields a $6\frac{2}{3}$-approximation. Later Arya et al. improved the result to a bound of $3 + \frac{2}{k}$ in [3] by means of a local search heuristic.

Another approach of solving the metric k-median problem is to tackle it by means of facility location algorithms. This basically implies the relaxation of the number of facility constraint 2.8, moving it to the objective function and imposing some cost on opening facilities. By varying this cost factor the desired number of open facilities can be approached. Jain and Vazirani provided such an attempt in [29]. Their algorithm achieves a 6-approximation with low running time of $\mathcal{O}(m \log m(L + \log n))$ where m and n represent the number of edges and vertices respectively and L is the number of bits needed to represent an assignment cost. Their approach combines the heuristic procedure concerning the opening cost described above with a primal-dual schema based on LP-relaxation.

A completely different approach was applied in [11] by Chrobak et al. The reverse greedy algorithm starts by placing facilities on every node, then removing one by one until the desired number of open facilities is reached. The approximation bound which is achieved is of order $\mathcal{O}(\log n)$. This reverse procedure resembles a little bit the technique which is applied and described in this thesis from Section 2.5.1 on.

2.3.4 Network graphs and the k-median problem

Distances may also be derived from graphs. In 1979 O. Kariv and S.L. Hakimi, the authors of [31], used the concept of networks for this purpose. In their paper they defined a network as a connected undirected graph with a nonnegative number associated with each vertex (vertex weight) and a positive number associated with each edge (edge length, distance of adjacent vertices). The distance concept is generalized in a natural way to arbitrary pairs of vertices by means of the shortest path between them. The length of a path is readily defined by the sum of the lengths of the edges the path runs along.

However, their formulation of the k-median problem differs slightly form those given above. The cited paper is actually the second part of a comprehensive study of algorithmic approaches to network location problems. In the first part, [30], p-centers of networks are defined and discussed. Among other things a p-center is a set of p points on the graph of the network. But, a point on a graph is not necessarily a vertex. It is allowed to be any point located on an edge of the graph. Consequently, the problem of finding a p-center is not a discrete optimization problem.

Secondly, a p-center minimizes the **maximum** of the weighted distance between all vertices of the graph and a set of p points for all sets of p points on the graph. It is this second property where the p-center definition departs from the definition of p-medians in the second part of the

paper [31]. A p-median minimizes the **sum** of weighted distances between all vertices of the graph and a set of p points for all sets of p points on the graph. Otherwise, the definition of p-center and p-median coincide which implies that a p-median is not necessarily a subset of the set of vertices of the network.

In the concepts of Kariv and Hakimi the crucial difference between centers and medians is the norm of what has to be minimized, the maximum distance or the average distance between vertices and a set of points. However, as it turns out and Hakimi had already proven in 1965 in his paper [25], there always exists a set of p vertices of the graph which also forms a p-median to solve the p-median problem.[24].

In his paper Hakimi gave an illustration and motivation for the concept of p-medians which was already announced in the title: "Optimum Distribution of Switching Centers in a Communication Network ...". The formulation of the example resembles the airfreight example[25] for the metric k-median problem in details. The communication network is the telephone interconnection system, the backbone of a telephone system which connects all the local loops with each other. Within this interconnection system switching centers (distribution bases) are to be established which carry out all "national"interconnecting traffic. "Local"interconnection traffic is kept between a switching center and the local loops which are assigned to it. The vertex weight represents the number of wires (lines) that must be connected between the vertex and its switching center. The length of an edge is the price for establishing a connection for one unit, i.e. one wire or line. The goal is to minimize the total length of wires which have to be connected to the switching centers. Surprisingly, the connection of the switching centers is considered to be negligible. The reason therefore is not given.

Hakimi gave a second application in his paper from 1965 to study the vertex covering problem: policing a highway network. An adaptation of this example illustrates the idea of the concept of p-center in contrast to the p-median problem. A system of highways is represented by a network. Vertices are intersections, exits, cities, or motorway stations. The connecting highways are the edges whose lengths coincide with the distance the highway spans between two locations. Since the goal is to distribute p policemen along the highways in such a way that any location may be in reach of the police by no more than a fixed and specified distance, vertex weights are not important. Therefore they are all set equal to 1. It is clear from the description that in this case the appropriate norm which has to be minimized is the maximum of the distances of all locations to their closest policemen.

These two examples applied to a simple network illustrate why in one case the desired location may be positioned in the middle of the edge (1-center), whereas in the other case the optimum is rather found at one of the vertices. This simple network consists of two vertices and the edge, which connects them. The policeman is best placed right in the middle along the highway

[24] Actually, that is the main result of the paper with respect to the k-median problem: it is reduced to a discrete optimization problem. For small networks and just a few medians the problem might be solved by brute force.

[25] The airfreight example may also be expressed in terms of a network by means of the complete graph connecting all involved airports. However, this shows that graphs are not only useful to derive assignment costs, but also to add additional structure to the problem. The graph expresses rules for locations, how they can be directly assigned to each other, and which locations need some intermediate points.

between the two cities to reach each of the cities equally well. The optimal location for the switching center is the location with higher weight, i.e. with more wires which have to be connected.

A definition of the k-median problem in general graphs (networks) which is different from the one given by Kariv and Hakimi is found in [10].

> *In the k-median problem we are given a graph G in which each node u has a non-negative weight w_u and each edge (u, v) has a non-negative length d_{uv}. Our goal is to find a set F of k vertices that minimizes* $\text{Cost}(F) = \sum_{x \in G} \min_{u \in F} w_x d_{xu}$. *We think of the nodes of G as customers, with each customer x having a demand w_x for some service. F is a set of k facilities that provide this service. It costs d_{xu} to provide a unit of service to customer at x from a facility located at u. We wish to place k facilities in G so that the overall service cost is minimized. This optimal set is called the k-median of the weighted graph G.*

In [31] Kariv and Hakimi also proved that the k-median problem is NP-hard on general networks. Lin and Vitter described several approximation algorithms for this and other problems in [39]. Thorup studied the shortest path metric, i.e. the metric on a graph, for the k-median, k-center and facility location problems on sparse graphs in [54] and derived approximation algorithms with linear running time in the number of edges.

An interesting alternative approach to solve k-median problems in network graphs is to approximate distances in the network by a tree graph and then solve the k-median problem in the approximation tree (see the following section). Bartal presented such an approximation technique in [4] and improved it late in [5].

2.3.5 Undirected trees and the k-median problem

The runtime complexity drastically improves for the special case of tree networks. Kariv and Hakimi studied such graphs in [31]. They gave a solution algorithm and estimated the runtime complexity by $\mathcal{O}(n^2 * k^2)$. Tamir undertook a careful analysis of the complexity bound and improved it to $\mathcal{O}(n^2 * k)$ in [53]. In addition he considered facility opening costs which is not typically done when studying k-median problems.

During the preprocessing of the dynamic programming algorithm which Tamir described in [53] for every node of the tree a sorted listed of the distances to all other nodes of the tree is produced. Then a "leaves to root"strategy is applied: i.e. starting at the leaves for every node of the tree several values of two types of cost functions are computed. Basically, the cost functions are evaluated for all integers between 0 and k — the maximum number of facilities which is desired — and with respect to all other nodes of the tree.

Benkoczi and Bhattacharya were able to improve the algorithm even further. The runtime could be brought down to sub-quadratic time. In [6] they applied a special decomposition technique to the tree — called spine decomposition (see also Section 2.3.11) — and showed that the worst case runtime is then $\mathcal{O}(n\ log^{k+2} n)$. A detailed discussion of the spine decomposition technique and its application to facility location problems can be found in the thesis of Benkoczi [8].

In principal, these algorithms can be used to solve the k-median problem in rooted, directed trees. However, because of the additional requirement for directed trees that customers may be served only by facilities which lie on the directed path which leads from the customer to the root there is some hope that faster algorithms exist. For example, there is no need to compute the distances between every pair of nodes as it is done during the preprocessing in [53]. This may not improve the runtime for the worst case, but it may improve the average runtime and the runtime in practice respectively. Therefore, it makes sense to investigate the case of k-median problems on rooted, directed trees separately.

2.3.6 Directed trees and the k-median problem

In [10] Li, Golin, Italiano, Deng, and Sohraby are cited to have introduced the k-median problem for rooted and directed trees in their paper [37] from 1999 "as a mathematical model for optimizing the placement of web proxies to minimize average latency". In fact, in [37] the term k-median is not mentioned at all, nor are there any references to preceding works concerning k-medians found. However, they introduced a concept which eligibly could carry this term and gave an algorithm to solve it. One year later the first three of the mentioned authors together with Vigneron and Gao published a paper [56] where they improved this algorithm and called the problem by its proper name. Moreover, they established a connection to the algorithm which Kariv and Hakimi had given in [31] for undirected graphs.

The k-median concept for undirected graphs is transferred to rooted directed trees nearly without alteration. Only the specific role of the root is reflected by the fact that the root is always part of the set of medians. So, a k-median is a set of k vertices always containing the root of the tree. Unlike in the undirected or metric case the 1-median problem is trivial on directed trees. The only effort results from calculating the value of the cost function.

Furthermore, customers may only be assigned to facilities where a directed path leads from the customer site to the facility. The direction in these trees is understood in the following way: the root of the tree is the forefather of all nodes, and the edges are directed from children towards their parents.

Later the definition of k-medians for directed trees changed slightly, probably to simplify the formulation of algorithms. In [10] the definition becomes:

> *We consider the k-median problem for rooted directed trees. Let T be a directed tree with root r. If y is the father of x, denote by d_{xy} the length of edge (x, y). The length function extends in a natural way to any pair x, y where y is an ancestor of x. As in the undirected case, each node x is assigned a non-negative weight w_x. Our goal is to find a set of nodes F of size k that minimizes the quantity:*
>
> $$\mathrm{Cost}(F) = \sum_{x \in T} \min_{u \in F + r} w_x d_{xu},$$

where $F + r$ stands for $F \cup \{r\}$.

The root is not called a median anymore. The root, however, stays part of the solution, i.e. the cost function is evaluated with respect to the k vertices of the k-median set and additionally

the root. There are k medians but $k+1$ facilities considered. The 0-median problem is the trivial analogue for the 1-median problem of the previous definition.

In [37] the authors gave a detailed account of the application for which they suggested the k-median method as the appropriate solution: Web caching in the internet by means of web proxies. A (web -) clients' request is passed through the internet to the addressed server. The longer this path is and the more clients address the same server with similar requests, the higher the risk of internet congestion becomes which leads to increased response times for the clients, i.e. increased latency.

On the way through the internet a request encounters other servers. If one or some of these servers would keep copies of the requested documents, the request could be answered earlier, response times could be reduced and the risk for internet congestion could be minimized. This is the key idea of web proxies. The placement of the proxy servers has to be arranged with respect to the server whose information they are supposed to mirror. The clients' requests take the shortest path through the internet to the server. This way the structure of a directed tree with the server as its root and clients somewhere in the periphery[26] is generated. According to the authors the weight of vertices represents "the traffic traversing this node"and the distance between two vertices "can be interpreted as either latency, link cost, hop count and etc."The distance function is generalized by summing the lengths of edges of a path according to the concept already given above.

SARU provides another example for an application of the k-median problem in rooted directed trees: the determination of copper centers in a copper access network. The definition of k-median will be understood according to [10]; i.e. the root of the tree always carries an open facility, but it is not counted as a median. The terminology reflects the fact that in SARU facilities are called access remote units where the term "remote"implies that facilities are situated aside the central office[27].

Since in general it cannot be expected that the trail graph of the copper network can be described as a tree, some inspection, modeling and alteration of the trail graph and individual supply paths will be necessary to prepare the network for k-median algorithms. The conflicts of the copper network with the tree concept are demonstrated in detail in Chapter 1 Section 1.3.

In any case, the central office is the root. Weights are defined by the demand of the customers, e.g. the weights count the number of wires which have to be connected[28]. Distance is derived from cable lengths or alternatively form transmission rates or damping[29]. However, the distance function will give reason for new variations of this problem which are motivated by the following question: How can transmission rates be expressed as sums of edge weights?

[26] In fact, all leaves of this tree must be clients.
[27] See Chapter 1 Rule 8 (Limitation of ARU capacities, 1.4.8) for a detailed discussion.
[28] Compare also Chapter 1 Rule 3 (Customer demand, 1.4.3).
[29] Compare Chapter 1 Rule 10 (Distance rule, 1.4.10).

2.3.7 Description of the basic algorithm

The algorithm which solves the k-median problem in directed trees as described by Chrobak, Larmore and Rytter in [10] is of a recursive nature. To find the k medians in a directed and rooted tree T_r with root r basically for every subtree all j-medians for any value of j between 0 and k have to be determined.

0-median
It is not difficult to calculate the cost of the 0-median solution of the original tree T_r. The only open facility is found at the root r. The cost is given by the sum of the weighted distances of all customers to the root.

But, it turns out that in order to compute the KMP for arbitrary values of k, the cost for the 0-median has to be computed for every subtree T_x of the original tree. T_x is the subtree of T_r which contains all descendants of x including x itself. For the 0-median solution of T_x it is assumed that one facility is opened at x. The calculation of the cost is equivalent to the procedure given above.

1-median
To find the 1-median solution the algorithm decomposes the tree into proper subtrees. For simplicity and without loss of generality the tree is assumed to be binary: any vertex which is not a leaf is of in-degree ≤ 2. A detailed description of how to produce a binary tree is given, for example, by Tamir in [53][30]. Hence, the root r has got two children d and s. Together with the root they define two subtrees, T_d^r and T_s^r, by adding the edge (x, r) to the tree T_x leaving r to be the root of the tree T_x^r. The intersection of T_d^r and T_s^r contains only the root r, and their union is the original tree again.

Horizontal decomposition step
Certainly, the only median of the 1-median solution of T_r lies either inside of T_d^r or inside of T_s^r but not in both, since the only facility which can be situated in both trees resides at r. But, by definition this facility is not called a median. More important, this median solves the 1-median problem for the subtree in which it is located. This is due to the fact that the tree is directed and its root separates the two subtrees with a facility which is always open. Moving facilities around in one of the trees does not effect the provisioning of customers in the other tree.

Conversely, the 0-median and 1-median solutions of the trees T_d^r and T_s^r determine the 1-median solution of the original tree T_r. Either the median lies in the daughter tree, then the son tree cannot contain a median, or the roles of the two trees are interchanged. The assignment costs will give the clue which one induces the 1-median solution for the entire tree.

The expression $\mathrm{Cost}_x^j(r)$ denotes the cost of the j-median solution in the subtree T_x^r with root r. The minimum

$$\min\{\mathrm{Cost}_d^1(r) + \mathrm{Cost}_s^0(r); \mathrm{Cost}_d^0(r) + \mathrm{Cost}_s^1(r)\}$$

is the assignment cost $\mathrm{Cost}_r^1(r)$ for the 1-median solution of the tree T_r and its minimizers determine the solution set.

By the following definition the cost function is generalized for later use.

[30]The author gave $2n - 3$ as an upper bound for the number of nodes in the derived binary tree where n is the number of nodes in the original tree.

Definition 12 (Cost function) *For a directed tree T_r with root r and a subset V of its vertices* $\mathrm{Cost}(\mathbf{T_r}, \mathbf{V})$ *denotes the sum of weighted distances between all nodes of the tree and the set $V \cup \{r\}$, where the weighted distance between a node and a set is given by the minimum of the distances between this node and all nodes of the set:*

$$\mathrm{Cost}(T_r, V) = \sum_{c \in vert(T_r)} \min_{v \in V \cup \{r\}} w(c) * d(c, v).$$

Furthermore, for any positive integer j the term $\mathbf{M^j(T_r)}$ *denotes a set of j vertices in T_r which solve the j-median problem in T_r, and* $\mathbf{Cost^j(T_r)} := \mathrm{Cost}(T_r, M^j(T_r))$. *Consequently,* $\mathrm{Cost}_x^j(r) = \mathrm{Cost}^j(T_x^r)$.

This constitutes the first decomposition step: a horizontal move along the offspring of a vertex. To get the recursion in momentum a second, a vertical decomposition step is necessary, a decomposition of trees of type T_x^r where the root r is of in-degree 1.

Vertical Decomposition Step
There are two cases: The median of the 1-median solution of T_x^r is either located at vertex x or not.

If x is the median, the problem is solved. The median is found. The cost of the 0-median problem in T_x equals the assignment cost $\mathrm{Cost}_x^1(r)$ of the 1-median solution in T_x^r. In the general case of the quest for j medians the task becomes one of finding $(j-1)$ medians in the tree T_x with one open and immobile facility situated at its root quite alike the original problem.

In the second case — the optimal solution does not situate a facility at x — vertex x becomes redundant. It will not carry an open facility. It may be removed from the graph. This situation is best expressed by the union of the two trees $T_{d_x}^r$ and $T_{s_x}^r$

$$T_{d_x}^r \cup T_{s_x}^r.$$

This graph is constructed from the graph T_x^r by removing x and all its incident edges, and connecting the offspring — d_x and s_x — of x to r by edges whose lengths correspond to the lengths of the paths which connect the offspring with r in the original tree.

In the second case the 1-median solution of $T_{d_x}^r \cup T_{s_x}^r$ is the 1-median solution for T_x^r. The costs of the two solutions, however, are not necessarily equal. Particularly, if vertex x is of positive weight, i.e. it is a customer vertex, the assignment cost of x is not considered in the cost for the 1-median solution of $T_{d_x}^r \cup T_{s_x}^r$. But, this is easily adjusted.

The actual solution for the 1-median problem in the tree T_x^r is again found by comparing the cost values of two alternative assignments:

$$\min\{\mathrm{Cost}_x^0(x); \mathrm{Cost}^1(T_{d_x}^r \cup T_{s_x}^r) + w(x) * d(x, r)\}.$$

Recursion
After this second decomposition the algorithm arrives at trees for which the respective median problems are either already solved or are of the same type as in the original tree, i.e. the trees

are binary with an open and immobile facility at their roots. The decomposition may start over again by applying horizontal and vertical steps to the newly constructed trees. The recursion must end, because each vertical decomposition step produces trees whose height is reduced by at least one edge compared to the trees they are derived from.

At some point during the recursion vertex x will become a leaf. For leaves x and trees of the type T_x^r any j-median is easily constructed and the assignment cost computed.

j-medians
The key to generalize the stated procedure to arbitrary values of j lies in the horizontal decomposition step. The j medians of an optimal assignment in the tree T_r are separable into two subsets, one contained in the daughter's subtree of T_r and the other in the son's. The two sets — the daughter's is of size p and $q = j - p$ is the size of the son's — are disjoint, since the two trees T_d and T_s are disjoint. And again, because of the optimality of the j medians in T_r, their subsets must be optimally situated in the trees T_d^r and T_s^r. Any alteration of the location of facilities in the subtrees T_x^r results in an equivalent change of the value of the cost function in the original tree T_r.

Conversely, the solution for the j-median problem in T_r has to be found among the $j + 1$ combinations of p-median solutions for the subtrees T_d^r and T_s^r for the integer variable p ranging from 0 to j. The minimizer of the assignment costs of the $j + 1$ combinations will again give the clue of the j-median in T_r:

$$\mathrm{Cost}_r^j(r) = \min_{p+q=j} \{\mathrm{Cost}_d^p(r) + \mathrm{Cost}_s^q(r)\}.$$

2.3.8 General recursion formula

M. Chrobak, L. Larmore and W. Rytter gave a very compact formulation of the general recursion formula in [10].

$$\mathrm{Cost}_x^j(u) = \min \begin{cases} \min\limits_{p+q=j} \left\{\mathrm{Cost}_d^p(u) + \mathrm{Cost}_s^q(u)\right\} + w(x) * d(x,u), \\ \mathrm{Cost}_x^{j-1}(x) \end{cases} \tag{2.11}$$

x is any vertex in the network, u is an ancestor of x; d and s are the offspring of x, if it has got any.

If x is a leaf, then $\mathrm{Cost}_x^0(u) = w(x) * d(x,u)$ and $\mathrm{Cost}_x^j(u) = 0$ for $j \geq 1$. In any case the recursion ends at this point.

If $j = 0$, then $\mathrm{Cost}_x^{j-1}(x)$ is formally set to infinity, so that the calculation of $\mathrm{Cost}_x^0(u)$ reduces to the calculation of the cost for the 0-median solutions in the trees of the two children of x. However, $\mathrm{Cost}_x^0(u)$ can also be calculated directly by adding the weighted distances of all customers contained in the tree T_x to vertex u. This is the seconde case in which the recursion ends.

In Recursion Formula (2.11) horizontal and vertical decomposition are performed simultaneously. The algorithm makes use of subtrees of the type T_x^u which is the tree T_x together with

an additional edge which connects x with one of its ancestors $u \in T_r$. The length of this edge equals the length of the path from x to u in the original tree.

The root u may be chosen to be equal to x. In this case the second line in the recursion formula, which is $\mathrm{Cost}_x^{j-1}(x)$, becomes redundant. The $(j-1)$-median solution will certainly be at least as expensive as the j-median solution[31]. In fact, they will be of equal cost if and only if the assignment cost of the $(j-1)$-median solution is zero[32]. Hence, if the cost of the $(j-1)$-median solution is positive, it will be higher than the cost for the j-median solution in T_x and the recursion formula actually reduces to

Proposition 2 (Cost P1)

$$\mathrm{Cost}_x^j(x) = \min_{s+t=j} \left\{ \mathrm{Cost}_d^s(x) + \mathrm{Cost}_s^t(x) \right\}. \tag{2.12}$$

However, this does not invalidate the recursion formula.

The two properties of the cost function which were used have to be stated and proved, too.

Proposition 3 (Cost P2 and P3) *For any vertex x of T_r, any of its ancestor u and any non negative integer j*

$$\mathrm{Cost}_x^j(u) \geq \mathrm{Cost}_x^{j+1}(u) \tag{2.13}$$

and

$$\mathrm{Cost}_x^j(u) > \mathrm{Cost}_x^{j+1}(u) \tag{2.14}$$

if and only if

$$\mathrm{Cost}_x^j(u) > 0.$$

Proof:
(1) $M^j(T_x^u)$ denotes a set of j medians in the tree T_x^u. If there is a vertex not occupied by a facility yet, it shall be denoted by v. Then

$$\begin{aligned} \mathrm{Cost}_x^j(u) &= \mathrm{Cost}(T_x^u, M^j(T_x^u)) \\ &\geq \mathrm{Cost}(T_x^u, M^j(T_x^u) \cup \{v\}) \\ &\geq \mathrm{Cost}(T_x^u, M^{j+1}(T_x^u)) \\ &= \mathrm{Cost}_x^{j+1}(u). \end{aligned} \tag{2.15}$$

If all vertices in T_x^u are occupied by facilities, then the assignment cost $\mathrm{Cost}_x^j(u) = 0$ and the investigation ends there.

(2) Since assignment cost are sums of products of non-negative weights and distances, they are non negative, too. Hence,

$$\mathrm{Cost}_x^{j+1}(u) \geq 0$$

[31] See Proposition 3 below.
[32] See Proposition 3 below.

for any non-negative integer j. It follows from Inequality (2.14) that

$$0 \leq \text{Cost}_x^{j+1}(u) < \text{Cost}_x^j(u).$$

Conversely, $\text{Cost}_x^j(u) > 0$ implies that there is at least one customer v in T_x^u with positive distance to its supplying facility. To open a facility at this customer site reduces the cost at least by the product of the customers' weight and its distance to the facility and produces a situation with $j+1$ open facilities. So, in Inequality (2.15) strict inequality holds.
□

2.3.9 Iterative version

An iterative version of the algorithm can be derived from the recursion formula, too. For this purpose it has to be determined for which triple of vertex, its ancestor and the number of medians (x, u, j) the j-median problem has to be solved and in what order they have to be computed.

The recursion sets out to find the value of $\text{Cost}_r^k(r)$. The first application of the recursion formula results in queries for the cost of j-median solutions for every j between 0 and k and for both children of r:

$$\text{Cost}_c^j(r) \text{ for } 0 \leq j \leq k \text{ and } c \in \{d_r, s_r\}.$$

A second application of the recursion formula results in similar queries for the grandchildren of the root. Since any vertex u is some descendant of the root, the repeated application of the recursion formula will eventually make it necessary to determine the j-median solution in the tree T_u^r for any j and especially for j equal to k:

$$\text{Cost}_u^k(r).$$

To proceed in the recursion among other things the $(k-1)$-median problem has to be solved in the tree T_u. This tree contains vertex x, since u is an ancestor of x. By the very same argument as before the recursion eventually will ask for all the j-median solutions in T_x^u for j between 0 and $k-1$:

$$\text{Cost}_x^j(u) \quad 0 \leq j \leq k-1.$$

This is true for $u \neq r$. If u is the root of the tree, then j has to be chosen up to k.

Algorithm 3 (KMP by ascending iteration)

Preprocessing	P1	The vertices of the tree T_r are sorted and enumerated according to the height of the subtree whose root they are (= their level).
Main	**X-Loop**	x runs through all vertices of the graph according to their level.
	U-Loop	u runs through all ancestors of x starting at x and stopping at r. So u traverses the directed path which connects x to r.
	J-Loop	j is chosen between 0 and $k-1$, or k in case $u = r$, always starting at 0.
	Minimum Loop	According to Recursion Formula (2.11) $j+1$ pairs of p and q-median solutions with $p+q=j$ have to be examined to determine the optimal combination.
	Computation	$\mathrm{Cost}_x^j(u)$ is calculated according to Recursion Formula (2.11).

For the iterative version of the KMP algorithm the vertices x are sorted in ascending order according to the height of the subtree T_x whose root they are. Hence, first in line are the leaves of the tree, and the sequence is closed by root r. Four loops are necessary to run the algorithm.

X-Loop and U-Loop may be placed in reversed order. The values for x are then chosen from all vertices in T_u. By making these choices some care has to be taken. A vertex from T_u may be chosen as the next value for x if and only if all its descendants have already been chosen.

The Minimum Loop depends on the values of x, u and j. Therefore it hast to be the innermost loop.

The J-Loop may be freely positioned with respect to the X- and U-Loops: before, in between or after. However, to calculate the cost of the j-median with respect to x and u in the tree T_x^u, only information involving u and the offspring of x is utilized. Once all necessary j-medians with respect to x and u are determined, the information

$$\mathrm{Cost}_{d_x}^p(u) \text{ and } \mathrm{Cost}_{s_x}^q(u) \text{ for } 0 \leq p,\ q \leq k-1 \text{ or } k$$

is not needed anymore. To keep memory complexity low it is therefore advisable to place the J - Loop as the innermost loop of the three and delete this information after moving on to a new value for u or x.

2.3.10 Time complexity

The description of the iterative algorithm displays on first sight that the worst time complexity of the KMP algorithm is at most $\mathcal{O}(n^2 * k^2)$. The graph may be a path. Then the U-Loop basically runs through all the vertices of the graph. This can be improved by careful analysis as it was done in [10] and [53]. The bound for the worst time complexity can be reduced to $\mathcal{O}(n^2 * k)$.

However, for realistic trees which are derived from local loop networks this worst case estimate is certainly too pessimistic. In [56] the authors estimated the worst case time complexity by $\mathcal{O}(P * k^2)$ and the space complexity by $\mathcal{O}(n * k)$. P is the path length of the tree which is the sum of the lengths of all shortest paths connecting vertices to the root. The length of a path is measured by the number of edges contained in it[33]. Or in other words, it is the sum of the number of all ancestors for all vertices of the tree. Clearly, this is exactly the number of combined iterations of the X- and the U-Loop.

2.3.11 Tree decomposition strategies

The following two sections provide an overview of some variants of the basic k-median algorithm for directed trees. The first set of variants are produced by different strategies of decomposing the underlying tree.

In [56] A. Vigneron et al. used a post-order traversal of an arbitrary directed tree, i.e. not necessarily a binary tree, to enumerate the vertices for the decomposition procedure. The numeration starts with an arbitrary leaf[34].

Algorithm 4 (Post-order traversal)

Initializing	Step 0.1	One of the leaves is chosen as the first vertex of the ordering.
	Step 0.2	The vertex last visited (VLV) is initialized with the first vertex.
Main	Loop 1	Runs as long as there exist vertices not visited yet.
	Step 1.1	The parent of VLV is determined.
	Step 1.2	Are all the children of the parent enumerated?
	Yes	
	Step 1.3	The parent of VLV is enumerated with the next number.
	Step 1.4	VLV is set to the vertex just enumerated.
	No	
	Step 1.5	A vertex which was not visited yet and which is furthest away from the parent of VLV is determined,
	Step 1.6	and labeled with the next number.
	Step 1.7	VLV is set to the vertex just enumerated.

At the beginning of the iteration step the parent of the vertex which was enumerated during the previous step is visited. If all children of the parent are already enumerated, then the vertex itself receives its number and the iteration starts over again. Otherwise, the vertex furthest away from the parent is chosen and labeled with the next number. The iteration starts over again. The algorithm stops with the root of the original tree receiving the highest number. Alternative descriptions for a post-order traversal can be found in [1] or [14][35].

[33] For a balanced binary tree with n nodes $P = \Theta(n \log n)$, for random general trees $P = \Theta(n\sqrt{n})$ and in the worst case $P = \Theta(n^2)$. See [56].

[34] It may start with the leaf with greatest distance to the root.

[35] Dale and Lilly described a recursive procedure for a binary tree: considering a non-leaf, first visit the subtree of the left child, then the subtree of the right child, then the node itself.

Ordering and labeling the vertices of a tree like that enables to describe any subtree as a closed interval of integers. If $m(x)$ denotes the lowest numbered vertex in the tree T_x, then the set of vertices of this tree is given by the set of integers between $m(x)$ and x: $[m(x), x]$.

For x together with one of its ancestors u a decomposition of the ancestral tree T_u is conducted resulting in three different sub graphs whose vertex sets can be determined as

$$\text{vert}(T_x) = [m(x), x]\,,\ \text{vert}(L_{u,x}) = [m(u), m(x))\ \text{ and } \text{vert}(R_{u,x}) = (x, u]\,.$$

Round brackets indicate (semi-) open intervals. Since x is a descendant of u, T_x is a subtree of T_u. $R_{u,x}$ turns out to be a tree, too, whereas $L_{u,x}$ is in general a disconnected sub graph of T_u.

If the tree T_u is drawn from left to right starting with the lowest numbered vertex at the left side, the tree $R_{u,x}$ lies to the right of T_x and the graph $L_{u,x}$ to its left.

Cost functions are associated with any of the three sub graphs. Their values are calculated by means of recursion formulas again. The solution of the KMP is found by iterating u through all vertices of the tree according to their post-order enumeration.

Another decomposition strategy is found in [7] (and also in [6], a very detailed discussion can be found in [8].). The authors (Benkoczi, Bhattacharya, Chrobak, Larmore, Rytter) applied a spine-decomposition to a binary tree. A spine-decompositions requires labeling the offspring of every vertex as left and right descendant. Additionally, in [7] the right offspring is chosen such that its subtree is at least as large as the left subtree.

A spine of a tree is the longest path in the tree containing the root and all its right descendants.

The decomposition of the spine works similar to a binary search. An appropriate vertex somewhere in the "middle"of the spine is chosen[36]. The spine splits into two disjoint segments. Certain sub graphs are associated with each of the segments. The two segments are split again with more sub graphs being associated with each resulting sub segment. This procedure continues, until a sub segment consists only of a single vertex anymore. In this case, the decomposition moves to the left descendant of that vertex. The left descendant features a subtree which contains another spine. The decomposition continues until leaves are reached.

The succession of horizontal and vertical decomposition steps is interchanged.

The KMP is solved by associating certain cost functions with the sub graphs which are generated during the deconstruction of the graph. The cost functions are calculated recursively following the spine decomposition. Moreover, the authors of [7] additionally applied a technique which was also presented in [10] and given the name depth-based algorithm. It is also called undiscretized dynamic programming and presented, for example, in [47] to solve facility location problems in trees (Compare Chapter 1 Section 1.5.2 Definition UFLP).

[36] This choice is not arbitrary. For details see [7].

2.3.12 Ancestor-based versus depth-based algorithm

One of the key features of the KMP algorithms described so far is that for every vertex x and all of its ancestors u several j-median problems have to be solved and the corresponding cost values have to be stored. This characterizes ancestor-based algorithms. Depth-based algorithms omit the explicit calculation of assignment costs depending on the ancestors. Instead, the assignment cost is expressed as a piecewise linear function in one variable which specifies the distance between vertex x and the closest ancestor where a facility is situated.

This concept and its potential advantage are best illustrated with 0-median solutions. Ancestor-based algorithms calculate the value for the 0-median solutions for any vertex x and every one of its ancestors u as the sum of the weighted distances of all the customers contained in the tree T_x and u. For a vertex x this leads to as many values as there are ancestors of x. So, in total P — the path length of the tree — values have to be calculated and stored.

For a leaf, however, all these numbers can be compressed into one single value. Following the notation in [10] — which abuses the notion of the already introduced cost function slightly — the cost function $\mathrm{Cost}_x^j(\alpha)$ becomes a function in one variable α. This variable represents the distance between x and the closest open facility situated at an ancestor u of x. So, $\mathrm{Cost}_x^j(u)$ becomes now $\mathrm{Cost}_x^j(d(x,u))$. Especially, for leaves the only meaningful cost function has got the form

$$\mathrm{Cost}_x^0(\alpha) = w(x) * \alpha$$

where $w(x)$ is the weight of the leaf and α is a variable which may be replaced by the distance $d(x,u)$ between x and any ancestor u.

For a vertex whose offspring are leaves, i.e. whose tree T_x is of height 1, the corresponding function is easily calculated using the analogous functions of its children:

$$\mathrm{Cost}_x^0(\alpha) = \mathrm{Cost}_d^0(\alpha + d(d,x)) + \mathrm{Cost}_s^0(\alpha + d(s,x)) + w(x) * \alpha.$$

Resolving the right hand side shows that the resulting function is linear:

$$\mathrm{Cost}_x^0(\alpha) = [w(x) + w(d) + w(s)] * \alpha + [w(d) * d(d,x) + w(s) * d(s,x)]\,.$$

Since substituting variables of linear functions by linear functions and adding linear functions always produce linear functions again, the assignment cost for 0-median solutions can be expressed by a linear function for every vertex:

$$\mathrm{Cost}_x^0(\alpha) = a(x) * \alpha + b(x).$$

The coefficients of this linear function can be nicely interpreted. The intercept $b(x)$ gives the assignment cost, if all customers in T_x are assigned to a facility at x. The slope $a(x)$ is the sum of the weights of all customers in the tree T_x. So $a(x) * \alpha$ determines the additional assignment cost, if no facility is situated at x and the distance to the next facility is α.

The extension of this concept to the 1-median problem becomes more complicated. The 1-median cost function for leaves is still trivial. It is 0 and hence again a linear function. The same is true for leaves and any further value of j. However, the calculation of the 1-median

cost function for arbitrary vertices asks for another operation: forming the minimum of linear functions which leads to piecewise linear functions.

The set of functions in which these operations take place is the set of lower envelopes of a finite number of linear functions. This set is closed with respect to the following operations:

1) Adding a lower envelope and a linear function
2) Substituting the variable of a lower envelope by a linear function
3) Forming the minimum of two lower envelops.

Therefore, this concept can be generalized to determine assignment cost functions for any vertex x and any j-median problem. The minima which have to be formed in this course depend on the two children of the vertex and the value of j. Hence, the number of operands is at most $k + 2$.

However, the efforts are thereby redirected from the ancestors to calculating and maintaining vertices of polygons. Moreover, as the distance between x and the global root r decreases, i.e. the more the number of ancestors of x decreases, the more complex the lower envelops potentially become. Nonetheless, the authors of [10] expressed their belief that the depth-based version is faster than the ancestral one.

For SARU the concept of depth-based algorithms is not useful, because two of the three possible distance measures — transmission rate and damping — are not linear.

2.4 Theory: Properties and improvements

Since the final goal during project SARU was to implement and apply the k-median concept, the properties of the algorithm were studied and possible improvements were sought. The key findings are collected and presented in the following sections.

2.4.1 Algorithmic Improvements

1) As soon as $\text{Cost}_x^j(u)$ is equal to zero, it is dispensable to proceed the iteration for higher values of j, since a value of 0 can not be underbid anymore. Leaves provide a typical example. One median in the trivial one-vertex-tree T_x guarantees full supply. In general the point of zero cost will be reached as soon as there remain fewer vertices of positive weight, i.e. customers, in a tree than j accounts for.

2) A second improvement is ascribed to the following property. For a fixed vertex x and fixed value of j the ancestor vertex u runs through the path which connects x to the overall root r. Then the assignment costs $\text{Cost}_x^j(u)$ rise until eventually a vertex u_0 is reached, such that no customer is assigned to u_0 in the j-median solution of $T_x^{u_0}$. The assignment costs for all ancestors of u_0 are identical, and the solution of the j-median problem in the respective trees is always the same.

If such a vertex u_0 exists for given x and j, then the U-Loop terminates prematurely.

A typical reason why no customers may be assigned to the root u of a tree T_x^u can be determined from the Recursion Formula (2.11). If the value of the first line of the recursion formula exceeds the value of the second line, a facility is situated at vertex x. Consequently, no customer can be assigned to the root anymore.

This is not the only possibility that such a vertex u_0 may exist. Facilities could be located at the children of x leading to the same consequences without a facility at x. Should this be the case, then the value of the first part of the recursion formula has not exceeded the second part.

To prove the consequences of the existence of such a vertex u_0 it is first shown that

Proposition 4 (Cost P4) *For any two ancestors u_1 and u_2 of x:*

If u_2 is an ancestor of u_1, then

$$\text{Cost}_x^j(u_1) \leq \text{Cost}_x^j(u_2) \quad \forall j \in \mathbb{N}_0 \tag{2.16}$$

Proof:
The set $\mathcal{M} := M^j(T_x^{u_2})$ of j medians which minimizes $\text{Cost}_x^j(u_2)$ provides also a set of facilities in the tree $T_x^{u_1}$, which together with a facility at root u_1 supply all customers in this tree:

- All customers previously assigned to u_2 are now assigned to u_1 which is possible, since the supply paths of all customers which lie in the tree T_x contain the subpath from x over u_1 to u_2.

- All other customers stay assigned to the same facility as before.

The cost of this assignment decreases by the length of the path from u_1 to u_2 times the total sum of the weights of all customers who are assigned to u_2. But, for later use it is better to express this fact slightly more cumbersome:

$$\mathrm{Cost}_{\mathcal{M}} + \sum_{c \in C^j(u_2)} w(c) * [d(c,u_2) - d(c,u_1)] = \mathrm{Cost}_x^j(u_2)$$

where $\mathrm{Cost}_{\mathcal{M}}$ denotes the cost of the assignment in $T_x^{u_1}$ as it was described before and $C^j(u_2)$ comprises all customers of T_x which are assigned to u_2 in the j-median solution of $T_x^{u_2}$. It follows that

$$\mathrm{Cost}_{\mathcal{M}} \leq \mathrm{Cost}_x^j(u_2)$$

if the expression

$$\sum_{c \in C^j(u_2)} w(c) * [d(c,u_2) - d(c,u_1)] \tag{2.17}$$

is non negative which is due to the simple but important observation that $d(c,u_1) \leq d(c,u_2)$ for any triple of vertices such that u_2 is an ancestor of u_1 and u_1 an ancestor of c, since in trees distances are defined as path-lengths.[37]

Of course,

$$\mathrm{Cost}_x^j(u_1) \leq \mathrm{Cost}_{\mathcal{M}}$$

since the j-median in $T_x^{u_1}$ minimizes the assignment cost over all sets of j facilities. In case $x = u_1$ and $x \in \mathcal{M}$, i.e. x is a member of the j-median of $T_x^{u_2}$, one facility is lost and $\mathrm{Cost}_{\mathcal{M}}$ reflects the assignment of customers to a total of j facilities including the root. But, an additional facility for the j-median solution in T_x may only bring costs further down.
□

Proposition 5 (Cost P5) *If an ancestor u of x exists, such that no customer is assigned to u through the j-median solution in T_x^u, then this property is true for all ancestors of u, their cost is equal to $\mathrm{Cost}_x^j(u)$ and the j-median solutions are identical. The nearest ancestor of x with this property is denoted by u_0^j. Then*

$$\mathrm{Cost}_x^j(u) = \mathrm{Cost}_x^j(u_0^j) \tag{2.18}$$

for any ancestor u of u_0^j.

Proof:
The j medians of the j-median solution in $T_x^{u_0^j}$ constitute a feasible solution for the assignment problem for all ancestors u of u_0^j, since all customers contained in T_x are assigned to one of them. Again, if u_0^j itself is a customer, it has no impact on the assignment nor on the cost, since u_0^j is always assigned to itself for zero cost.

[37] For the general facility location problem this property does not hold.

The assignment costs do not change, and since according to Proposition 4 the assignment cost of optimal solutions are increasing for subsequent vertices u, the assignment derived from the j-median solution in $T_x^{u_0^j}$ is optimal for all T_x^u.
□

Propositions 4 and 5 are combined in the statement

$$\mathrm{Cost}_x^{j_1}(u_1) \leq \mathrm{Cost}_x^{j_2}(u_2)$$

if and only if

$$j_1 > j_2 \text{ or } u_2 \text{ is an ancestor of } u_1, \text{ if } j_1 = j_2.$$

3) A third improvement is achieved by eliminating the Minimum Loop under certain circumstances. To see how this is done, the minimum problem solved during the Minimum Loop is transformed into a maximum problem and the differences of the j-median cost-function for consecutive values of j have to be studied.

2.4.2 Inspection of consecutive assignment costs

During the Minimum Loop the j-median solution for the tree $T_d^u \cup T_s^u$ is determined:

$$\mathrm{Cost}^j(T_d^u \cup T_s^u) := \min_{p+q=j} \left\{ \mathrm{Cost}_d^p(u) + \mathrm{Cost}_s^q(u) \right\}.$$

By multiplying the objective function by -1 a minimum problem is transformed into a maximum problem:

$$-\mathrm{Cost}^j(T_d^u \cup T_s^u) = \max_{p+q=j} \left\{ -\mathrm{Cost}_d^p(u) - \mathrm{Cost}_s^q(u) \right\}.$$

Adding a constant effects only the value of the objective, but not its maximizer:

$$\begin{aligned} &\mathrm{Cost}^0(T_d^u \cup T_s^u) - \mathrm{Cost}^j(T_d^u \cup T_s^u) = \\ &= \max_{p+q=j} \left\{ \mathrm{Cost}^0(T_d^u \cup T_s^u) - \mathrm{Cost}_d^p(u) - \mathrm{Cost}_s^q(u) \right\}. \end{aligned}$$

It is a simple reflection that $\mathrm{Cost}^0(T_d^u \cup T_s^u) = \mathrm{Cost}_d^0(u) + \mathrm{Cost}_s^0(u)$, and therefore the objective function can be rewritten as:

$$\max_{p+q=j} \left\{ \mathrm{Cost}_d^0(u) - \mathrm{Cost}_d^p(u) + \mathrm{Cost}_s^0(u) - \mathrm{Cost}_s^q(u) \right\}.$$

Adding and subtracting the values of all cost functions $\mathrm{Cost}_d^{j_1}(u)$ and $\mathrm{Cost}_s^{j_2}(u)$ for all values of j_1 and j_2 between 1 and p and q respectively, effects only the appearance of the formula and not the solution:

$$\max_{p+q=j} \left\{ \sum_{j_1=1}^{p} \left(\mathrm{Cost}_d^{j_1-1}(u) - \mathrm{Cost}_d^{j_1}(u) \right) + \sum_{j_2=1}^{q} \left(\mathrm{Cost}_s^{j_2-1}(u) - \mathrm{Cost}_s^{j_2}(u) \right) \right\}.$$

Suitably, the first differences of the cost function are defined and denoted by

$$\Delta(x,u,j) := \mathrm{Cost}_x^{j-1}(u) - \mathrm{Cost}_x^{j}(u)$$

for positive j. Because of Proposition 3 the first differences are non-negative. They can be interpreted as the savings which are gained, if an additional facility is used in the tree T_x^u to solve the j-median problem there. The maximization problem based on the first differences can be interpreted as finding the combination of p facilities in T_d and q facilities in T_s which maximizes the savings compared to the solution without any facilities. Now the maximization formula reads as follows

$$\max_{p+q=j} \left\{ \sum_{j_1=1}^{p} \Delta(d,u,j_1) + \sum_{j_2=1}^{q} \Delta(s,u,j_2) \right\}.$$

A tempting approach to solve this maximization problem without forming combinations of sums is to

1. form one single sequence from the sequences of first differences for vertices d and s by merging them
2. sort them according to their values in descending order,
3. pick the j top most values which now are the j largest values, too, and
4. solve the j-median problem with the two highest indices p_0 and q_0 found in the set of j largest values of first differences.

For example: The daughter's sequence is

100 20 3

and the son's is

31 22 13

then their mixture is

100 31 22 20 13 3

which implies that the first median is placed in the daughter's tree, median number 2 and 3 are situated at the son's, number 4 in the daughter's, 5 - son, 6 - daughter.

Unfortunately, this does not work in general. Additionally, it must be taken into account that for any delta which is chosen from one of the sequences all its predecessors have also to be picked. The sums always include all deltas starting with the first. So, it does not suffice to pick the j deltas with largest value, they also must be among the first j deltas.

For example: If the son's sequence is

11 32 13

then the mixed sequence

100 32 20 13 11 3

does not give the correct indication for the second facility which in this case is situated in the daughter's tree again.

However, if both sequences of first differences are monotonically decreasing, then the property of being contained in the set of the j largest valued deltas and the property of belonging to a set of first j deltas, are equivalent.

Moreover, it can be shown that

Proposition 6 (Cost P6) *If the j_0-median problem in $T_d^u \cup T_s^u$ for given d, s and u, where the first two are descendants of the latter, is solved by p_0 medians in T_d^u and q_0 medians in T_s^u, and if the first differences $\Delta(d,u,j)$ and $\Delta(s,u,j)$ are both decreasing at least up to the index j_0+1, then the j_0+1 median problem is solved either by the combination of*

- *p_0+1 medians in T_d^u and q_0 medians in T_s^u or*
- *p_0 medians in T_d^u and q_0+1 medians in T_s^u.*

Proof:
Even if the sequences are not decreasing for all indices j, the stated assumption of monotonic decline suffices, because to solve the j_0+1 median problem only the first j_0+1 first differences from each sequence are permitted. Any delta of higher index asks for a higher number of medians.

Since p_0 and q_0 solve the j_0-median problem, the sum

$$\sum_{j_1=1}^{p_0} \Delta(d,u,j_1) + \sum_{j_2=1}^{q_0} \Delta(s,u,j_2)$$

is maximal. Consequently,

$$\Delta(d,u,p_0) \geq \Delta(s,u,q_0+1)$$

and

$$\Delta(d,u,p_0+1) \leq \Delta(s,u,q_0).$$

Otherwise, the sum can be improved: e.g. if $\Delta(d,u,p_0) < \Delta(s,u,q_0+1)$, then the sum is improved by removing $\Delta(d,u,p_0)$ and adding $\Delta(s,u,q_0+1)$ leading to an increased sum $\sum_{j_1=1}^{p_0-1} \Delta(d,u,j_1) + \sum_{j_2=1}^{q_0+1} \Delta(s,u,j_2)$.

It follows from this observation and the assumed monotonic decline of the first differences of both sequences that

$$\Delta(d,u,p_0) \geq \Delta(s,u,j) \quad \forall\, j:\ j \in \{q_0+1,...,j_0+1\}.$$

So, $\Delta(d,u,p_0)$ still belongs to the j_0+1 largest values of first differences at least in the set of the first j_0+1 differences from both sequences. Or in other words, it does not make sense to

use less than p_0 medians in the tree T_d^u and use more than q_0+1 medians in T_s^u, instead, since for a positive $j \leq p_0$

$$\sum_{j_1=1}^{p_0-j} \Delta(d,u,j_1) + \sum_{j_2=1}^{q_0+1+j} \Delta(s,u,j_2) \leq \sum_{j_1=1}^{p_0} \Delta(d,u,j_1) + \sum_{j_2=1}^{q_0+1} \Delta(s,u,j_2).$$

By symmetry the same holds for the second tree. For any positive $j \leq q_0$

$$\sum_{j_1=1}^{p_0+1+j} \Delta(d,u,j_1) + \sum_{j_2=1}^{q_0-j} \Delta(s,u,j_2) \leq \sum_{j_1=1}^{p_0+1} \Delta(d,u,j_1) + \sum_{j_2=1}^{q_0} \Delta(s,u,j_2).$$

The two candidates which are left for the solution of the j_0+1 median problem in $T_d^u \cup T_s^u$ are

$$\sum_{j_1=1}^{p_0} \Delta(d,u,j_1) + \sum_{j_2=1}^{q_0+1} \Delta(s,u,j_2) \quad \textit{and} \quad \sum_{j_1=1}^{p_0+1} \Delta(d,u,j_1) + \sum_{j_2=1}^{q_0} \Delta(s,u,j_2).$$

And the winner is the tree which offers higher savings by adding an additional median, because either

$$\begin{array}{rcl} \Delta(d,u,p_0+1) & \leq & \Delta(s,u,q_0+1) \\ & \textit{or} & \\ \Delta(d,u,p_0+1) & > & \Delta(s,u,q_0+1). \end{array}$$

□

According to the recursion formula the values of $\text{Cost}_d^{j+1}(u)$ and $\text{Cost}_s^{j+1}(u)$ have to be calculated before $\text{Cost}_x^{j+1}(u)$ can be determined for given x and u. So, it is known in advance, whether the first differences with respect to d and u and s and u respectively are monotonically decreasing. If they are, then the Minimum Loop can be replaced by the comparison of two values: $\Delta(d,u,p_0+1)$ and $\Delta(s,u,q_0+1)$. This is especially helpful, if the final goal of j is quite large.

Moreover, the verification of the monotonicity of the two sequences may become the only purpose for calculating $\text{Cost}_d^{j+1}(u)$ and $\text{Cost}_s^{j+1}(u)$. One or both of the values are probably never used to calculate j-median solutions in T_x^u. In fact, if it is known in advance, that the first differences of the cost functions with respect to any pair (x,u) are monotonically decreasing, this knowledge can be used to design a very efficient algorithm. Unfortunately, Proposition 6 does not hold in general which is demonstrated by a counterexample in Figure 2.5.

To see what happens the first differences are calculated with respect to the children of root r: T_y^r and T_z^r:

Tree	# of medians	Set of medians	Δ	$\sum\Delta$
T_y^r	1	$\{y\}$	1.5	1.5
T_z^r	1	$\{z\}$	6	6
	2	$\{z, o_1\}$	1	7
	3	$\{o_1, o_2, o_3\}$	2	9

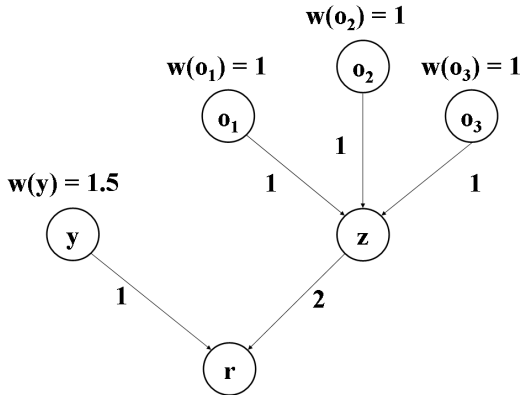


Figure 2.5 Counterexample

The sequence for T_y^r is trivial.

The first median of T_z^r is situated at z, because at z a total weight of 3 accumulates and the edge between z and r is relatively long compared to the other edges. The second median is placed at o_1, because facilities at o_1 and o_2 instead of o_1 and z lead to remaining costs of 3 for assigning o_3 to r which is more than assigning o_2 and o_3 to z for 2 units. Clearly, by the use of 3 medians all the customers can be assigned to an individual facility. Since z is not a customer, there is no need for a facility at its site anymore. The resulting sequence of first differences is not monotonically decreasing.

Still, according to Proposition 6 the choice for the first 2 medians can be made according to the sequence of first differences, because the first differences $\Delta(z, r, j)$ are monotonically decreasing for j less than 3.

Tree	# of medians	Set of medians	Δ	$\sum \Delta$
T_r	1	$\{z\}$	6	6
	2	$\{z, y\}$	1.5	7.5
	3	$\{o_1, o_2, o_3\}$	1.5	9
	4	$\{o_1, o_2, o_3, y\}$	1.5	10.5

However, in the 3-median problem $\Delta(z, r, 3)$ exceeds the savings gained by placing a facility at vertex y during the previous step. It makes sense to remove this median again and use it in the tree T_z instead. Finally, a forth median is situated at the last customer not directly supplied yet.

It is noteworthy, that this effect does not simply rely on the counter motion of the first differences with respect to z. It is a necessary condition, that the value of first difference $\Delta(y, r, 1)$ lies in the interval generated by the sequence of first differences which increases. In this example the interval is $[1; 2]$ and the weighted distance 1.5 between y and r is contained in this interval. If the weighted distance for y happens to lie outside this interval, the problem dissolves.

It should also be observed, that it is an artefact of this example that the resulting sequence of first differences in T_r is decreasing again. By a different choice for the weight of y it can be forced to change directions.

It remains to clarify what causes the counter motion in the sequence of first differences with respect to z. To study this effect the first differences of the trees $T^r_{o_1} \cup T^r_{o_2} \cup T^r_{o_3}$ and T_z are listed. For the tree T_z the notion of first differences is thereby extended by a zero - delta value, i.e. $\Delta(z, z, 0)$ or shorter $\Delta(z, 0) := \text{Cost}^0_r(r) - \text{Cost}^0_z(z)$ and z is counted as a median, not in T_z but in T_r.

Tree	# of medians	Set of medians	Δ	$\sum\Delta$
$T^r_{o_1} \cup T^r_{o_2} \cup T^r_{o_3}$	1	$\{o_1\}$	3	3
	2	$\{o_1, o_2\}$	3	6
	3	$\{o_1, o_2, o_3\}$	3	9
T_z	1 (0 *inside* T_z)	$\{z\}$	6	6
	2 (1)	$\{z, o_1\}$	1	7
	3 (2)	$\{z, o_1, o_2\}$	1	8
	4 (3)	$\{z, o_1, o_2, o_3\}$	1	9

The j-median solutions for T^r_z are determined by comparing the first and second part of the recursion formula. From the inspection of the two lists it becomes clear that the second part dominates the first for j equal to 1 and 2. A facility at vertex z is always part of the solution. For $j = 3$ the first part becomes dominant, and the facility at z is obsolete.

So, one important reason for increasing values of the first differences in T^r_z is, that the solution for T^r_z switches between first and second part of the recursion formula for at least one value of j. In fact, if this switch never occurs, then the first differences in T^r_z inherit their behavior from that part of the recursion formula the solutions are calculated with.

For the switch to occur, vertex z must be an interesting location for a facility. In the example the distance between z and the root is large enough to make z interesting. If the length of edge (z, r) is less than 0.5, then it is better to situate facilities directly at the customer sites o_i.

But, z must not be too interesting, i.e. it should not be necessary to always situate a facility at z. In the example this is the case, because z is not a customer's site. So, for $j = 3$ the facility at z becomes dispensable.

Still, this is only a necessary reason for increasing first differences in T^r_z. If the distances of o_2 and o_3 to z are for example both equal to 0.5, the resulting first differences will be monotonically decreasing.

In general, first differences have a strong tendency to decrease, since the j^{th} difference has to be less or equal to the $(j-1)^{st}$ cost value by definition and cost values are strictly decreasing until they are zero. Large values of first differences, which in practice hopefully occur for the first medians situated in a tree, force subsequent differences to be rather small. At least for a large number k of medians which have to be placed in a tree an inspection of the first differences probably speeds up running times significantly.

2.4.3 More improvements and special purpose algorithms

The cost $\text{Cost}_x^0(u)$ for 0-median solutions are of special interest. The basic cost information for the entire tree T_r is stored in the 0-median cost function. The recursion formula merely recombines these values for different vertices u and trees T_x to find the optimal solution for a j-median problem. Therefore, $\text{Cost}_x^0(u)$ has to be computed for any pair of vertices x and u. The value can be computed directly.

In order to do this efficiently, the strategy of the depth-based algorithms can be applied as it is introduced in Section 2.3.12. A linear function $a(x) * \alpha + b(x)$ is recursively computed for every vertex x by visiting every vertex x exactly once. Function $a(x)$ may very well be interpreted as the weight of the tree T_x, since it is the sum of the weights of all customers in T_x. Function $b(x)$ may be called the assignment cost of the tree T_x, since $b(x)$ is the contribution to the assignment cost in the overall tree T_r, if exactly one facility is situated in the tree T_x which is located at vertex x.

The number $a(x) * d(x, r)$, however, can be interpreted as the maximal amount of assignment cost in tree T_r which is saved by opening a facility at vertex x. If there is no facility situated in T_r between x and r, and if there is no facility situated inside the tree T_x, then $a(x) * d(x, r)$ is exactly the amount of savings for situating a facility at x. Otherwise, either the weight $a(x)$ is reduced by the weight of customers already supplied inside the tree T_x, or the distance $d(x, r)$ is replaced by the distance between x and the closest facility to x. The product $a(x) * d(x, r)$ will be denoted by $wd(T_x, r)$ as the weighted distance of tree T_x to vertex r.

This last thought leads to an alternative approach to solve 1-median problems. Since assignment costs are minimized by maximizing savings by properly situating an additional facility (see Section 2.4.2), the facility which solves the 1-median problem is also found by identifying the vertex which maximizes $a(x) * d(x, r)$.

Furthermore, in order to solve higher median problems in T_r, the 1-median problems have to be solved for every tree of type T_x^r. For the solution of the 1-median problem in T_x^r it is not necessary to compute more terms like $a(x) * d(x, r)$. All that is needed is to find the maximum of a subset of these values, namely, for the vertices of tree T_x. These vertices are easily identified, if a post-order traversal is used as it is described in Algorithm 4 to enumerate the vertices of the tree. If $m(x)$ denotes the smallest index of this enumeration in the tree T_x, then the integer interval $[m(x), x]$ determines the set of vertices of T_x.

The enumeration of the vertices according to a post-order traversal may be done in advance or performed simultaneously with the calculation of the cost values in T_r. In any case, following a post-order traversal of the vertices x of tree T_r the subsequent computations are performed.

Algorithm 5 (Solution of 1-MP)

Main	Loop 1	Through all vertices x of the tree T_r
	If x is a leaf, then	• $a(x) = w(x)$, where $w(x)$ is the weight of x. • $b(x) = 0$ • $m(x) = x$ • $wd(T_x, r) = a(x) * d(x, r)$ • $wd(T_x, r)$ is inserted in a sorted list LWD of all weighted distances between r and the visited trees T_x so far. • $M^1(T_x^r) = \{x\}$, where $M^1(T_x^r)$ is the set of the median which solves the 1-median problem in T_x^r. • $\text{Cost}_x^1(r) = 0$
	x is not a leaf and y and z are its children, then	• $a(x) = w(x) + a(y) + a(z)$ • $b(x) = a(y) * d(y, x) + b(y) + a(z) * d(z, x) + b(z)$ • $m(x) = \min(m(y), m(z))$ • $wd(T_x, r) = a(x) * d(x, r)$ • $wd(T_x, r)$ is inserted in the sorted list LWD. • $M^1(T_x^r) = arg \max_{y \in [m(x), x]} wd(y, r)$ • $\text{Cost}_x^1(r) = wd(T_x, r) + b(x) - wd(T_{m_1}, r)$, where $m_1 \in M^1(T_x^r)$ is the median of the 1-median solution.

The list LWD of weighted distances between trees T_x and root r can be further utilized to find the 2-median or even higher solutions in T_r. The 1-median solution of T_r — $x_{max} \in M^1(T_r)$ — heads the list LWD. What is its relationship to the solvers $M^2(T_r)$ of the 2-median problem in T_r? And how are the second, third and so forth entries of LWD related to that problem?

Fact 1) x_{max} is an ancestor or a descendant of at least one of the medians of the 2-median problem in T_r.

If no facility is opened on the path from x_{max} to r (i.e. x_{max} is not an descendant of a facility), then the total distance $d(x_{max}, r)$ can be accounted for savings by situating a median at x_{max}.

If no facility is situated inside the tree $T_{x_{max}}$ (i.e. x_{max} is not an ancestor of a facility), then the total weight $a(x_{max})$ accumulated at x_{max} can be accounted for savings by situating a facility at x_{max}. Since $wd(T_{x_{max}}, r) = a(x_{max}) * d(x_{max}, r)$ maximizes these savings not only for x_{max} but for the tree T_r, at least one of the two conditions must be true. Otherwise, the assignment to the two vertices in $M^2(T_r)$ can be improved, which contradicts their optimal choice.

Fact 2) If one of the two medians in $M^2(T_r)$ is not related to x_{max}, i.e. the median and x_{max} are not connected by a directed path, then x_{max} is the other median.

Because of the assumption and Fact 1) that there is at least one median related to x_{max}, it follows that there is exactly one median contained in the subgraph which is derived from the union of the path between r and x_{max} and $T_{x_{max}}$. Since also in this graph the 1-median solution is found by maximizing $wd(T_x, r)$, x_{max} has to be the median.

Therefore, a pair of candidates for the 2-median solution is found in x_{max} and the first vertex following x_{max} in the list LWD which is not related to x_{max}. If the second in line is not related to x_{max}, then the 2-median solution is found. Actually, any successive vertex with the same property, i.e. not being related to any of the preceding vertices, solves an equivalently higher median problem.

Unfortunately, it is more likely that the second vertex in line is an ancestor of the first, because the weight of its tree is at least as high as the one of x_{max} and the distance between the second vertex in line and r has probably not decreased very much. Anyway, a check may pay off.

Fact 3) If x_{max} is related to both medians in $M^2(T_r)$, then

- either the two vertices in $M^2(T_r)$ are not related, both different from x_{max} and contained inside the tree $T_{x_{max}}$,
- or the medians are related and x_{max} lies on the path which connects the two facilities.

If the medians m_1 and m_2 are related and w.l.o.g. m_1 is the ancestor of m_2, then there exists a path which contains all three vertices m_1, m_2 and x_{max}, since by assumption x_{max} is related to both medians, too. It has to be shown, that x_{max} cannot be a true ancestor of m_1 or a true descendant of m_2.

The total savings gained by the two medians can be written as:

$$a(m_1) * d(m_1, r) + a(m_2) * d(m_2, r) - a(m_2) * d(m_1, r).$$

The first two summands result from the maximal savings gained by situating facilities at both locations. But, the weight of the customers in the tree of the second median is counted twice along the path between m_1 and r, since the one median is contained in the tree of the other. This is corrected by the third summand.

Substituting m_1 by x_{max} in this expression has got a potential to increase the first summand. If, however, x_{max} is an ancestor of m_1, then the term $a(m_2) * d(m_1, r)$ which is necessary to produce the correct value of the savings decreases, since x_{max} is closer to r than m_1. So in

total the expression increases - or better - cannot decrease which allows to choose m_1 equal to x_{max}.

Substituting m_2 by x_{max} in the savings' formula has got a potential to increase the second summand. If, however, x_{max} is a descendant of m_2 then the correcting term decreases again, now because $a(x_{max})$ is smaller or equal to $a(m_2)$.

Consequently, if m_1 and m_2 are related, then x_{max} is either one of them or lies on the path which connects them.

The only possibility left for x_{max} to be related to both medians is, that they are both contained in the tree $T_{x_{max}}$. If the medians m_1 and m_2 are related, then one of them has to be x_{max} which corresponds to the case considered just before. So, if the medians are assumed to be contained in the the tree of x_{max}, then they may also be assumed to be different from each other and from x_{max}.

This second possibility for unrelated solvers of the 2-median problem, i.e. they both lie in $T_{x_{max}}$, is theoretically possible, but practically very unlikely. Local loops of copper networks were purposefully designed with several copper centers, especially for densely populated areas. Several very thick branches of copper cables leave the central offices. The cardinal directions of the compass may give a good guess for their number. But, more than four copper centers may very well be the case. Moreover, it was standard strategy to bundle the individual twisted pairs as far as possible, i.e. not to split them up at the location which was as close to the customer as possible, even if that implied a longer distance between central office and customer, and therefore installation costs increased[38]. This strategy leads to well known switching nodes (LV, KV) which are easy to access, concentrate many twisted pairs and are consequently natural locations for facilities. Hence, rarely a local loop will be encountered where the best locations for the 2-median problem are found in the subdistrict of the vertex which provides the best location for the 1-median problem.

The 2-median problem in T_r can be solved by inspecting the list LWD which was produced to solve the 1-median problem in the following way.

1. The first vertex in LWD which is not related to x_{max} is determined an denoted by x_{norel}.
   - If this vertex is the second in line, the problem is solved.
   - Further medians can be found in subsequent vertices, if they are also not related to any of their predecessors.
   - If x_{norel} is not second in line, then a first lower bound for the savings gained by the 2-median problem is found by
   $$a(x_{max}) * d(x_{max}, r) + b(x_{max}) + a(x_{norel}) * d(x_{norel}, r) + b(x_{norel}).$$
2. The list LWD is split into two sublists LWD_A and LWD_D.
   - LWD_A contains all true ancestors of x_{max} which are preceding x_{norel} in LWD.

[38] Compare backward supply in Chapter 1 Section 1.3.2.

- LWD_D contains all true descendants of x_{max} which are preceding x_{norel} in LWD.

3. If LWD_D contains one or more pairs of unrelated vertices (x_1, x_2), another set of candidates for the 2-median solver is found by choosing x_1 and x_2 such that

$$a(x_1) * d(x_1, r) + b(x_1) + a(x_2) * d(x_2, r) + b(x_2)$$

is maximized, i.e. the first two vertices in LWD_D according to the ordering of LWD which are not related are chosen.

4. Any pair of one vertex x_1 from $\{x_{max}\} \cup \mathrm{LWD}_A$ and one vertex x_2 chosen from $\{x_{max}\} \cup \mathrm{LWD}_D$ with $x_1 \neq x_2$ is checked whether

   (a) their maximal savings

$$a(x_1) * d(x_1, r) + a(x_2) * d(x_2, r)$$

   exceed the best savings gained so far,

   (b) and if this test is successful, their true savings are inspected

$$a(x_1) * d(x_1, r) + a(x_2) * d(x_2, r) - a(x_2) * d(x_1, r).$$

2.5 Application: Descending k-median

2.5.1 Considerations concerning the distance function

In Section 2.2.2 the quality of an access network is conceptualized by the term supply rate which is defined as the percentage of customer's demands whose supplying meets a predetermined quality standard. SARU knows several ways to define quality standards:

- by restricting the maximal cable distance between customer and supplying facility,
- by limiting the maximal damping or
- by demanding a minimal transmission rate which is achieved by supplying a customer from a proper location.

(See Rule 10, distance rule, 1.4.10 in Chapter 1.)

One way of introducing supply rates to the k-median problem is as a means of reporting. For every value of j between 0 and k the j-median solution is determined for which subsequently the supply rate is calculated. Supply rates are then reported as a function of the number of medians j.

Another way of bringing the two concepts together is by defining the distance function of the k-median problem directly based on supply rates. The distance between a customer and a potential facility location is set to 0, if the quality standard is met by supplying the customer from a facility there. Otherwise, if the quality standard is violated, the distance is set to 1:

Definition 13 (Quality threshold) *The **quality threshold** is a numeric threshold by which the assignment of a customer to a facility is classified as a **good** or **bad** assignment. The quality threshold is applied to the distance function in use in the directed tree. In case distances are measured in terms of cable lengths or damping, the assignment quality is good, if the cable distance or the damping respectively between customer and supplying facility does not exceed the **quality cable threshold** or the **quality damping threshold**. In case distances are measured by transmission rates, the assignment quality is good, if the transmission rate achieved by supplying the customer from the facility does not drop below the **quality transmission threshold**.*

The definition of the quality thresholds is obviously closely related to Rule 10 (Distance rule, Chapter 1 Section 1.4.10).

2.5.2 Discrete distance functions

The so defined distance function is not complete yet. Its values for pairs of

- non-customers, i.e. vertices of zero weight, or
- non-customer and customer, or
- customer and potential facility vertex which the customer is not allowed to be assigned to

are missing. However, the first line of Recursion Formula (2.11)

$$\min_{p+q=j} \left\{ \text{Cost}_d^p(u) + \text{Cost}_s^q(u) \right\} + w(x) * d(x,u)$$

shows that the value of the distance function of two vertices matters only if the first of the two is a customer and the second one is a potential location for a facility which the customer may be assigned to. Hence, at least for the ancestor based algorithms their distance may be chosen arbitrarily, e.g. 1.

This concept for the distance function does not work with depth based algorithms. The cost functions for the 0-median solution which are the fundamental building block for any k-median algorithm are linear functions in one variable which stands for the distance between the tree rooted at x and the closest open facility ascendent to x.

The validity of the concept of depth based algorithms grounds on an additive property of the distance function which is normally associated with distances in trees:

> If u, v and x are three vertices in a tree with x a descendant of v and u an ancestor of v, then the distance of x and u is the sum of the distances of x to v and v to u
>
> $$d(x,u) = d(x,v) + d(v,u).$$

In general this additivity does not hold for discrete distance functions. If location u is a potential and "good"facility location for customer v, their distance is equal to zero. The same is true, if vertex v offers an admissible location for customer x. But, if u is too far away from customer x, then their distance must be positive.

2.5.3 Discrete versus additive distance functions

It is not surprising that the two types of distance functions produce different concepts of k-median solutions. This is illustrated by a tree with 10 customer vertices of weight 1.

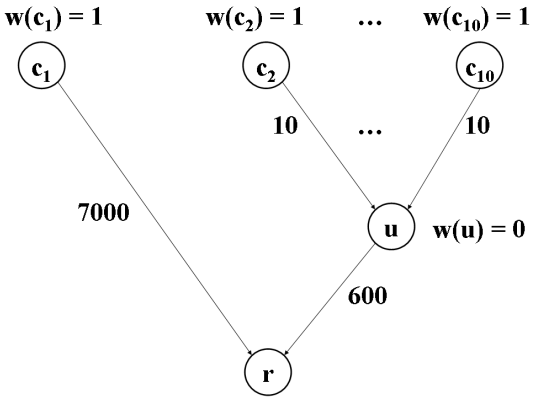
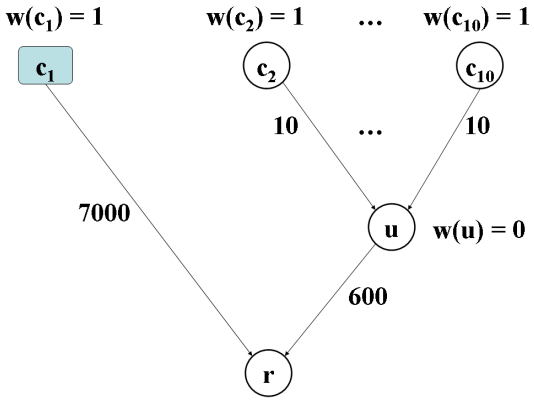


Figure 2.6

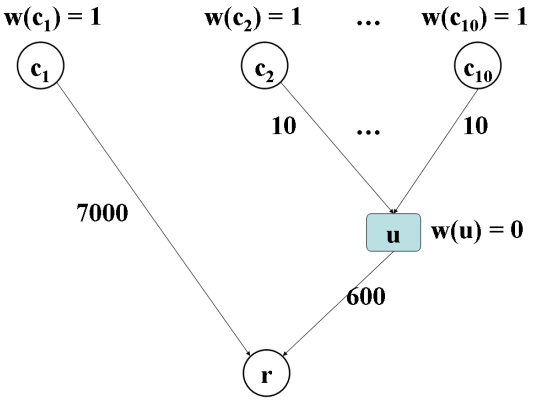
The first of the customers is directly connected to the root with a distance of 7000 meter. The other 9 customers are all rooted at an additional and intermediate vertex u. Their distance to u is 10 meters. The distance of u to the central office is 600 meters. The quality threshold for the distances between facilities and assigned customers is 600 meters.

The 0-median solution for the version with the additive distance function starts with a supply rate of 0%. Since the first customer is so far away from the root, the first median will be

located at customer c_1 due to a reduction of assignment costs of 7000 compared to 600 times 9 for locating the first facility alternatively at u or 610 for situating it at any one of the other customers. The supply rate increases to 10%. Clearly, the second "additive"median is situated at vertex u bringing the supply rate up to 100%. The median problem with additive distances continues until all customers are supplied by an individual facility located at their own site. Certainly, the supply rate is not improved anymore.



1-median solution with additive distance. Supply rate = 10%.
Figure 2.7



1-median solution with discrete distance. Supply rate = 90%

The discrete version situates the first median at vertex u, because this increases the supply rate to 90% instead of 10% for locating it at any of the customer vertices. The 2-median solution of the discrete problem coincides again with the one of the additive version and full supply is achieved. The algorithm which uses the discrete distance stops at this point, because the assignment cost are zero.

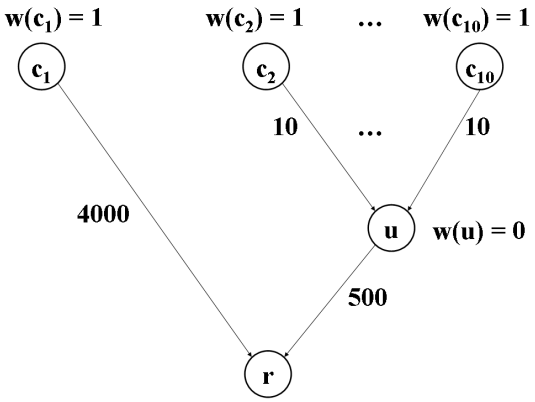
The difference of the two approaches is essential. Far away customers even of small weight are very attractive for medians, if additive distances are applied. This may cause that solutions with low supply rates are delivered first, or worse, solutions with high supply rate are never presented, like the 90% solution in the example.

A different edge weighting in the same graph (see Figure 2.8) demonstrates that the additive formulation improves the average quality in terms of lowering the average distance, whereas the quality in terms of the supply rate may not be altered at all.

The 0-median solution has got a supply rate of 90% already. The first median with respect to the additive distance is situated at u now, because 9 times 500 meters outvalues 4000 meters. The corresponding supply rate, however, does not change.

Of course, it could be argued, that it is better to improve the service quality for 9 customers beyond the predefined quality level instead of rising the supply rate to 100% by installing one facility for one customer.

This is certainly debatable.



The example makes use of the same graph with different distances and the same quality threshold. The distance between u and the root is now 500 meters. The distance between the first customer c_1 and the root is 4000 meters. The distances between the remaining customers and their parent is 10 meters.

Figure 2.8

2.5.4 Discrete distance function and transmission rates

However, on the one hand the goal of applying the k-median problem in SARU is to study solutions which establish high supply rates and high utilization of the opened facilities at the same time, and not to determine the solution with 100% supply rate.

On the other hand it turns out that the distance function derived from transmission rates is not additive at all. For example, in VDSL2+ which is the actual transmission protocol used in FTTC solutions at the time of the development of SARU[39], the maximal achievable transmission rate is already reached at some distance to the end customer. It cannot be increased by situating the supplying facility closer to the customer.

Secondly, the transmission rate is not a linear function of the cable distance between the customer and the facility. It may even be discontinuous depending on changes of the diameter of the copper adders along the copper line which connects the customer to the central office.

And finally, the most problematic property of VDSL2+ in this context is that from a certain cable distance on the transmission rate between customer and facility is zero. It cannot be decreased any further, even if the cable distance between customer and facility is increased.

Cable distance, damping and transmission rate are alternately discussed as potential distance measures for SARU[40]. However, the actual and factual goal of SARU is to situate ARUs in such a way that transmission rates are maximized. Transmission rate is the quintessence of the quality of the facility location in SARU. It is not cable distance nor damping.

Cable distance serves as an additive approximation distance for the transmission rate. This approximation may work well on a range of probably 50 to 2000 meters. Outside it provides a bad model. The deviance between the actual distance function and its approximation may still be acceptable for small distances from 50 meters down to zero, because the deviance itself can only be small. But, it is unacceptable for distances beyond the upper limit.

[39] See Chapter 1 Section 1.1.4 for details.
[40] See Chapter 1 Rule 10 (Distance rule, 1.4.10) for details.

Based on an additive approximation of the transmission rate the k-median algorithm may try to improve the placement of another facility by rearranging the facilities in such a way that for some customers the distances to their supplying facilities are reduced from 7000 to 4000 cable meters. This may constitute an excellent improvement of the assignment cost for the version with the additive distance function, but in terms of transmission rates nothing has changed for the affected customers. Transmission is still zero for them.

This consideration on distance functions used in SARU demands that the values of the distance function which exceed a certain limit are leveled. There are more arguments to underline this requirement.

- Even if the transmission rate is positive, it may not be high enough to enable Sales and Marketing to offer any of the new products for which ARUs are erected. So, what is the point of differentiating transmission rates below a certain threshold?

- Alternative transmission protocols like ADSL 2+ are still offered in parallel and provide acceptable transmission rates even over longer distances. So, what is the point of considering a facility to supply a customer with VDSL 2+, if the same or a better transmission rate may be achieved by an alternative which operates from the central office like ADSL 2+?

- Similar arguments may be found for very high transmission rates. Is it really necessary to situate another facility to improve the transmission rates for some customers from 32 to 33 Mbps, whereas the quality threshold lies around 25 Mbps?

Consequently, some thought has to be given designing the distance function for SARU properly, especially if it is not chosen to be discrete.

2.5.5 Semi discrete distance functions

Because of the preceding considerations it seems in place to find a compromise between the concepts of discrete and additive distance functions. A distance function with the property that it is constant over certain ranges and continuously increasing over others may be called a **semi discrete distance function**.

Given an additive distance like cable distance, and a discrete version of it a semi discrete distance function may be easily constructed by the product of the two. Alternatively, a semi discrete distance function is carefully designed based on a thorough analysis of the underlying problem as it was done in SARU.

The algorithm which is presented in what follows may be applied to formulations of the k-median problem with any one of the stated distance functions. However, real advantages are achieved for discrete and semi discrete distance functions. Consequently, there is no depth based version for this algorithm available.

2.5.6 Solving KMP by descending iteration (I)

It was discussed in Section 2.2.6 that for decision support in SARU it is more beneficial to explore ranges of high values of the numbers of medians, i.e. for high values of the supply rate, than for low values. It is of higher relevance to know how many facilities are needed to provide a supply rate of 100%, than to determine the supply rate for the 1-median solution[41]. If a final k-median solution with a supply rate around 80% seems desirable, then it is of higher interest to trace the supply rates down by reducing the number of medians rather than to observe the rise of supply rates by increasing the number of medians starting at 1.

Based on the Recursion Formula for KMP described in Section 2.3.8 the solution of the k-median problem in a directed tree T_r with root r starts with a call of

$$\mathrm{Cost}_r^k(r).$$

To resolve this expression — among many other things — $\mathrm{Cost}_v^j(r)$ has to be calculated for every child v of r and any integer j between 0 and k. Once these values are known — which, however, is the main work of the KMP algorithm — it is just a little bit more effort to solve the j-median problems in T_r for any integer j less than k. The values for $\mathrm{Cost}_r^j(r)$ are easily derived from the cost values already determined. Moreover, although it is not imperative to compute $\mathrm{Cost}_r^j(r)$ for $j < k$ to solve the k-median problem in T_r, the 0-median, the 1-median and so on and so forth up to the $(k-1)$-median can be calculated, before the k-median problem is finally solved.

So, viewing this procedure from the perspective of an iterative algorithm the k-median problem in T_r is solved by initially determining the cost of the 0-median problem in T_r. Then the number of medians is successively increased, until the desired amount is reached. The algorithm ascends through the parameter space of k — the number of facilities to be used.

In analogy a descending strategy for an iterative KMP algorithm is formulated: The algorithm starts with a number of facilities and the corresponding solution of the KMP which grants full supply, i.e. assignment cost are equal to 0. Afterwards the algorithm attempts to solve the k-median problem by decreasing the number of facilities step by step. It solves the KMP by descending through the parameter space of k and thereby producing solutions for the KMP for any number of medians greater than k.

However, the crucial question beside the formulation of appropriate formulas for solving the KMP by descending iteration is, where will this kind of iteration start? Or better, what is and how to determine the initial value of k and the corresponding set of medians for a descending iteration and how are they determined?

There is no doubt, that such an initial solution exists. If every customer is equipped with his personal facility, there is no need for any additional median, since in this case assignment costs are zero and cannot be improved anymore. But, is it possible that assignment cost vanish and not all the customers obtain their individual facility? This depends on the distance function

[41] This is a bit different for the 0-median solution which is equivalent to the special interest project VDSL@CO. See Chapter 1 Section "What is fiber to the x" 1.1.4 and rules R8, R9 and R12.

in use. For a regular metric the distance between two objects is zero if and only if the two objects are identical. So, if such a metric is used as a distance function in a KMP, then the total assignment cost will indeed be equal to zero if and only if customer and facility locations are identical. In SARU a different distance concept is utilized.

2.5.7 Distance function for SARU

As usual T_r denotes a directed tree with r being its root.

Definition 14 (Pseudo-quasi metric) *If c is a customer vertex in T_r, i.e. a vertex of positive weight $w(c)$, and p the path which connects c and r, then the distance function which is used in T_r*

$$d(c, f)$$

is expected to provide the following properties for any vertex $f \in T_r$:

- *(D1) If f is a vertex on path p, the value of $d(c, f)$ is defined as a non-negative number:*

$$d(c, f) \geq 0.$$

- *(D2) For at least one vertex f on path p the distance to c is zero:*

$$d(c, f) = 0.$$

- *(D3) (Monotonicity) For two vertices f and f' on path p such that f' is an ancestor of f it holds that*

$$d(c, f) \leq d(c, f').$$

- *(D4) If f is not a vertex on path p, then the value of $d(c, f)$ is set to $+\infty$:*

$$d(c, f) = +\infty.$$

And if c is not a customer, then

- *(D5) the distance between c and any other vertex f is set to $+\infty$:*

$$d(c, f) = +\infty.$$

A finite value for the distance function $d(c, f)$ is only needed, if c is a customer and f a potential facility location which c may be assigned to, i.e. f is a vertex on path p. Therefore, in case (D4) and (D5) any value could be used to define $d(c, f)$ instead.

Symmetry is not needed for this distance function, since if f is a vertex on the path from customer c to root r, then c is certainly not a vertex on the path from f to root r. Therefore, symmetric values of the distance function are never needed. However, if desirable the distance function could always be extended to a symmetric distance.

It is more important to note, that reflexivity is not required: from $d(c, f) = 0$ it does not necessarily follow, that $c = f$. In fact, for an efficient application of the algorithm which is

going to be described in what follows, the distance between a customer and many subsequent vertices should be zero.

The main purpose of property (D2) is to ensure that the value of the assignment cost for the KMP is always equal to zero for some value of k.

The triangle inequality does not hold for this distance function, too. For example, the distance function is derived from a check, whether the cable distance between two vertices exceeds the quality cable threshold t. Three vertices c_1, c_2 and c_3 lie on a common path such that c_2 lies between the other two vertices and the cable distance between c_1 and c_2 and between c_2 and c_3 is t. Consequently, the cable distance between c_1 and c_3 is $2t$. The derived distance between c_1 and c_3 is therefore 1, whereas for the other two pairs it is 0. Hence, this contradicts the triangle inequality. It is substituted by a weaker version in property (D3).

The interpretation of a finite and non-negative distance between a customer and a potential facility is that the customer is allowed to be assigned to that facility. An infinite value implies that assigning the customer to that location contradicts specification Rule 6 (Admissible locations for ARUs, 1.4.6 in Chapter 1). A positive distance may be seen as some sort of penalty for the assignment of a customer to a specific facility. Now, the distance rule is violated. But, the k-median approach takes care of it by minimizing the "amount"of violation.

For any vertex f' for which property (D2) holds, properties (D1) and (D3) imply that the distance between any of its descendants on path p, f, and c is zero, too. Especially, this guarantees that $d(c, c)$ is always zero. Conversely, if f and c have got a positive distance, then any ancestor of f has got positive distance to c. This conclusion resembles property (A2) from Chapter 1 Section 1.6.3 Definition 9 for admissible facility locations.

Discrete distance functions — as they were defined in Section 2.5.2 — which are based on quality thresholds according to Definition 13 fulfill properties (D1) to (D3) and can be extended to distance functions which also satisfy (D4) and (D5). Since discrete distance functions only take values 0 and 1, Property (D1) is trivially true. Property (D2) holds for properly chosen values of the threshold. If quality cable threshold or quality damping threshold are positive numbers, then at least the discrete distance of a customer site to itself is zero. In case of the quality transmission threshold a little bit more knowledge about the transmission technology in use is necessary to make a proper choice for the threshold. Its value should be chosen less than the maximal transmission rate which can be achieved by the given technology to guarantee the validity of (D2) for such a distance function.

Property (D3) is derived from the monotonicity of cable length, damping and transmission rate. Cable length and damping increase, transmission rates decrease — even if not necessarily strictly — by situating the supplying facility further and further away from the customer.

2.5.8 Effect of the pseudo-quasi metric on Proposition 2 to 6

Before exploring the impacts of this kind of distance function on the k-median problem and its solution, it is necessary to show that the propositions stated and proved so far stay valid, if this weaker distance concept is used.

Proposition 2 and 3. Inequality (2.13) stays valid because its proof does not use any explicit properties of a distance function. In fact, this inequality is true, even if a completely arbitrary distance function is used in the tree, e.g. distances are negative, or the distance between child and parent is greater than the distance between grandchild and grandparent. Adding a facility to a median set cannot increase the assignment cost, because in worst case the assignment of customers to medians is kept as it was before the additional facility was opened. And the $j+1$ medians of the $(j+1)$-median solution minimize the assignment cost by definition.

The proof for Inequality (2.14), however, explicitly uses that the distance function is non negative. But, this property is provided by the weaker version of the distance function. Equation (2.12) of Proposition 2 is a direct consequence of Proposition 3.

Proposition 4. The proof of Inequality (2.16) makes use of an explicit computation of the cost function at least for some of the customers for different facility locations. These costs are subsequently compared, and it is important that their difference is non negative (Equation (2.17)) which is guaranteed by Property (D3) for the pseudo-quasi metric.

Proposition 5. The key ingredients for the proof of Proposition 5 are the presentation of a candidate for the solution of the j-median problem in T_x^u for all ancestors u of the vertex u_0^j and the application of Proposition 4. Additionally, in case the root of the tree T_x^u is itself a customer's site, it is necessary to guarantee that total assignment cost does not change. This is achieved by Property (D2) and (D1) of the pseudo-quasi metric which forces $d(u,u)$ to be equal to zero.

Proposition 6. This proposition and its proof basically state how to pick j_0+1 numbers from 2 decreasing sequences of non negative numbers, such that the sum of these j_0+1 numbers is maximized. For the optimal choice of the numbers it is irrelevant how they were generated. Moreover, like Equation (2.13) in Proposition 3 this proposition is true for arbitrary distance functions, because it reflects the behavior of the cost function when facilities are successively opened in a tree.

The main property of Proposition 6 — that the sequences of first differences are decreasing — is part of the assumption and does not have to be true.

2.5.9 Distance function, admissibility and the CU Net algorithm

In the following section a connection between the KMP and the theory presented in Chapter 1, especially with the CU Net algorithm is established. Based on the stated properties of the distance function, admissibility according to Chapter 1 Section 1.6.3 Definition 7 can be defined for all customers of T_r.

Definition 15 *For a customer c, a vertex $f \in T_r$ is an admissible facility location, if f lies on the path which connects c to root r and $d(c,f)=0$, where $d(.,.)$ is a pseudo-quasi metric.*

Since Property (A1) from Chapter 1 corresponds to (D2) and Property (A2) holds as it is shown above, the CU Net algorithm can be applied to solve the facility location problem in T_r the way it is stated in Chapter 1 Section 1.6.3.

Three questions arise and are subsequently answered. Are the facilities which are derived by the CU Net algorithm medians of a k-median problem for some value of k? What is the value of k? What is the total assignment cost?

All the customers are assigned to admissible locations by the CU Net algorithm. Admissibility is defined by zero cost of assignment. Consequently, the total cost of the assignment is zero.

At most one of the facilities which are produced by the CU Net algorithm may be situated at the root r. If a facility is situated at the root, it is not counted as a median according to the terminology of the k-median problem in directed trees (see Section 2.3.6). So, the number of potential medians is given by the number of the facilities located inside the tree T_r, i.e. not at its root. This number shall be denote by k_0 and is the motivation for the following definition:

Definition 16 (ZCMS) *The **zero cost median solution** (ZCMS) of a graph is a k-median solution of the given directed tree whose assignment cost is zero and the number of medians is minimal. The set of medians of the zero cost solution is called the **zero cost medians**. The number of zero cost medians of the given graph is called the **zero cost number** (ZCN). The zero cost number function on the set of subtrees is denoted by $ZCN(T)$ and returns the zero cost number of the given subtree. In the special case of T_x the zero cost number function is also written as $zcn(x) := ZCN(T_x)$.*

According to this definition $k_0 := \mathrm{ZCN}(T_r)$.

Any customer is assigned to a facility by the CU Net algorithm which he is actually allowed to be assigned to, since the distance between customer and facility is finite. The k_0 facilities found inside the tree obviously constitute a feasible solution of the k_0-median problem. Since the corresponding assignment cost is zero, the assignment cost for an optimal solution is at most zero. And because assignment costs are defined as a sum of products of positive and non-negative numbers, the optimal cost is at least 0. In summary:

Proposition 7 (ZC P1) *The k_0 facilities which are determined inside the tree T_r by the CU Net algorithm constitute the solution set for the k_0-median problem in the tree T_r. The assignment cost is zero, and k_0 is the smallest number of medians for which the assignment costs are zero.*

Proof:
It remains to prove the minimality of k_0.

The minimality of k_0 is shown by means of the customers c_i which are assigned on maximal distance to the facilities f_i inside the tree T_r by the CU Net algorithm (See Chapter 1 Section 1.6.4). Since all f_i lie inside the tree, they are different from the root and they have got a parent. The maximality of the distance between f_i and c_i implies that the parent of f_i is not an admissible location for the corresponding customer, i.e. that the value of the distance function between customer and parent vertex is positive.

The (k_0-1)-median solution consists of k_0 facilities: the k_0-1 medians inside the tree and one facility at the root. All of the k_0 customers c_i are assigned to one of these facilities[42]. If one of the customers is assigned to the root, this assignment is not admissible in the sense of the CU Net algorithm, since the root is not an admissible location for any of these customers. Hence, there is one vertex with positive weight and positive distance to its facility. Consequently, the total assignment cost must be positive and k_0 is the smallest number of facilities to produce a zero cost assignment.

In case none of the k_0 customers c_i are assigned to the facility at the root, at least two of them have to be assigned to the same facility of the remaining (k_0-1) medians. Since the set of customers c_i is constructed by the CU Net algorithm in such a way, that no two of them may be admissibly assigned to the same location according to Theorem 1 part b) Chapter 1 Section 1.6.5, the assignment cost for at least one customer must be positive again which consequently has to be true for the total assignment cost, too.
□

By Proposition 3 the assignment costs are positive for any smaller value of k, because they increase with decreasing value of k.

2.5.10 First step to solve KMP by descending iteration

The development of an iterative algorithm which begins its iteration at the zero cost number — $zcn(r)$ — and successively decreases the number of medians, starts with an inspection of the behavior of the Recursion Formula (2.11) presented in Section 2.3.8.

Starting with the zero cost solution with zero cost number k_0 the first attempt is to find the (k_0-1)-median solution in T_r. The key idea to achieve this goal is to imitate the algorithm which solves the KMP by descending recursion using the cited formulas

$$\mathrm{Cost}_x^j(u) = \min \begin{cases} \min\limits_{s+t=j} \left\{ \mathrm{Cost}_y^s(u) + \mathrm{Cost}_z^t(u) \right\} + w(x) * d(x,u) \\ \mathrm{Cost}_x^{j-1}(x) \end{cases}$$

whose first call reduces to

$$\mathrm{Cost}_r^{k_0-1}(r) = \min_{s+t=k_0-1} \left\{ \mathrm{Cost}_y^s(r) + \mathrm{Cost}_z^t(r) \right\},$$

because of Proposition 2, where y and z are the children of the root r. Subsequently, j-median problems have to be solved in the trees T_x^r for $x \in \{y, z\}$.

By applying a similar argument as during the horizontal decomposition step in Section 2.3.7 it can be deduced that the zero cost median solution of the tree T_r splits into the zero cost median solutions for the two trees T_y^r and T_z^r.

[42]The term 'assignment' is differently used in the CU Net algorithm and the k-median algorithms. The CU Net algorithm assigns customers only to admissible vertices. In the current context this means the distance between vertices is zero. The k-median algorithms assigns customers to vertices with finite and possibly positive distance.

Any of the zero cost medians of T_r is contained in the one or the other tree. Ignoring one of the subtrees does not change assignment costs for the customers in the remaining subtree with respect to the medians there. So, the split of the medians produces p_0- and q_0-median solutions of zero cost in the respective subtrees. These assignments cannot be improved by removing a median from one of the subtrees and still remaining a zero cost solution. This would otherwise contradict the minimality of k_0.

Hence, the zero cost median solution of T_y^r and T_z^r are of size p_0 and q_0 respectively and $p_0 + q_0 = k_0$.

Since s and t are non-negative integers whose sum is $k_0 - 1$ and $k_0 = p_0 + q_0$, it may happen that s is chosen greater than p_0 or t is chosen greater than q_0. But, increasing, for example, the number of medians in T_y^r beyond p_0 does not produce any kind of advantage, because from p_0 upward all s-median solutions are of zero cost.

On the other side, the increasing number of medians in T_y^r has to be compensated in the tree T_z^r by reducing the number of medians. Since $s + t = p_0 + q_0 - 1$, t has to be less or equal to $q_0 - 1$. Because of the minimality of q_0, $\mathrm{Cost}_z^t(r)$ is positive in such cases. But, for t-median solutions with positive cost, $\mathrm{Cost}_z^t(r)$ increases strictly for decreasing t according to Proposition 3 Inequality (2.14). This produces a real disadvantage for the combination of the two solutions to a $k_0 - 1$ solution of the original tree T_r.

The minimum which is sought in the recursion formula will never be achieved for such values of s. Conversely, by symmetry the same property holds for values of t exceeding q_0. The values of s and t may therefore be limited by p_0 and q_0 respectively. The recursion formula can be rewritten as

$$\mathrm{Cost}_r^{k_0-1}(r) = \min_{\substack{s+t=k_0-1 \\ s \le p_0 \; and \; t \le q_0}} \left\{ \mathrm{Cost}_y^s(r) + \mathrm{Cost}_z^t(r) \right\}.$$

Next, a transformation of variables is applied to the recursion formula by setting

$$s = p_0 - p \; and \; t = q_0 - q.$$

Since s and t are both non-negative integers, p and q are integers and

$$p \le p_0 \; and \; q \le q_0,$$

and since s is less or equal to p_0 and t is less or equal to q_0, p and q are both non-negative integers, too.

The equation $s + t = k_0 - 1$ transforms to

$$p_0 - p + q_0 - q = k_0 - 1$$

which becomes

$$p + q = 1 + p_0 + q_0 - k_0.$$

In the case of the zero cost median of the root and its children this equation simplifies even further, because the zero cost numbers at the right hand side add up to zero. However, in general this is not the case which will be discussed later.

So, for the special case of removing one median from the zero cost solution the final version of the recursion formula at least for the root of the tree T_r reads

$$\mathrm{Cost}_r^{k_0-1}(r) = \min_{\substack{p+q=1 \\ 0 \le p \le p_0 \text{ and } 0 \le q \le q_0}} \left\{ \mathrm{Cost}_y^{p_0-p}(r) + \mathrm{Cost}_z^{q_0-q}(r) \right\}.$$

For the general case when trees of type T_x^u and an arbitrary reduction of j medians are considered, these considerations demonstrate that the keys to tackle the recursion formula to determine

$$\mathrm{Cost}_x^{j_0-j}(u), \text{ where } j_0 := \mathrm{ZCN}(T_x^u),$$

are a variable substitution and the ability of finding the j_0 zero cost medians for any of these trees.

2.5.11 Zero cost median solutions for all subtrees

To solve the k-median problem by descending iteration the zero cost number of every tree of type T_x^u has to be computed. Certainly, this could be done by determining the zero cost solution of these trees by applying the CU Net algorithm to all of them. But, such an approach produces long running times and turns out to be unnecessary anyway, because:

Proposition 8 (ZC P2) *For any tree T_x a zero cost median solution can be derived from the zero cost median solution of the tree T_r which the CU Net algorithm produces. If $ZCM(T_r)$ denotes the zero cost median of T_r resulting from the CU Net algorithm, then a zero cost median of T_x is given by*

$$ZCM(T_x) = ZCM(T_r) \cap \{vert(T_x) \setminus \{x\}\}.$$

By definition the root vertex x is not part of any k-median solution, even if it is a median in the zero median solution for T_r.

Proof:
According to Corollary 1 Chapter 1 the CU Net algorithm is optimal on rooted directed trees, i.e. it delivers the minimum number of facilities $\mathrm{ZCM}(T_r)$ which are necessary to fully supply all customers.

The CU Net algorithm has assigned every customer which lies in T_x to a facility inside T_x, on x or outside of T_x. For all customers assigned to facilities inside T_x this assignment is kept. For facilities on x or outside of T_x which supply customers in T_x the vertex x is a descendant. Consequently, property (A2) (see Chapter 1 Definition 9) makes x an admissible facility location for all these customers. They are assigned to the facility at vertex x. The result is a feasible assignment of all customers in T_x to facilities in T_x.

By Theorem 1 any facility f of the zero cost median solution of T_r which is situated inside of T_x represents a different customer c which was determined by the CU Net algorithm and assigned to f with maximal distance. These customers were chosen such that for any two of them no vertex can be found in T_r to which both of them may be admissibly assigned. If there is no such vertex in T_r, there is certainly no such vertex in T_x. This proves that all the facilities

situated inside T_x are indispensable in order to ensure a feasible assignment. The minimality of this set is a consequence of the fact that the facilities are indispensable.

Since admissibility is defined as zero distance between vertices, the facilities inside T_x constitute a zero cost median solution for T_x.
□

Zero cost median solutions for trees of type T_x^u are closely related to the zero cost median solutions of T_x. Basically, the zero cost median solution ZCM(T_x) from above is also a zero cost solution for T_x^u. Under certain conditions an additional facility has to be situated at x. To be able to state the condition when the additional facility is needed, the concept of the zero cost ancestor is introduced.

Definition 17 (ZCA) *Given a rooted, directed tree T_r and a vertex x in T_r, the **zero cost ancestor** of x — in short **ZCA(x)** — is the ancestor u of x which fulfills the following properties:*

- *there exists a zero cost median solution ZCM in T_x which is also a zero cost median solution for T_x^u,*
- *the path distance between x and u is maximal.*

The zero cost ancestor is well defined, since the maximum over a non-empty and finite set of ancestors which are of different distance to x exists and is unique.

Some care has to be taken in the choice of the zero cost solution ZCM in Definition 17. In general, k-median solutions and especially zero cost median solutions are not unique. Moreover, not every zero cost solution of T_x is useful for the concept of the zero cost ancestor. In Figure 2.9 a simple example illustrates how easily different solutions can be constructed.

Three different customers c_1, c_2 and c_3 have to be supplied. The solid lines are the edges of the copper tree. The dotted lines indicate to which facility locations the customers may be admissibly assigned, i.e. by an assignment an optimal transmission rate is provided. For customer c_1 all ancestors are admissible locations. For customer c_2 only his own location and the location of his parent z are admissible. For the third customer c_3 the ancestors z and x are admissible locations.

The two solutions depicted in Figure 2.9 are both zero cost median solutions with respect to the tree T_x. In the left tree customer c_3 is assigned to the facility at vertex z, whereas in the right tree c_3 is assigned to the root x of tree T_x. The situation in the left tree even constitutes a zero cost solution for the tree T_x^u. But, the facility selection in the right tree is not of zero cost with respect to T_x^u anymore. Still, according to Definition 17 vertex u may be the zero cost ancestor of x.

With the concept of the zero cost ancestor it is straight forward to formulate the zero cost solutions for trees of type T_x^u.

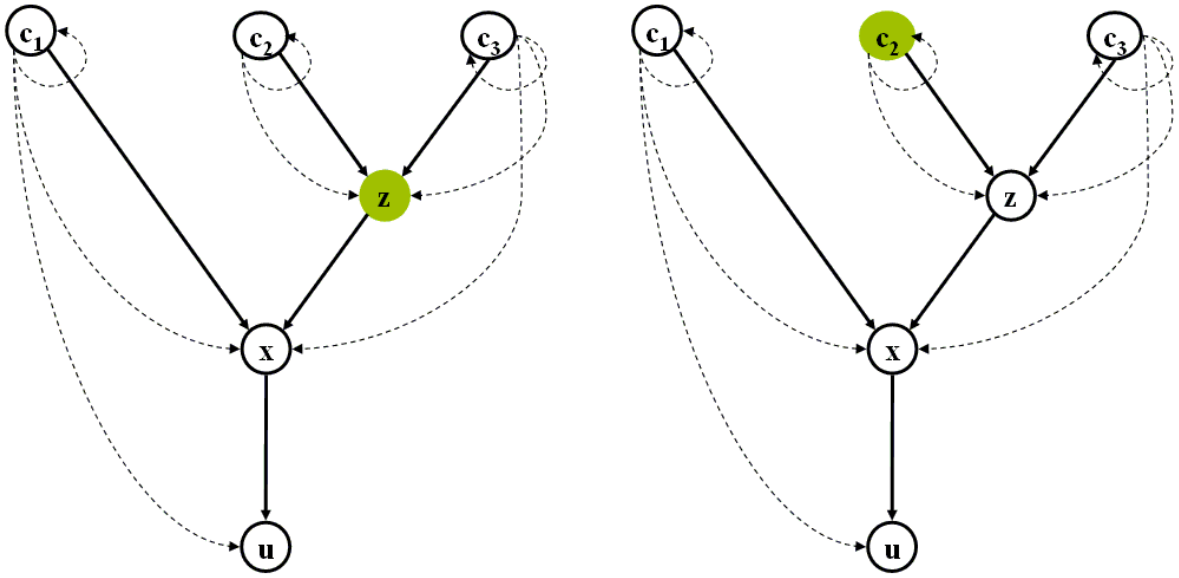


Figure 2.9: Tree with two different facility locations which both induce zero cost median solutions with respect to the subtree T_x. Left tree with facility at vertex z. Right tree with facility at vertex c_2.

Proposition 9 (ZC P3) *For the tree T_x^u, where u is an ancestor of x, a zero cost median solution is constructed in the following manner:*

If u is a descendent of ZCA(x) or equal to ZCA(x), then

$$ZCM(T_x^u) = ZCM(T_x^{ZCA(x)}). \tag{2.19}$$

If u is a true ancestor of ZCA(x), then

$$ZCM(T_x^u) = ZCM(T_x^{ZCA(x)}) \cup \{x\}. \tag{2.20}$$

Proof:
Equation (2.19) is trivially true, if $u = \text{ZCA}(x)$. If u is a descendent of ZCA(x), then all the customer assignments can be kept, except for those who are assigned to the root ZCA(x). They are now reassigned to the new root u. Clearly, the assignment cost are still zero. The number of medians cannot be reduced. Otherwise, the zero cost median solution of T_x could be improved which leads to a contradiction.

If u is a true ancestor of ZCA(x) and the zero cost solution of T_x^u does not contain vertex x as a median, this zero cost solution is also a zero cost solution of T_x. But, this contradicts the maximality of the path distance between x and ZCA(x), since the path between x and u is even longer: it contains the path between x and ZCA(x). Consequently, for true ancestors of ZCA(x) the zero cost median always contains vertex x. But, then any zero cost solution of T_x together with vertex x constitutes a zero cost solution of T_x^u, especially the zero cost solution

$\mathrm{ZCM}(T_x^{\mathrm{ZCA}(x)})$.
□

Proposition 3 is not constructive, since it refers to a zero cost solution of T_x which is proven to exist, but its construction is unknown. This flaw is resolved in the preceding section, where an algorithm is stated which delivers all necessary information.

2.5.12 Computation of the zero cost ancestor function

The importance of the concept of the zero cost ancestor is due to the fact that by means of the zero cost ancestor function the zero cost medians and consequently, also their number can be easily determined for any tree of type T_x^u. Moreover, T_r is also a subtree of this kind ($T_r = T_r^r$) which entails that the zero cost median solution for the original problem is implicitly solved by the computation of the zero cost ancestor function. An application of the CU Net algorithm becomes superfluous.

The zero cost medians of T_x are the fixed-points of the zero cost ancestor function restricted to the vertices of T_x. If an enumeration of the vertices by a post-order traversal is used, the selection of all vertices of T_x is quite comfortable to realize. As already mentioned in Section 2.3.11 the vertex-set of a tree of type T_x is identified by the integer interval $[m(x); x]$, where $m(x)$ denotes the smallest ID of a vertex in T_x.

To find the zero cost medians and the zero cost number of trees T_x^u an inspection of the ancestral relation of u and $\mathrm{ZCA}(x)$ is necessary. Again, by using the enumeration by a post-order traversal this becomes an easy task. Vertex u is an ancestor of $\mathrm{ZCA}(x)$ and consequently, x is the location of an additional facility, if and only if $u > \mathrm{ZCA}(x)$.

First the algorithm is stated and subsequently its correctness is proved.

Algorithm 6 (Det ZCM, ZCN and ZCA)

Preprocessing	Algorithm 4	The vertices of the tree T_r are enumerated according to a post-order traversal.
Initializing	Step 0	The zero cost ancestor array $\mathrm{ZCA}(x)$ is initialized with the id of the root r for any vertex x in T_r.
Main	Loop 1	c iterates through all customers of T_r.
	Step 1.1	$\mathrm{ZCA}(c)$ is determined as the unique admissible facility location for customer c with maximal distance to c.
	Step 1.2	$\mathrm{zcn}(c)$ is set equal to 0.
	End Loop 1	
	Loop 2	x iterates through all non-leaves of T_r according to the post-order enumeration.
	Step 2.1	The children of x — y and z — are determined.
	Step 2.2	$\mathrm{ZCA}(x) = \min\{\mathrm{ZCA}(x), \min_{v \in \{y,z\},\ \mathrm{ZCA}(v) \neq v} \mathrm{ZCA}(v)\}$

Main	Step 2.3	$\begin{aligned} \mathrm{zcn}(x) &= \mathrm{zcn}(y) + \mathcal{I}(y, \mathrm{ZCA}(y)) + \\ &\quad \mathrm{zcn}(z) + \mathcal{I}(z, \mathrm{ZCA}(z)) \end{aligned}$
	Step 2.4	$\begin{aligned} \mathrm{ZCM}(x) &= \mathrm{ZCM}(y) \cup (\{y\} \cap \{\mathrm{ZCA}(y)\}) \cup \\ &\quad \mathrm{ZCM}(z) \cup (\{z\} \cap \{\mathrm{ZCA}(z)\}) \end{aligned}$
	End Loop 2	

where the indicator function $\mathcal{I}(a, b)$ is defined as

$$\mathcal{I}(a,b) = \begin{cases} 0 \text{ if } a \neq b \\ 1 \text{ if } a = b. \end{cases}$$

If the customer vertices are basically all leaves, as it is the case for access networks, then the total number of vertices visited during Loop 1 and 2 is the number of vertices in T_r. In any case loop 1 and loop 2 can be merged into one single loop through all nodes of T_r.

For customers who are located at leaves the vertex and value determined in loop 1 correspond to the definition of $\mathrm{ZCA}(c)$ and $\mathrm{zcn}(c)$. If c is a leaf, then T_c is a trivial tree. The zero cost median solution in T_c is a facility at c. Since this facility is not counted as a median, the zero cost number is zero. $\mathrm{ZCA}(c)$ is chosen properly according to the definition of ZCA, since c is the only customer in T_c which is assigned to the facility at c in the zero cost median solution of T_c. The zero cost median solution $\mathrm{ZCM}(c)$ which corresponds to the zero cost ancestor of c is easily determined since it is given by the empty set.

For customers who are not leaves the choice may be wrong. However, the initial value $\mathrm{ZCA}(c)_{initial}$ contributes to the correction of $\mathrm{ZCA}(c)$ during Step 2.2. The value of $\mathrm{zcn}(c)$ is later corrected independently of the initial setting.

The following theorem states the correctness of the algorithm and describes how the locations of the ZCM for T_r are found.

Theorem 2 (Det ZCM, ZCN and ZCA) *Algorithm* ***Det ZCM, ZCN and ZCA*** *determines the correct values of the functions $zcn(x)$ and $ZCA(x)$ for all vertices x in T_r and delivers the zero cost medians for the tree T_x. They are identified as the fixed-points of the function $ZCA(.)$ over the set $vert(T_x) - \{x\}$:*

$$ZCM(x) = \{v \in vert(T_x) - \{x\} : v = ZCA(v)\}$$

Proof:
As mentioned subsequent to the algorithm the theorem is true for leaves of the tree T_r.

Since the desired values for any non-leaf x are recursively computed in steps 2.2, 2.3 and 2.4 by using the respective values of the children of x, the correctness of the theorem is proved by assuming that the stated properties hold for y and z, i.e. for v equal to y or z

- the zero cost number $\mathrm{zcn}(v)$ is correctly computed,
- $\mathrm{ZCA}(v)$ — the zero cost ancestor of v — is correctly determined, i.e. as the furthest ancestor of v for which a zero cost solution in T_v exists which is also a zero cost solution in $T_v^{\mathrm{ZCA}(v)}$,
- and this zero cost solution is identified by those vertices w in $T_v - \{v\}$ for which $\mathrm{ZCA}(w) = w$.

The proof of this theorem is a repeated application of Proposition 9 and the tree decomposition strategy which is presented in Section 2.3.7: horizontal and vertical decomposition steps.

With respect to zero cost solutions the horizontal decomposition implies that any zero cost solution ZCM in T_x can be divided into two disjoint sets which are zero cost solutions in T_y^x and T_z^x — where y and z are the children of x — simply by restricting the given zero cost solution to the respective subtree, i.e. $\mathrm{ZCM}_y = \mathrm{vert}(T_y^x) \cap \mathrm{ZCM}$ and $\mathrm{ZCM}_z = \mathrm{vert}(T_z^x) \cap \mathrm{ZCM}$. In turn, the union of any two zero cost solutions for T_y^x and T_z^x is a zero cost solution in T_x.

A similar property can be associated with the vertical decomposition step. A zero cost solution of the tree T_x^u is either a zero cost solution of T_x minus a facility at vertex x, or it is a zero cost solution of the tree $T_y^u \cup T_z^u$, i.e. a facility at x is not necessary. Again in turn, if there exist a zero cost solution in T_x^u which does not situate a facility at vertex x, then the union of any two zero cost solutions of T_y^u and T_z^u gives a zero cost solution in T_x^u.

It is clear that the computation of the set $\mathrm{ZCM}(x)$ in Step 2.4 of the algorithm

$$\mathrm{ZCM}(x) = \mathrm{ZCM}(y) \cup (\{y\} \cap \{\mathrm{ZCA}(y)\}) \cup \mathrm{ZCM}(z) \cup (\{z\} \cap \{\mathrm{ZCA}(z)\})$$

coincides with the definition of $\mathrm{ZCM}(x)$ as it is given in the theorem

$$\mathrm{ZCM}(x) = \{v \in \mathrm{vert}(T_x) - \{x\} : v = \mathrm{ZCA}(v)\},$$

since $\mathrm{ZCM}(y)$ and $\mathrm{ZCM}(z)$ are by assumption the fixed points of the zero cost ancestor function inside of T_y and T_z respectively. The vertices y and z are added to $\mathrm{ZCM}(x)$, only if they are identical with their respective zero cost ancestor.

It has to be verified, that the so defined set $\mathrm{ZCM}(x)$ constitutes a zero cost median for T_x and $T_x^{\mathrm{ZCA}(x)}$.

By assumption the set $\mathrm{ZCM}(y)$ is a zero cost solution of T_y and $T_y^{\mathrm{ZCA}(y)}$. Hence, according to Proposition 9 the set

$$\mathrm{ZCM}(y) \cup (\{y\} \cap \{\mathrm{ZCA}(y)\})$$

is a zero cost median of the tree T_y^x. Since, if $y = \mathrm{ZCA}(y)$, then x is a true ancestor of the zero cost ancestor of y and according to Equation (2.20) a facility at y becomes necessary.

Otherwise — if $y \neq \mathrm{ZCA}(y)$ — x is a descendent of $\mathrm{ZCA}(y)$ and no facility is needed at y according to Equation (2.19). By symmetry the same is true for the second child z of x. From the observation about zero cost solutions and horizontal decomposition it follows that the set $\mathrm{ZCM}(x)$ is a zero cost median in T_x.

As a first consequence the computation of the zero cost number $\mathrm{zcn}(x)$ in Step 2.3

$$zcn(x) = \mathrm{zcn}(y) + \mathcal{I}(y, \mathrm{ZCA}(y)) + \mathrm{zcn}(z) + \mathcal{I}(z, \mathrm{ZCA}(z))$$

turns out to be correct.

Next, the candidate for the zero cost ancestor of x is introduced based on the following three vertices:

$$u_x = \begin{cases} \text{admissible facility location} \\ \text{of greatest path distance to } x, & \text{if } x \text{ is a customer vertex} \\ r, & \text{otherwise} \end{cases}$$

$$u_y = \begin{cases} \mathrm{ZCA}(y) & \text{if } y \neq \mathrm{ZCA}(y) \\ r, & \text{otherwise} \end{cases}$$

$$u_z = \begin{cases} \mathrm{ZCA}(z) & \text{if } z \neq \mathrm{ZCA}(z) \\ r, & \text{otherwise} \end{cases}$$

Clearly, these three vertices lie on the path which connects x to r. Consequently, one of them is the descendent of all of them and lies closest to x. This vertex will be denoted by u_{ZCA} and is the candidate for the zero cost ancestor of x. If the vertices of the tree T_r are enumerated according to a post-order traversal, then u_{ZCA} can be determined as the minimum of the three vertices u_x, u_y and u_z.

Next, it is shown that the set $\mathrm{ZCM}(x)$ provides a zero cost assignment in $T_x^{u_{\mathrm{ZCA}}}$, i.e. all customers in this tree can be assigned to some vertex in $\mathrm{ZCM}(x)$ or to u_{ZCA} at zero cost:

- If x is a customer site, then it can be assigned to u_{ZCA} at zero cost, since u_{ZCA} is a descendent of u_x and therefore an admissible facility location for x.
- In case $y = \mathrm{ZCA}(y)$ every customer c from the tree T_y is assigned to a facility in the set $\mathrm{ZCM}(y) \cup \{y\}$ at zero cost, since $\mathrm{ZCM}(y)$ constitutes a zero cost solution in T_y. Otherwise, if $y \neq \mathrm{ZCA}(y)$, then u_{ZCA} is a descendent of $\mathrm{ZCA}(y)$ and it follows from Proposition 9 (2.19) that $\mathrm{ZCM}(y)$ is a zero cost solution in the tree $T_y^{u_{\mathrm{ZCA}}}$. In any case, every customer c is assigned to some facility in

  $$\mathrm{ZCM}(y) \cup (\{y\} \cap \{\mathrm{ZCA}(y)\})$$

  which is a subset of $\mathrm{ZCM}(x)$ by construction.
- By symmetry the same argument holds for all customers in T_z, such that every customer is assigned to some facility in

  $$\mathrm{ZCM}(z) \cup (\{z\} \cap \{\mathrm{ZCA}(z)\})$$

  which is also a subset of $\mathrm{ZCM}(x)$ by construction.

This proves that ZCM(x) provides a zero cost assignment in $T_x^{u_{\mathrm{ZCA}}}$. Proposition 9 implies that zero cost medians in $T_x^{u_{\mathrm{ZCA}}}$ use at least as many facilities as zero cost medians in T_x do. Since ZCM(x) is a zero cost median in T_x, it follows that ZCM(x) is also a zero cost median in $T_x^{u_{\mathrm{ZCA}}}$. Consequently, u_{ZCA} is a descendent of ZCA(x).

The proof concludes by showing that the zero cost ancestor of x is a descendent of u_x, u_y and u_z and hence, identical with u_{ZCA}.

- (A customer site at vertex x) A zero cost median of $T_x^{\mathrm{ZCA}(x)}$ which is also a zero cost median in T_x does not contain vertex x. So, if x is a customer site, then it must be possible to assign x to the facility at ZCA(x) at zero cost. Consequently, ZCA(x) is a descendent of u_x.

- (Vertical decomposition) By definition of the zero cost ancestor of x there exists a zero cost median ZCM in T_x which is also a zero cost median in $T_x^{\mathrm{ZCA}(x)}$. Since ZCM is a zero cost median in T_x, it does not contain x. Therefore, the set ZCM can be written as the union of the two disjoint subsets

  $$\mathrm{ZCM}_y = \mathrm{ZCM} \cap \mathrm{vert}(T_y)$$

  and

  $$\mathrm{ZCM}_z = \mathrm{ZCM} \cap \mathrm{vert}(T_z).$$

- (ZCA(x) with respect to ZCA(y)) Since ZCM is a zero cost median in T_x and $T_x^{\mathrm{ZCA}(x)}$, the set ZCM_y is a zero cost median in $T_y^{\mathrm{ZCA}(x)}$. Assuming now that ZCA(x) is a true ancestor of ZCA(y), then by Proposition 9 Equation (2.20) a zero cost solution in $T_y^{\mathrm{ZCA}(x)}$ is given by ZCM(y) $\cup \{y\}$. Since the zero cost number is unique, the equation

  $$\mid \mathrm{ZCM}_y \mid = \mid \mathrm{ZCM}(y) \mid +1$$

  holds. Consequently, the zero cost number of T_y^x is equal to

  $$\mid \mathrm{ZCM}(y) \mid +1.$$

  Hence, ZCM(y) is not a zero cost median of T_y^x. So, ZCM(y) $\cup \{y\}$ is a zero cost median of T_y^x which implies based on Proposition 9 that x like ZCA(x) is a true ancestor of ZCA(y). In other words:

  $$\text{If } \mathrm{ZCA}(x) \text{ is a true ancestor of } \mathrm{ZCA}(y), \text{ then } \mathrm{ZCA}(y) = y.$$

  Conversely, if ZCA(y) $\neq y$, then ZCA(x) is equal to or a descendent of ZCA(y). In any case, ZCA(x) is a descendent of u_y.

- By symmetry the same argument holds for ZCA(x), ZCA(z) and u_z.

This concludes the proof.
□

2.5.13 Recursion formula for the zero cost number

Before turning to the general recursion formula for the KMP it is of interest to study the behavior of the zero cost number for trees of type T_x^u with fixed root u and varying vertex x. The main interest lies in the relation between the zero cost numbers for the tree T_x^u and the corresponding trees for the children of x: T_y^u and T_z^u.

For the root r and its children it was already argued that the set of zero cost medians of T_r splits into two subsets, each one of them contained in one of the two subtrees T_y and T_z, such that both of these sets constitute zero cost medians for the two subtrees.

Unsurprisingly, it can be shown for arbitrary subtrees T_x by repeating this argument that

Proposition 10 (ZCN P1) *For any vertex $x \in T_r$ and its children y and z the equation*

$$zcn(x) = ZCN(T_y^x) + ZCN(T_z^x) \tag{2.21}$$

holds.

Unfortunately, this results does not generalize to the equation

$$\mathrm{ZCN}(T_x^u) = \mathrm{ZCN}(T_y^u) + \mathrm{ZCN}(T_z^u).$$

The vertical decomposition step which removes x from the tree T_x^u and produces two new trees T_y^u and T_z^u may remove a facility from site x because the vertex is removed, and at the same time open two new facilities at vertices y and z.

The following proposition collects the conditions and shows the exact relationship between the three zero cost numbers.

Proposition 11 (ZCN P2) Recursion for ZCN
For a pair of vertices x and u of the rooted directed tree T_r, where u is an ancestor of x, the set $P_{x,u}$ denotes the set of vertices of the path which connects x and u excluding the vertex u itself[43].

a) If $\mathrm{ZCA}(y) \in P_{x,u}$ and $\mathrm{ZCA}(z) \in P_{x,u}$, then

$$\mathrm{ZCN}(T_x^u) = \mathrm{ZCN}(T_y^u) + \mathrm{ZCN}(T_z^u) - 1. \tag{2.22}$$

b) If $\mathrm{ZCA}(x) \in P_{x,u}$,
$\mathrm{ZCA}(y) \notin P_{x,u}$ and
$\mathrm{ZCA}(z) \notin P_{x,u}$, then

x is a customer site with $d(x,u) > 0$ and

$$\mathrm{ZCN}(T_x^u) = \mathrm{ZCN}(T_y^u) + \mathrm{ZCN}(T_z^u) + 1. \tag{2.23}$$

[43] So, $P_{x,x} = \emptyset$.

c) Otherwise,

$$\mathrm{ZCN}(T_x^u) = \mathrm{ZCN}(T_y^u) + \mathrm{ZCN}(T_z^u). \tag{2.24}$$

Proof:
The proof utilizes Proposition 9 and Proposition 10. Several cases have to be inspected:

Case: $\mathrm{ZCA}(\mathbf{x}) \in \mathbf{P_{x,u}}$

Because of Proposition 9 this assumption implies that

$$\mathrm{ZCN}(T_x^u) = \mathrm{zcn}(x) + 1$$

which together with Equation 2.21 from Proposition 10 leads to

$$\mathrm{ZCN}(T_x^u) = \mathrm{ZCN}(T_y^x) + \mathrm{ZCN}(T_z^x) + 1.$$

The question is now, what is the relation between $\mathrm{ZCN}(T_v^x)$ and $\mathrm{ZCN}(T_v^u)$ where v is anyone of the children of x? If additionally to $\mathrm{ZCA}(x)$ also $\mathrm{ZCA}(v) \in P_{x,u}$, then

$$\mathrm{ZCN}(T_v^x) = \mathrm{zcn}(v) = \mathrm{ZCN}(T_v^u) - 1,$$

since x is a descendent and u is a true ancestor of $\mathrm{ZCA}(v)$. If this is the case for both children of x, then

$$\begin{aligned} \mathrm{ZCN}(T_x^u) &= \mathrm{ZCN}(T_y^x) + \mathrm{ZCN}(T_z^x) + 1 \\ &= \mathrm{ZCN}(T_y^u) - 1 + \mathrm{ZCN}(T_z^u) - 1 + 1 \\ &= \mathrm{ZCN}(T_y^u) + \mathrm{ZCN}(T_z^u) - 1 \end{aligned}$$

which gives Equation (2.22) and corresponds to case a) of Proposition 11.

If for both children $\mathrm{ZCA}(v) \notin P_{x,u}$, then $\mathrm{ZCN}(T_v^x) = \mathrm{ZCN}(T_v^u)$ and consequently

$$\begin{aligned} \mathrm{ZCN}(T_x^u) &= \mathrm{ZCN}(T_y^x) + \mathrm{ZCN}(T_z^x) + 1 \\ &= \mathrm{ZCN}(T_y^u) + \mathrm{ZCN}(T_z^u) + 1 \end{aligned}$$

which gives Equation (2.23) and corresponds to case b) of Proposition 11. Moreover, this constellation implies that $\mathrm{ZCA}(x) \neq \mathrm{ZCA}(v)$ for both the children of x. Since $\mathrm{ZCA}(x)$ is a true descendent of u and therefore necessarily a true descendent of the overall root r it follows from Algorithm 6 Step 0, 1.1 and 2.2 that x must be a customer site.

If only one of the zero cost ancestors of the children of x is contained in $P_{x,u}$, then

$$\begin{aligned} \mathrm{ZCN}(T_x^u) &= \mathrm{ZCN}(T_y^x) + \mathrm{ZCN}(T_z^x) + 1 \\ &= \mathrm{ZCN}(T_y^u) + \mathrm{ZCN}(T_z^u) - 1 + 1 \\ &= \mathrm{ZCN}(T_y^u) + \mathrm{ZCN}(T_z^u) \end{aligned}$$

which leads to Equation (2.24) and case c) of Proposition 11.

Case: $\mathrm{ZCA}(\mathbf{x}) \notin \mathbf{P_{x,u}}$

From Proposition 9 it follows that

$$\mathrm{ZCN}(T_x^u) = \mathrm{zcn}(x)$$

which together with Proposition 10 leads to

$$\mathrm{ZCN}(T_x^u) = \mathrm{ZCN}(T_y^x) + \mathrm{ZCN}(T_z^x).$$

According to Algorithm 6 Step 2.2 the zero cost ancestors of the children of x are either the children themselves or they are ancestors of $\mathrm{ZCA}(x)$. Since by assumption $\mathrm{ZCA}(x)$ is an ancestor of u, both zero cost ancestors — $\mathrm{ZCA}(y)$ and $\mathrm{ZCA}(z)$ — are not contained in $P_{x,u}$. So, either

$$\mathrm{ZCN}(T_v^x) = \mathrm{zcn}(v) = \mathrm{ZCN}(T_v^u)$$

because $\mathrm{ZCA}(v)$ is an ancestor of x and u or

$$\mathrm{ZCN}(T_v^x) = \mathrm{zcn}(v) + 1 = \mathrm{ZCN}(T_v^u),$$

for $\mathrm{ZCA}(v) = v$, but, in any case they are equal. So, if $\mathrm{ZCA}(\mathbf{x}) \notin \mathbf{P_{x,u}}$, then Equation (2.24) and case c) of Proposition 11 always hold.
□

An immediate consequence of this results is, that the zero cost numbers are monotonously decreasing with descendants.

Corollary 3 (Monotonicity of ZCN) *For any descendant x' of a vertex x it is true that*

$$ZCN(T_x^u) \geq ZCN(T_{x'}^u). \tag{2.25}$$

The conclusion is obvious for Equation (2.23) and (2.24), since zero cost numbers are non-negative. In Equation (2.22) the children's zero cost numbers — $\mathrm{ZCN}(T_y^u)$ and $\mathrm{ZCN}(T_z^u)$ — are both at least 1, which follows from the fact that both zero cost ancestors are true descendants of u.

2.5.14 Another recursion formula to solve KMP

The goal is to find the k-median solution in a directed tree by starting at a zero cost median solution of the tree and approaching the k-median solution by decreasing the number of medians.

$$\mathrm{Cost}_r^k(r) = \mathrm{Cost}_r^{k_0-J}(r)$$

has to be solved, where k_0 is the $\mathrm{ZCN}(T_r)$ and k is a non negative integer less than k_0. The recursion formula for the KMP stated in Section 2.3.8 is applied which leads to

$$\mathrm{Cost}_r^{k_0-J}(r) = \min_{s+t=k_0-J} \left\{ \mathrm{Cost}_y^s(r) + \mathrm{Cost}_z^t(r) \right\}$$

in the special case of the root r and its children y and z (Proposition 2). However, because of the recursive call to $\mathrm{Cost}_y^s(r)$ and $\mathrm{Cost}_z^t(r)$, cost values for trees of type T_x^r have to be calculated for any vertex x in T_r and some value of j as

$$\mathrm{Cost}_x^j(r) = \min \begin{cases} \min\limits_{s+t=j} \left\{\mathrm{Cost}_y^s(r) + \mathrm{Cost}_z^t(r)\right\} + w(x) * d(x,r) \\ \mathrm{Cost}_x^{j-1}(x) \end{cases}$$

according to the recursion formula. The value of j is not clear at this point and will be determined later. Finally, because of the second line of the stated formula it becomes evident that the recursion formula has to be studied for general trees of type T_x^u where u is any ancestor of x:

$$\mathrm{Cost}_x^j(u) = \min \begin{cases} \min\limits_{s+t=j} \left\{\mathrm{Cost}_y^s(u) + \mathrm{Cost}_z^t(u)\right\} + w(x) * d(x,u) \\ \mathrm{Cost}_x^{j-1}(x) \end{cases}.$$

For j equal to or greater than the zero cost number of the tree T_x^u there is no need to call the recursion formula, because the desired assignment cost is known. It is zero. As a consequence of Proposition 9 the zero cost number $\mathrm{ZCN}(T_x^u)$ is determined based on $\mathrm{zcn}(x)$ and $\mathrm{ZCA}(x)$ by

$$j_0 := \mathrm{ZCN}(T_x^u) = \begin{cases} \mathrm{zcn}(x), & \text{if } u \text{ is a descendant of } \mathrm{ZCA}(x) \\ \mathrm{zcn}(x)+1, & \text{if } u \text{ is a true ancestor of } \mathrm{ZCA}(x). \end{cases}$$

Consequently, the recursion formula can be rewritten for an appropriate non negative value of j as

$$\mathrm{Cost}_x^{j_0-j}(u) = \min \begin{cases} \min\limits_{s+t=j_0-j} \left\{\mathrm{Cost}_y^s(u) + \mathrm{Cost}_z^t(u)\right\} + w(x) * d(x,u) \\ \mathrm{Cost}_x^{j_0-j-1}(x) \end{cases}.$$

The zero cost numbers for T_y^u and T_z^u — p_0 and q_0 respectively — are easily determined, too (Theorem 2).

It was already argued for the special case when $u = r$, $j = 1$ and y and z are the children of r that it suffices to choose s and t less or equal to p_0 and q_0 respectively. It is also true in the general case that s and t can be limited by the corresponding zero cost number.

If s can be and is chosen[44] beyond p_0, then $\mathrm{Cost}_y^s(u)$ cannot decrease anymore because its value is zero already (Proposition 3). Therefore, this choice of s does not produce any advantage for the $(j_0 - j)$-median solution in the union of the two trees T_y and T_z.

At the same time, if s is chosen beyond p_0, then t has to be chosen less or equal to q_0. This follows from Proposition 11 — the recursion formula for ZCN. According to this result $j_0 \le p_0 + q_0 + 1$. By construction $s + t = j_0 - j$. The combination leads to

$$\begin{aligned} s + t &\le p_0 + q_0 + 1 - j \\ t &\le p_0 - s + q_0 + 1 - j \\ t &\le q_0 + 1 - j \quad for\ s \ge p_0. \\ t &\le q_0 \quad for\ j \ge 1. \end{aligned}$$

[44] Such a choice of s may force t to be chosen below 0 which is not permissible.

Clearly, if the value of s is further increased, the value of t decreases by the same amount. But, then t becomes less than q_0 and the assignment cost $\mathrm{Cost}_z^t(u)$ are positive because of the definition of the zero cost number q_0 (Definition 16). As a consequence (Proposition 3 Equation (2.14)) the $\mathrm{Cost}_z^t(u)$ continues to grow for decreasing t which implies a disadvantageous contribution of the t-median solution of T_z.

A combination of a non-advantageous contribution of the s-median solution in T_y and a disadvantageous contribution of the t-median solution in T_z leads to an increasing sum $\mathrm{Cost}_y^s(u) + \mathrm{Cost}_z^t(u)$. Consequently, for such choices of s and t the desired minimum will certainly not be achieved. It follows,

$$s \leq p_0 \quad \text{and} \quad t \leq q_0.$$

The recursion formula can be rewritten again:

$$\mathrm{Cost}_x^{j_0-j}(u) = \min \begin{cases} \min\limits_{\substack{s+t=j_0-j \\ s \leq p_0 \wedge t \leq q_0}} \left\{\mathrm{Cost}_y^s(u) + \mathrm{Cost}_z^t(u)\right\} + w(x) * d(x,u) \\ \mathrm{Cost}_x^{j_0-j-1}(x) \end{cases}.$$

As usual s and t are non-negative integers.

Repeating the strategy from Section 2.5.10, a transformation of variables is applied to the recursion formula by setting

$$s = p_0 - p \text{ and } t = q_0 - q.$$

Since s and t are both non-negative integers, p and q are integers and

$$p \leq p_0 \text{ and } q \leq q_0,$$

and since s is less or equal to p_0 and t is less or equal to q_0, p and q are both non-negative integers, too.

The equation $s + t = j_0 - j$ transforms to

$$p_0 - p + q_0 - q = j_0 - j$$

which becomes

$$p + q = j + p_0 + q_0 - j_0.$$

Because of the recursion formula for ZCN (Proposition 11, the difference $p_0 + q_0 - j_0$ does not necessarily equal zero. It may be -1, 0 or 1. Hence, this difference can not be removed from the equation. To simplify notation the difference is substituted by $\Delta_0 = p_0 + q_0 - j_0$.

So, the final version of the recursion formula for the k-median problem which reflects the zero cost numbers of the subtree in question can be formulated:

Theorem 3 (Recursion formula for the KMP reflecting ZCNs)
The zero cost numbers of T_x^u, T_y^u and T_z^u are denoted by j_0, p_0 and q_0 respectively. Then for a positive integer $j < j_0$ the cost of the $(j_0 - j)$-median solution can be calculated by

$$\mathrm{Cost}_x^{j_0-j}(u) = \min \begin{cases} \min\limits_{p+q=j+\Delta_0} \left\{\mathrm{Cost}_y^{p_0-p}(u) + \mathrm{Cost}_z^{q_0-q}(u)\right\} + w(x) * d(x,u) \\ \mathrm{Cost}_x^{j_0-j-1}(x) \end{cases}, \qquad (2.26)$$

where $\Delta_0 = p_0 + q_0 - j_0$. *Additionally,* $0 \leq p \leq p_0$ *and* $0 \leq q \leq q_0$ *has to be observed.*

For $j = j_0$ *the cost of the* 0*-median solution of* T_x^u *is calculated directly as the sum of weighted distances of all customers in* T_x *to* u*:*

$$\mathrm{Cost}_x^0(u) = \sum_{\substack{v \in \mathit{vert}(T_x) \\ w(v) > 0}} w(v) * d(v,u). \tag{2.27}$$

It has to be noted for the subsequent call of the recursion formula by $\mathrm{Cost}_x^{j_0-j-1}(x)$ that the exponent $j_0 - j - 1$ is not stated in a correct way. The recursion formula asks for the difference between the zero cost number of the tree in question and the number of medians which are supposed to be removed from the tree. But, j_0 is not necessarily the zero cost number of T_x which may be 1 less than j_0 and depends on the relation between u and the zero cost ancestor $\mathrm{ZCA}(x)$ of x (Proposition 9). The correct expression for the subsequent call in the second part of Recursion Formula (2.26) is therefore

$$\mathrm{zcn}(x) - (j + 1 + \mathrm{zcn}(x) - j_0). \tag{2.28}$$

Yet, the Recursion Formula (2.26) of Theorem 3 does not initiate the desired solution algorithm by descending iteration. It is actually nothing but a different notation of Recursion Formula (2.11). Hence, their behavior is identical, i.e. they ask for j-median solutions of smaller and smaller size, until eventually 0-median solutions are called.

According to the first part of Recursion Formula (2.26) if the value of Δ_0 is positive, then the algorithm attempts to remove one more facility from T_y^u and T_z^u than there are supposed to be removed from T_x^u. According to the second part if the expression $-1 + j_0 - \mathrm{zcn}(x)$ in Equation (2.28) is not zero, i.e. if $j_0 = \mathrm{zcn}(x)$, then the number of facilities in T_x has to be reduced by $j + 1$ compared to its zero cost number. This increased reduction of the number of facilities is then inherited by subsequent trees and may be increased even further.

Moreover, for many trees either the first or the second part of the recursion formula attempts to remove more facilities than the original call asks for. It follows from Proposition 9 that $j_0 = \mathrm{zcn}(x)$ if and only if the zero cost ancestor of x is equal to u or an ancestor of u, i.e. if x and u are relatively close to each other. Consequently, the second part of the recursion formula tries to remove one more facility from T_x.

If the two vertices are far enough apart from each other, i.e. if the zero cost ancestor of x is a true descendent of u, the effect vanishes. There have to be as many facilities to be removed from T_x, than there are to be removed from T_x^u. However, in this case at least one of the children of x carries an open facility in the zero cost solution of $T_y^u \cup T_z^u$, since the zero cost ancestor of x coincides at least with the zero cost ancestor of one of them. If it is necessary to open a facility at the second child's location, too, then case a) of Proposition 11 is entered. Equation (2.22) forces Δ_0 to be equal to 1 and one more facility has to be removed from the children's trees.

For an iterative algorithm as it was described in Section 2.3.9, which depends on the k-median solutions of the children of a vertex x to solve the corresponding KMPs in T_x^u, it is necessary to know in advance how many KMPs have to be solved in total for the given subtree. This knowledge determines the size of the **J - Loop** described in Section 2.3.9. But, Recursion Formula (2.26) does not state how many facilities have to be removed from an arbitrary subtree T_x^u to enable the solution of the $(k_0 - k)$-median problem in the original tree T_r, at least not if the iteration starts at the leaves and works bottom-up.

Another question is why could it at all be necessary to remove more facilities from an arbitrary subtree than from the original tree?

2.5.15 Solving KMP by descending iteration (II)

To simplify notation a little bit the concepts of removing facilities from a solution and of reducing the number of medians is formalized by the following definition.

Definition 18 (Reduction of medians) *For a given directed and rooted tree T with zero cost number k_0 the **k-reduction** is the solution of the $(k_0 - k)$-median problem. To **reduce a given solution of the KMP in T by k medians/facilities** means to solve a KMP for the number of medians of the original solution minus k, where it is assumed that k is always less or equal to the number of medians in question.*

The clue of overcoming the problem which was addressed at the end of the previous section is that a k-reduction in the tree T does not utilize a $(k+2)$-reduction or higher in any of its subtrees. Therefore, a second kind of cost function is associated with k-median solutions, or better, with k-reductions for the subtrees which are generated by the recursion formulas.

Definition 19 (${}^k\mathrm{Cost}$) *For any subtree T_x^u of T_r (which includes the case T_x for $x = u$) with zero cost numbers $j_0 = ZCN(T_x^u)$ and $k_0 = ZCN(T_r)$, k and j non negative integers both less or equal to k_0 the function ${}^k\mathrm{Cost}$ is defined*

for $j > j_0$ or $j \geq k+2$ by

$$ {}^k\mathrm{Cost}_x^{j_0-j}(u) = +\infty. \tag{2.29}$$

If $j = j_0$ and $j < k+2$ then

$$ {}^k\mathrm{Cost}_x^0(u) = \sum_{\substack{v \in vert(T_x) \\ w(v) > 0}} w(v) * d(v,u). \tag{2.30}$$

If $j < j_0$ and $j < k+2$

$$ {}^k\mathrm{Cost}_x^{j_0-j}(u) = \min \left\{ \begin{array}{l} \gamma^j(u,y,z) + w(x) * d(x,u) \\ {}^k\mathrm{Cost}_x^{j_0-j-1}(x) \end{array} \right. \tag{2.31}$$

where

$$ \gamma^j(u,y,z) = \min_{p+q=j+\Delta_0} \left\{ {}^k\mathrm{Cost}_y^{p_0-p}(u) + {}^k\mathrm{Cost}_z^{q_0-q}(u) \right\}, \tag{2.32}$$

p_0 and q_0 are the zero cost numbers of T_y^u and T_z^u respectively, $\Delta_0 = p_0 + q_0 - j_0$ and additionally, $0 \leq p \leq p_0$ and $0 \leq q \leq q_0$ has to be observed.

The essential difference between the definition of the ${}^k\text{Cost}$ function and the behavior of the function Cost from Equation (2.26) is that ${}^k\text{Cost}$ is not defined for $(k+2)$-reductions or higher. Formally the definition of the ${}^k\text{Cost}$ function is just a trick to reformulate the Recursion Formula (2.26) in such a way that $(k+2)$-reductions and higher are not considered by setting their values to plus infinity. So, they will never appear as the minimizers in Equation (2.31) nor in (2.32). ${}^k\text{Cost}$ values for $(k+1)$-reductions and lower are calculated without reference to ${}^k\text{Cost}$ values for $(k+2)$-reductions and higher. The usefulness of this trick, however, is expressed by the following theorem. The proof is deferred to p. 147 below.

Theorem 4 (Identity of ${}^k\text{Cost}$ and Cost function) *For any non negative integer j which is less or equal to k and less or equal to j_0*

$$\text{Cost}_x^{j_0-j}(u) = {}^k\text{Cost}_x^{j_0-j}(u) \tag{2.33}$$

for any subtree T_x^u of T_r.
For the $(k+1)$-reduction in T_x^u it holds that

$$\text{Cost}_x^{j_0-k-1}(u) \leq {}^k\text{Cost}_x^{j_0-k-1}(u) < +\infty. \tag{2.34}$$

If the value of the ${}^k\text{Cost}$ function is strictly larger than the Cost value in Inequality (2.34), then because the $(k+1)$-reduction in T_x^u makes use of a $(k+2)$-reduction in some subtree T_v. (2.35)

In other words Theorem 4 states that the optimal assignment cost for k-reductions and lower can be computed without knowing the optimal assignment cost for $(k+2)$-reductions and higher.

But, what about $(k+1)$-reductions? According to Definition 19 Equation (2.31) they are explicitly computed and not automatically set to plus infinity. It will be shown that the ${}^k\text{Cost}$ values of $(k+1)$-reductions are always finite and not always correct, i.e not identical with the optimal assignment cost of the $(k+1)$-reduction. Furthermore, if ${}^k\text{Cost}$ and Cost value are not identical, then because the $(k+1)$-reduction makes use of a $(k+2)$-reduction in some subtree. In this case the ${}^k\text{Cost}$ value is strictly larger than the Cost value. But, if — as Theorem 4 implies — k-reductions make no use of $(k+2)$-reductions, then no $(k+1)$-reduction which utilizes a $(k+2)$-reduction can ever be the minimizer in Equation (2.31) and (2.32) for a k-reduction in T_x^u. Then, of course, the wrongly computed value of ${}^k\text{Cost}$ for this $(k+1)$-reduction will not really matter simply because it is too high.

But, there are k-reductions which utilize $(k+1)$-reductions. This can be illustrated by a simple example. A tree with root u which has got one child x, two *grandchildren* y and z and several more offspring thereafter provides the basic network. The zero cost median solution of this tree comprises 2 medians which are located at the two grandchildren of u. What is the 1-median reduction of this tree going to be?

Either the median at y or z is removed, or they are both removed and replaced by a median at x. The latter case, however, implies that a 2-reduction has to be determined in the subtree T_x of T_u. Removing 2 medians from a subtree implies that there is a surplus of one facility

compared to the number of facilities which have to be removed from T_u originally. This surplus has to be compensated by inserting a facility somewhere else in the tree T_u. There is no better location than x.

The same network can be used to demonstrate that it is not useful to apply a 3-reduction to T_x to produce a 1-reduction in T_u. Assuming now that there are at least 3 medians contained in T_x and 3 medians are removed from T_x, then a surplus of 2 facilities is removed from T_u. One of these can be compensated by opening a facility at x. The second has to be compensated by opening a facility in $T_u - T_x$. But, this tree contains only vertex u and there is no need for an additional facility in this tree. Consequently, this strategy will not produce the desired 1-reduction in T_u. But, even if $T_u - T_x$ is a large tree, there are already enough facilities contained in it to provide a zero cost solution. An additional facility will not improve the assignment cost.

The following proposition generalizes these thoughts.

Proposition 12 (^{k}Cost P1) *Given a tree T_u with zero cost number j_0, a vertex x in T_u with zero cost number l_0, a positive integer $j < l_0$, then a j-reduction in T_u removes at most*

- *j facilities from inside of T_x, if there exists a zero cost solution in T_u which opens a facility at x,*
- *$j+1$ facilities from inside of T_x, if there exists no zero cost solution in T_u which opens a facility at x.*

Hence, a j-reduction in T_u makes no use of $(j+2)$-reductions or higher in any of its subtrees T_x.

Proof:
The cases $j = 0$ and $j \geq l_0$ are trivial, since in these cases at most j facilities are removed from T_x anyway.

A j-reduction in T_u places f_1 facilities in T_x and f_2 facilities in $T_u - T_x$. Consequently,

$$j_0 - j = f_1 + f_2.$$

Since the goal of this proposition is to express the reduction of facilities in T_x with respect to its zero cost number, it is better set up the balance equation in the following way:

$$j_0 - j = l_0 - \delta_{l_0} + t_0 - \delta_{t_0}. \tag{2.36}$$

The zero cost number of $T_u - T_x$ is denoted by t_0. So, $t_0 - \delta_{t_0}$ is supposed to express the number of facilities of the j-reduction in T_u which are contained in $T_u - T_x$. Equivalently, the term $l_0 - \delta_{l_0}$ counts the number of facilities of the j-reduction of T_u which lie inside of T_x, not counting an eventual facility which is placed by the j-reduction at vertex x. If the j-reduction under consideration opens a facility at vertex x, then the balance Equation (2.36) is wrong and

$$j_0 - j = l_0 - \delta_{l_0} + t_0 - \delta_{t_0} + 1 \tag{2.37}$$

has to be used instead which accounts for the additional facility placed at vertex x and is not counted especially by $l_0 - \delta_{l_0}$.

In any case δ_{l_0} and δ_{t_0} are non negative integers. Otherwise, the j-reduction in T_u would situate more facilities in the respective trees, than their zero cost numbers ask for. Since $T_u - T_x$ is independent of T_x, any number of facilities higher than t_0 does not provide an improvement of the assignment cost for customers in this tree.

In T_x any assignment of its customers to more than l_0 facilities can be improved or at least replaced by a zero cost solution of T_x combined with a facility at x which utilizes a total of $l_0 + 1$ facilities.

So, equations (2.36) and (2.37) can be combined in

$$j + l_0 + t_0 - j_0 + \mathcal{I}(x, T_u, j) = \delta_{l_0} + \delta_{t_0} \text{ with } \delta_{l_0} \geq 0 \text{ and } \delta_{t_0} \geq 0, \tag{2.38}$$

where the expression $\mathcal{I}(x, T_u, j)$ indicates, whether there is a facility opened at vertex x by the j-reduction in T_u (value is equal to 1) or not (value is equal to 0).

The focus is now on the relation between the zero cost numbers t_0, l_0 and j_0. Zero cost solutions are unique with respect to the number of facilities they use. Hence, the zero cost number of T_u which is denoted by j_0 is a fixed integer.

The zero cost number of x which is $l_0 = \text{zcn}(x)$ counts the number of facilities of the zero cost solution of T_u which lie inside of T_x. This number is again independent of the specific zero cost solution in T_u, i.e. every zero cost solution of T_u places the same number of facilities inside of T_x (see Proposition 8). Therefore, l_0 is also a fixed integer and independent of a specific zero cost solution $ZCS(T_u)$.

Consequently, the l_0 facilities inside of T_x are also accounted for by j_0. It remains to investigate where exactly the remaining $j_0 - l_0$ facilities of the $ZCS(T_u)$ are located and how t_0 does relate to this number. One of them may by situated at vertex x and the others are certainly found in $T_u - T_x$.

If there exists a zero cost solution $ZCS(T_u)$ which places a facility at x, then the zero cost supplement of customers from $T_u - T_x$ is independent of the facilities inside of T_x and at x. In this case the facilities of $ZCS(T_u)$ which lie in $T_u - T_x$ provide an assignment with zero cost for all customers of $T_u - T_x$. Furthermore, such an assignment cannot be achieved with less facilities. Otherwise, the zero cost solution of T_u could be improved. Hence, t_0 is exactly the number of facilities of $ZCS(T_u)$ which are contained in $T_u - T_x$. Together with the facility at x

$$j_0 = l_0 + t_0 + 1 \tag{2.39}$$

states the correct relation between the three zero cost numbers for this case.

Substituting Equation (2.39) in Equation (2.38) leads to

$$j + j_0 - 1 - j_0 + \mathcal{I}(x, T_u, j) = \delta_{l_0} + \delta_{t_0} \tag{2.40}$$

which resolves to

$$j \geq \delta_{l_0} + \delta_{t_0}, \tag{2.41}$$

since the indicator function $\mathcal{I}(x, T_u, j)$ is at most equal to 1. This bounds the value of δ_{l_0} from above by j and proves the first part of Proposition 12.

If there exists no zero cost solution in T_u which situates a facility at x, then

$$t_0 = j_0 - l_0, \tag{2.42}$$

i.e. every zero cost median which is placed in $T_u - T_x$ by a zero cost solution of T_u is indispensable to provide a zero cost solution for $T_u - T_x$.

This is true, since if less than $j_0 - l_0$ facilities could be placed in $T_u - T_x$ to provide a zero cost solution there, then a zero cost assignment can be constructed in T_u by opening a facility at vertex x. This facility has to be opened, because there might be some customers inside of T_x who are actually assigned to a facility outside of T_x by $ZCS(T_u)$. Such an assignment would provide a zero cost assignment with at most j_0 facilities contradicting the assumption that there are no zero cost solutions in T_u which open a facility at x.

Again, substituting Equation (2.42) in Equation (2.38) leads to

$$j + j_0 - j_0 + \mathcal{I}(x, T_u, j) = \delta_{l_0} + \delta_{t_0} \tag{2.43}$$

which resolves to

$$j + 1 \geq \delta_{l_0} + \delta_{t_0}. \tag{2.44}$$

Consequently, the value of δ_{l_0} cannot exceed $j + 1$ which proves the second part of Proposition 12.
□

Proof of Theorem 4:

If x is a leaf, then the only meaningful reduction — if at all — is the 1-reduction. In this case the 1-reduction corresponds to the 0-median solution. For both kinds of cost functions the 0-median solution is computed directly without recursion and this is done in the same way (compare Equation (2.27) and Equation (2.30)). Consequently, Theorem 4 is true for any tree of type T_x^u where x is a leaf.

Next, a non-leaf x with children y and z is considered. It is assumed that Theorem 4 holds for the children with respect to any of their ancestors u. The set of their true ancestors coincides with the union of x and the true ancestors of x.

The proof of Theorem 4 consists of two parts, each of which following three steps which are basically identical for both parts. In the first part the correctness of Theorem 4 is proven for trees of type T_x which is a prerequisite for the proof in general tress of type T_x^u during the second part.

[k]Cost function and Cost function are linked to the values of the children of x by means of Recursion Formula (2.31), (2.32) and (2.26) respectively. For the first part of the proof —

where $u = x$ — the second line of Recursion Formula (2.31) and (2.26) respectively can be ignored, since neither Cost value nor ${}^k\text{Cost}$ value can be identical for the j-reduction and the $(j+1)$-reduction in a tree T_x. In general trees T_x^u Cost and ${}^k\text{Cost}$ function are additionally linked to their respective values in T_x which actually makes it necessary to distinguish two parts during the proof.

The **first step** of both parts is to prove Theorem 4 for reductions of degree less than k. According to the Recursion Formula (2.31), (2.32) and (2.26) the ${}^k\text{Cost}$ function and the Cost function respectively make use of reductions of degree at most k.

First part: Since the cited recursion formulas refer only to ${}^k\text{Cost}$ and Cost values of the children of x, for which the values are assumed to be identical, ${}^k\text{Cost}$ values can simply be replaced by the corresponding Cost values in Equation (2.31) by which it is transformed into Equation (2.26). This establishes the correctness of Theorem 4 for this case.

Second part: The argument is exactly same. Additionally, the identities of Cost and ${}^k\text{Cost}$ values for reductions in T_x of degree at most k are used. But, they are established by the first part.

In the **second step** that part of Theorem 4 is provided which is concerned with $(k+1)$-reductions. To see that the ${}^k\text{Cost}$ value is finite the key expression of the first line of Recursion Formula (2.31) is studied. It reads as follows:

$$\gamma^{k+1}(u,y,z) = \min_{p+q=k+1+\Delta_0} \left\{ {}^k\text{Cost}_y^{p_0-p}(u) + {}^k\text{Cost}_z^{q_0-q}(u) \right\}.$$

If Δ_0 is equal to 0, the minimum is formed over a non empty set of finite values. The application of Equation (2.31) implies that $k+1 < j_0$, the zero cost number of T_x^u. From Proposition 11 it follows that $k+1 \leq p_0+q_0$. Consequently, p and q can be chosen as non negative integers which add up to $k+1$. Hence, the set over which the minimum is formed is not empty. By assumption the involved ${}^k\text{Cost}$ values are all finite. So, the resulting minimum has to be finite, too. Since $\gamma^{k+1}(u,y,z)$ is finite, the value of ${}^k\text{Cost}_x^{j_0-k-1}(u)$ which results from Equation (2.31) is also finite.

If Δ_0 is equal to 1, then according to Proposition 11 Equation (2.22) there are facilities situated at vertices y and z in the zero cost solution of $T_y^u \cup T_z^u$. In other words, both zero cost numbers — p_0 and q_0 — are at least 1. Consequently, $p = k+2$ and $q = 0$ or $p = 0$ and $q = k+2$ cannot be the only possible choices for p and q. Both indices — p and q — can be chosen greater than zero which forces both of them to be strictly less than $k+2$. From here the arguments are the same as before and the finiteness of ${}^k\text{Cost}_x^{j_0-k-1}(u)$ follows.

To conclude the proof of Inequality (2.34) all ${}^k\text{Cost}$ values in Equation (2.31) are replaced by their corresponding Cost value which results in Equation (2.26) again. By assumption and — in case of part 2 — by proof of part 1 these values are less or equal to the ${}^k\text{Cost}$ values.

Finally, Statement (2.35) has to be verified. If Inequality (2.34) is strict, i.e.

$$\text{Cost}_x^{j_0-k-1}(u) < {}^k\text{Cost}_x^{j_0-k-1}(u),$$

then the minimizer of Equation (2.26), which is the $(k+1)$-reduction in T_x^u, makes use of a $(k+1)$-reduction or a $(k+2)$-reduction. Otherwise, ^{k}Cost and Cost values are identical for the minimizer, since it uses reductions of degree at most k in trees for which such values are assumed to be identical or (second part) already proven to be identical (first part). Then the minimizer of Equation (2.26) is also the minimizer of (2.31) and (2.34) is not strict.

If the minimizer uses a $(k+2)$-reduction, then Statement (2.35) is obviously true and v is equal to y or z or additionally for general trees of type T_x^u may be equal to x.

If the minimizer uses a $(k+1)$-reduction in T_y^u, T_z^u or in general trees in T_x, then the values of the ^{k}Cost and Cost functions have to be different for this $(k+1)$-reduction. Otherwise, the minimizer of (2.26) is also the minimizer of (2.31). Consequently, the $(k+1)$-reduction in T_x^u uses a $(k+1)$-reduction in T_y^u, T_z^u or T_x (only second part of the proof; for $u \neq x$) which itself uses a $(k+2)$-reduction in some subtree. This follows from the assumption that Statement (2.35) is true for the children of x. So, the $(k+1)$-reduction in T_x^u uses a $(k+2)$-reduction in some subtree.

It remains to the **third step** of the proof to show that ^{k}Cost and Cost values of k-reductions are identical.

This is obviously true as long as Recursion Formula (2.31) makes no use of $(k+1)$-reductions which happens only if $\Delta_0 = 0$ and the zero cost number of T_x^u is different from the zero cost number of T_x. So, this happens rarely.

If, however, a $(k+1)$-reduction appears in Formula (2.31), it may only threaten the identity of ^{k}Cost and Cost function for the k-reduction, if its ^{k}Cost and Cost values are different. But, the difference of ^{k}Cost and Cost values of a $(k+1)$-reduction is an indicator, that the $(k+1)$-reduction in question uses a $(k+2)$-reduction in some subtree. Proposition 12 guarantees that a k-reduction makes no use of a $(k+2)$-reduction in any subtree. So consequently, it does not make use of a $(k+1)$-reduction which itself uses a $(k+2)$-reduction.

So, in this case the minimizer of (2.26) does not use the problematic $(k+1)$-reduction. It uses only reductions whose ^{k}Cost value coincide with the Cost value. Consequently, the minimizers of Equation (2.26) and (2.31) are identical. So are ^{k}Cost value and Cost value of the k-reduction in T_x^u.
□

2.5.16 Algorithm for solving KMP by descending iteration

In analogy to Algorithm 3 in Section 2.3.9 an algorithm can be set up to solve KMP by descending iteration. The main task is again to determine the triples (x, u, j) of vertex, its ancestor and the degree of reduction for which j-reductions in T_x^u have to be computed and their order of calculation.

Instead of the Cost function the ^{k}Cost function is evaluated which implies that ^{k}Cost values for reductions up to a degree of at most $(k+1)$ have to be computed. An inspection of the ^{k}Cost Formula (2.30), (2.31) and (2.32) shows that basically for any tree T_x^u with a zero cost

number greater than or equal to $(k+1)$ all these ${}^k\text{Cost}$ values have to be computed. This is due to the fact that the overall goal is to determine the k-reduction in T_r, i.e. to calculate

$$\text{Cost}_r^{\text{ZCN}(T_r)-k}(r) = {}^k\text{Cost}_r^{\text{ZCN}(T_r)-k}(r).$$

For trees T_x^u with a zero cost number less than $(k+1)$ all possible reductions have to be computed — from the 0-median solution up to the zero cost median solution.

Initially — before starting the reduction process — the zero cost solution for the entire tree T_r is determined by Algorithm 6. This algorithm delivers not only the desired zero cost solution in T_r but also zero cost numbers and zero cost ancestor for any vertex of the tree which are necessary information during the reduction process.

Algorithm 6 itself requires an ordering of the vertices according to a post-order traversal of T_r. So, first of all Algorithm 4 has to be applied.

The actual reduction process consists out of four loops (compare Algorithm 3).

Algorithm 7 (KMP by descending iteration)

Preprocessing	Binary tree Algorithm 4	The input tree has to be transformed into a binary tree. The vertices of the tree T_r are enumerated according to a post-order traversal.
Initializing	Algorithm 6	Determine the zero cost solution of T_r.
Main	**X-Loop**	x iterates through all vertices of T_r according to the post-order traversal.
	U-Loop	u runs through all ancestors of x starting at x and stopping at r. So, u traverses the directed path which connects x to r.
	J-Loop	j is chosen between 1 and the minimum of $k+1$ and $\text{ZCN}(T_x^u)$
	Minimum Loop	According to Recursion Formula (2.32) at most $j+1$ pairs of p and q-reductions with $p+q = j+\Delta_0$ have to be examined to determine the optimal combination.
	Computation	$${}^k\text{Cost}_x^{j_0-j}(u)$$ is calculated according to Definition 19.

Regarding the sequence of implementation of the four loops, the same comments as following the description of Algorithm 3 are in place.

Due to the special nature of discrete and semi-discrete distance functions (see Section 2.5.2 and Definition 14) for which Algorithm 7 is especially efficient, X-Loop and U-Loop offer some more potential to save run-time.

Trees T_x^u with zero cost number 0 don't have to be visited during X-Loop and U-Loop, since the zero cost solution coincides with the 0-median solution: costs are zero and reductions are not possible. There is nothing which has to be computed.

Additionally, the computation of the ${}^k\text{Cost}$ function can be omitted for trees T_x^u, if

$$\text{ZCN}(T_x^u) = \text{ZCN}(T_{\text{par(x)}}^u) = 1,$$

i.e. if the tree in question and the tree with the parent of x in place of x have both got zero cost number 1.

According to Equation (2.32) trees T_x^u with $x \neq u$ are needed only to compute the ${}^k\text{Cost}$ value for trees of of type $T_{\text{par(x)}}^u$ where par(x) is the parent of vertex x. If $\text{ZCN}(T_{\text{par(x)}}^u) = 1$, then the 1-reduction is the only possible reduction in this tree which leads to the 0-median solution of the tree and is computed directly by Equation (2.30). A reference to a reduction of tree T_x^u is not necessary anymore.

Trees of type T_x are eventually necessary to compute ${}^k\text{Cost}$ values for trees T_x^u according to (2.31) where u is an ancestors of x. Again, if the zero cost number of T_x^u is equal to 1, then there is no need to compute the zero nor — if applicable — the one median solution in T_x, since the 1-reduction in T_x^u is computed directly by (2.30). Consequently, if for all ancestors u of x the zero cost number is less or equal to 1, then no reduction of the tree T_x has to be computed.

Chapter 3

Empirical analysis

The following chapter presents a strategic analysis of a planning scenario as it could have been conducted in the preparation of an actual planning process. The main goal is to demonstrate how the developed planning models can be used to produce information which can be part of the specification of concrete planning rules. The scenario is fictitious and does not correspond to the actual Telekom Austria planning strategy in any detail. This is especially due to the fact, that the presented strategy reflects the FTTC approach, whereas realistic strategies consider a mix of network technologies (see Chapter 1 Section 1.1.7). Furthermore, the analyzed data allows a comparison of the two algorithms to solve the k-median problem which were presented in detail in Chapter 2. Therefore, both algorithms were implemented.

3.1 Implementation

The algorithms were implemented in MATLAB R2008b. Their iterative versions were realized (see Sections 2.3.9 and 2.5.16). A recursive implementation was also attempted, but later discarded, because of memory problems with large instances. The same pre- and postprocessing procedures were applied to both algorithms. Preprocessing comprised of

- the optional treatment of the CO circle, i.e. in case Rule 12 CO circle (Section 1.4.12) is enforced, all customers of the CO circle are removed from the data set.
- the extraction of a spanning tree from the graph describing the copper network,
- the enumeration of the vertices of the graph according to a post-order traversal, and
- the determination of the zero cost ancestor function by Algorithm 6.

The latter — the zero cost ancestor function — is actually not necessary for the ascending version of the k-median algorithm. However, the actual implementation of the ascending algorithm makes use of the zero cost number which is a certain advantage for this version. The knowledge of the zero cost number of vertices allows the algorithm to skip certain vertices during the X - and U - Loop (see end of Section 2.5.16).

Postprocessing comprised of

- the determination of the assignment of customers to facilities for all customers and all solutions[1]
- restoring the original coding of the vertices,
- reinsertion of customers of the CO circle in case they had been removed.

3.2 Planning scenario and sample

The following parameter values were chosen for the basic planning scenario

- R3 Customer demands: number of postal addresses
- R8 ARU capacities: no limitation
- R10 Distance rule: transmission rate. This made the determination of several technical parameters necessary. They were chosen according to standard values.
- R10 Distance rule threshold value: 20 Mbps
- R11 Individual distance rule: the same for all customers
- R12 Co cirlce: one scenario with and all others without the enforcement of the CO circle
- R13 TNAK length: 100m, the same value for all customers .

For the database a sample of 106 out of around 1,400 local loops were chosen. The selection was not random and is not representative for the entire set of local loops. The idea of the selection was to produce a diverse sample of local loops according to a classification of the settlement structure of the local loops in Austria. The classification was developed by R. Schnepfleitner, Z. Daroczi, R. Kalasek and W. Feilmayr in 2000 and published in [46]. It consists of five main categories and five subcategories. They are

- urban (Gr. Urban)
- suburban (Gr. Suburban)
- small town (Gr. Kleinstädtisch)
- touristy condensed (Gr. Touristisch verdichtet)
- rural (Gr. Rural).

Category rural is further differentiated into five subcategories:

- scattered settlements (Gr. Streulage)
- scattered settlements with center (Gr. Streusiedlung mit Zentrum)

[1] The algorithms deliver solutions for every integer between 0 and k.

- street village (Gr. Straßendorf)
- clustered settlement (Gr. Reihen- und Haufenort)
- polycentric (Gr. Polyzentrisch).

For details of the definitions see [46] page 12 and 13.

From the small town cluster an instance was moved to the urban cluster for all the analysis in this thesis. The reason was simple. With respect to the most relevant categories of local loops for the applied algorithms — number of vertices, demand and pathlength — it is an extreme outlier of cluster small town. Even within cluster suburban it would still be an extreme outlier. In the urban cluster it resembles the largest instances with respect to the three categories.

The following paragraphs and Tables 3.1 (number of vertices), 3.2 (pathlength), 3.3 (demand) and 3.4 (number of facilities) give an overview of the sample with respect to the four key features.

Clearly, there is a relationship between the cluster of settlement and the size and the structure of the settlement.

The urban cluster is the largest (many vertices and high demand), but not necessarily the most complex. In some of the suburban access areas the algorithms are confronted with greater pathlengths. By moving further down the list of clusters, the size and the complexity of instances drop making the rural cluster the smallest and most simple one. Only the polycentric cluster goes beyond this pattern. In size and complexity it is more like the touristy condensed cluster.

As appropriate as this summary is, it can be seen by an inspection of the three tables that **each of the 9 different clusters still constitutes a heterogenous set of instances**.

The smallest urban access area is smaller than at least 50% of the suburban instances. The medians of the number of vertices of suburban and small town clusters are nearly identical but the maximum of the latter is just half the size of the former.

With respect to complexity the urban cluster does not vary as much as the suburban. But, these two clusters are certainly the biggest and most complex.

The average access area of all other clusters does not achieve a comparable long pathlength, although some of the instances still reach into the top 50% of urban areas (Table 3.2).

The inspection of the demand table (Table 3.3) clearly shows that the urban cluster contains the most densely populated areas. Nearly, 60% of the total demand covered by the 106 access areas is concentrated there. The suburban cluster may be high in complexity, but, the suburban access area with the highest demand has got a lower demand than nearly 50% of the urban areas. The median of the demand in rural areas exceeds 1,000 which is due to the polycentric subcluster.

Cluster of settlement	n	# of vertices min	median	max
Total	106	129	1,166	16,704
Urban	16	1,512	7,001	16,704
Suburban	15	470	1,883	8,247
Small town	11	273	1,829	4,356
Touristy condensed	12	524	1,217	2,305
Rural	52	129	586	2,356
Scattered settlements	10	129	520	1,686
Scattered settlements with center	7	251	444	1,201
Street village	12	199	468	838
Clustered settlement	10	177	409	1,020
Polycentric	13	253	1,225	2,356

Table 3.1: Size of instances

Cluster of settlement	n	Pathlength min	median	max
Total	106	1,686	22,198	993,763
Urban	16	38,532	320,712	792,766
Suburban	15	5,738	52,015	993,763
Small town	11	2,624	34,393	92,507
Touristy condensed	12	9,055	21,279	47,716
Rural	52	1,686	9,838	60,320
Scattered settlements	10	1,686	10,174	54,827
Scattered settlements with center	7	4,481	5,993	30,857
Street village	12	2,568	6,213	22,401
Clustered settlement	10	3,409	7,312	17,998
Polycentric	13	3,228	22,302	60,320

Table 3.2: Structure of instances

The maximum number of facilities which is necessary to grant one hundred percent coverage in the base scenario is listed in Table 3.4. Although, there is one more urban than suburban cluster contained in the sample the total sum of facilities in suburban clusters exceeds the sum in urban ones. This shows again the higher complexity of suburban areas in comparison to the urban cluster.

However, at least in terms of the maximum number of facilities the suburban cluster is not the most demanding one. Urban and suburban clusters cover nearly 75% of the total demand of the sample. But, they only require one third of all facilities to cover that demand[2]. Most demanding in terms of the average need for facilities measured by the median is the small town cluster with 69 facilities, followed by the polycentric cluster with 65 facilities. Next in sequence is the suburban cluster with a median of 44 facilities. The inspection of the table also shows

[2]This result has to be put into perspective, since overutilization is not accounted for in this analysis.

Cluster of settlement	n	Demand sum	min	median	max
Total	106	518,593	110	1,801	73,316
Urban	16	301,214	1,266	13,729	73,316
Suburban	15	81,786	1,221	4,000	14,499
Small town	11	44,099	933	4,413	8,190
Touristy condensed	12	28,960	493	2,059	5,725
Rural	52	62,534	110	1,107	3,056
Scattered settlements	10	8,590	110	833	1,778
Scattered settlements with center	7	6,721	220	851	1,581
Street village	12	12,986	370	935	2,532
Clustered settlement	10	8,924	306	748	2,311
Polycentric	13	25,313	503	1,877	3,056

Table 3.3: Distribution of demands

Cluster of settlement	n	# of facilities sum	min	median	max
Total	106	4.986	5	38	195
Urban	16	743	11	39	131
Suburban	15	912	14	44	195
Small town	11	762	5	69	155
Touristy condensed	12	611	15	42	113
Rural	52	1.958	5	28	125
Scattered settlements	10	357	5	34	92
Scattered settlements with center	7	292	18	22	88
Street village	12	224	7	18	37
Clustered settlement	10	250	8	24	53
Polycentric	13	835	14	65	125

Table 3.4: Number of facilities for one hundred percent coverage

that the access areas are still very heterogeneous with respect to this key number even within clusters. Minima and maxima are far apart.

In summary, the clusters of settlement clearly differentiate with respect to size of instance, i.e. number of vertices, complexity of instance, i.e. pathlength, demand, and maximum number of facilities. But, the clusters overlap.

The two urban clusters promise to be the most demanding ones for the algorithms. A clear line can be drawn between these two clusters and the others.

3.3 Runtime analysis of the descending k-median algorithm

To study the runtime behavior four different scenarios were computed: based on the basic planning scenario from Section 3.2 a 1-median and a ZCN-median[3] solution with the ascending algorithm and a 1-reduction and a ZCN-reduction with the descending version. The scenarios will be addressed by ASC_1, ASC_{All}, $DESC_1$ and $DESC_{All}$ respectively.

The following analysis of the principle runtime refers to scenario $DESC_{All}$. The time for loading and saving data is not included in any runtime analysis.

The descending k-median algorithm runs rather fast on the 106 instances.

The total runtime of the algorithm is less than half a minute for three quarters of the data sets. The worst runtime lies below half an hour (26.2 min resulting from an urban access area which is the graph with the most vertices — 16,704 — and the second longest pathlength — 792,766).

As it is to be expected from the diversity of the sample, the variation of runtime is large. Mean runtime varies strongly between clusters. But still, the deviation of runtime within clusters is large, too.

Compared to the overall average runtime of 91 seconds the standard-deviation of 246 seconds is very high (refer to Table 3.5).

Runtime of scenario $DESC_{All}$

Cluster of settlement	Runtime in seconds				
	mean	std	min	median	max
Total	91	246	0,3	8	1,573
Urban	422	448	13,1	319	1,573
Suburban	139	278	1,2	21	893
Small town	23	26	0,4	17	99
Touristy condensed	13	11	1,8	8	35
Rural	7	10	0,3	3	47
Scattered settlements	7	10	0,3	5	34
Scattered settlements with center	9	12	1,	2	34
Street village	2	2	0,3	1	7
Clustered settlement	3	2	0,5	2	8
Polycentric	14	13	0,6	11	47

Table 3.5

The urban cluster takes the longest. The median of 5 minutes is clearly higher compared to the second highest median of 21 seconds in suburban areas, followed by 17 seconds in small towns and 11 seconds in polycentric rural areas. Except for this subcluster the median runtime for rural clusters does not exceed 5 seconds.

[3] ZCN is the zero cost number of the respective tree.

As can be seen from Table 3.5 the variation within clusters stays high. For the urban cluster runtime varies between a few seconds, over several minutes up to nearly half an hour. Standard-deviation rarely falls below the corresponding mean value. In case of suburban areas it is more than 200% of the average.

The analysis of the sample in Section 3.2 already showed how heterogenous the individual clusters with respect to number of vertices, pathlength and demand are. From the considerations of Chapter 2, Section 2.3.9 and especially Section 2.5.16, it is clear that pathlength is one of the main factors which influences runtime. The pathlength is equal to the number of pairs of vertices (x, u) where x is any vertex of the access network tree and u is any of x's ancestors. So, it basically corresponds to the number of combined iterations of X- and U-Loop according to Algorithm 7. Therefore, some space is given to the inspection of the relation between runtime and pathlength.

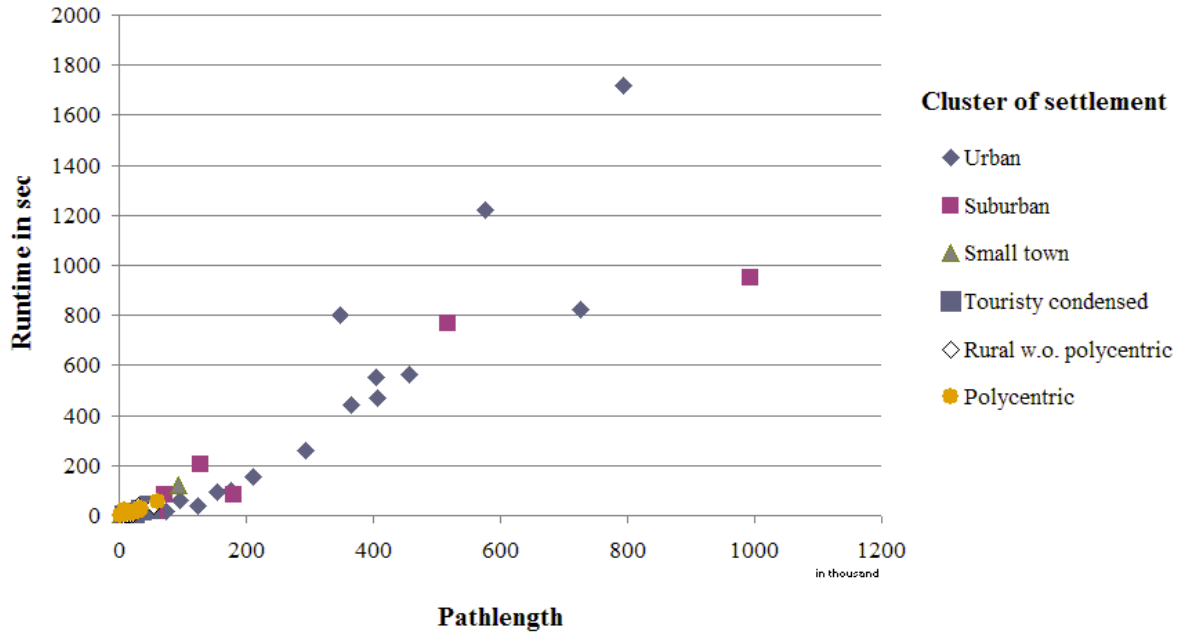


Figure 3.1: Runtime of scenario DESC_{All} by pathlength and cluster

The urban and suburban segments dominate the picture in Figure 3.1. A strong correlation between pathlength and runtime becomes visible, at least for the urban segments. As a consequence, and because of the insights from the analysis of the sample in Section 3.2 urban and non-urban clusters are analyzed separately.

Figure 3.2 shows the relation between runtime and pathlength for urban clusters. After a logarithmic transformation of both of the involved variables a linear regression is applied, i.e. a multiplicative model is determined. The resulting formulas together with the coefficients of determination are also depicted in the figure.

For both clusters a strong correlation is clearly visible which is supported by the corresponding coefficients of determination. Looking at the formulas it seems that the influence of the

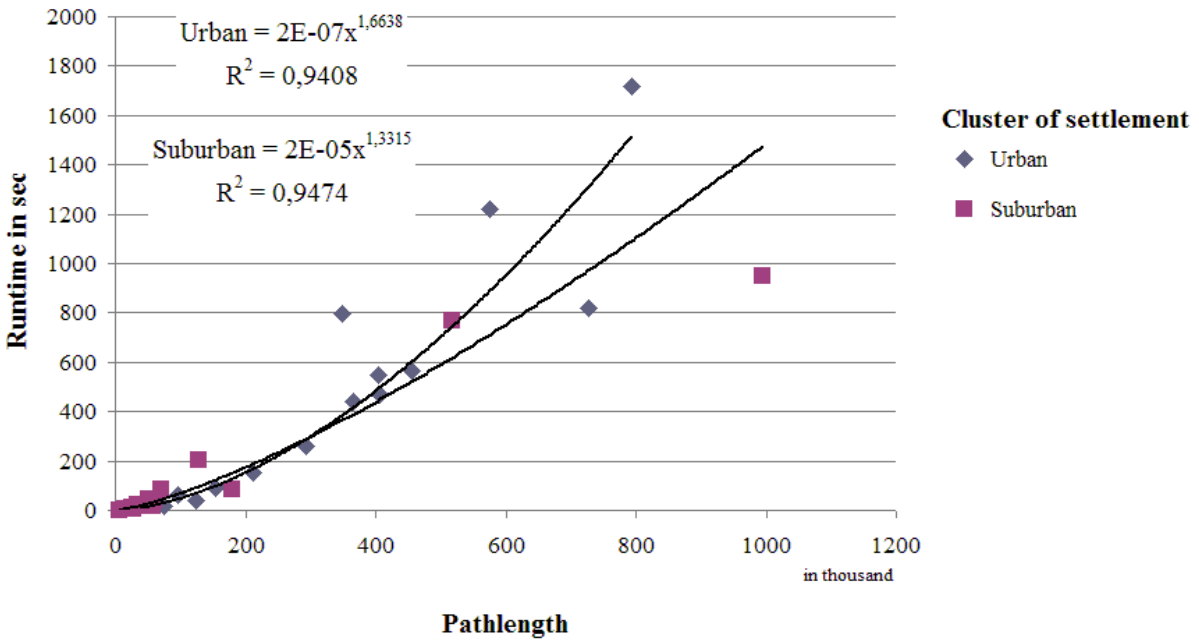


Figure 3.2: Runtime of scenario $DESC_{All}$ for urban clusters

pathlength on runtime is not as strong for suburban areas as for urban ones. This observation, however, is based on a single instance from the suburban cluster which lies on the extreme right too low for the pattern in the urban scatter cloud. The graphic also shows that the suburban cluster contains two instances with extremely high pathlengths which are quite untypical for that cluster.

The analogue analysis for non-urban clusters (Figure 3.3) presents a much more homogenous situation. The fitted curves are very similar. The scatter plot is more compact. Although, the coefficients of determination for rural and especially polycentric settlements are not as high as for the other types. This may be due to a few outliers.

Therefore, in Figure 3.4 a single curve is fitted to the cloud of rural access loops leading to a satisfying result.

The runtime analysis is finalized by multiple and multiplicative regression models based on pathlength, number of vertices, demand and the zero cost number ZCN of the access area:

The influence of pathlength on runtime remains strong. The zero cost number improves the estimation significantly. Some additional improvement is obtained for urban and suburban settlements by including the size of demand in the regression analysis.

The pathlength stays the most significant and influential factor in all models. Not surprisingly, the impact of the number of vertices turns out to be very low and not significant, since the number of vertices is already well represented by the pathlength. The influence of the demand depends on the type of settlement. For non-urban clusters the exponent of demand is even

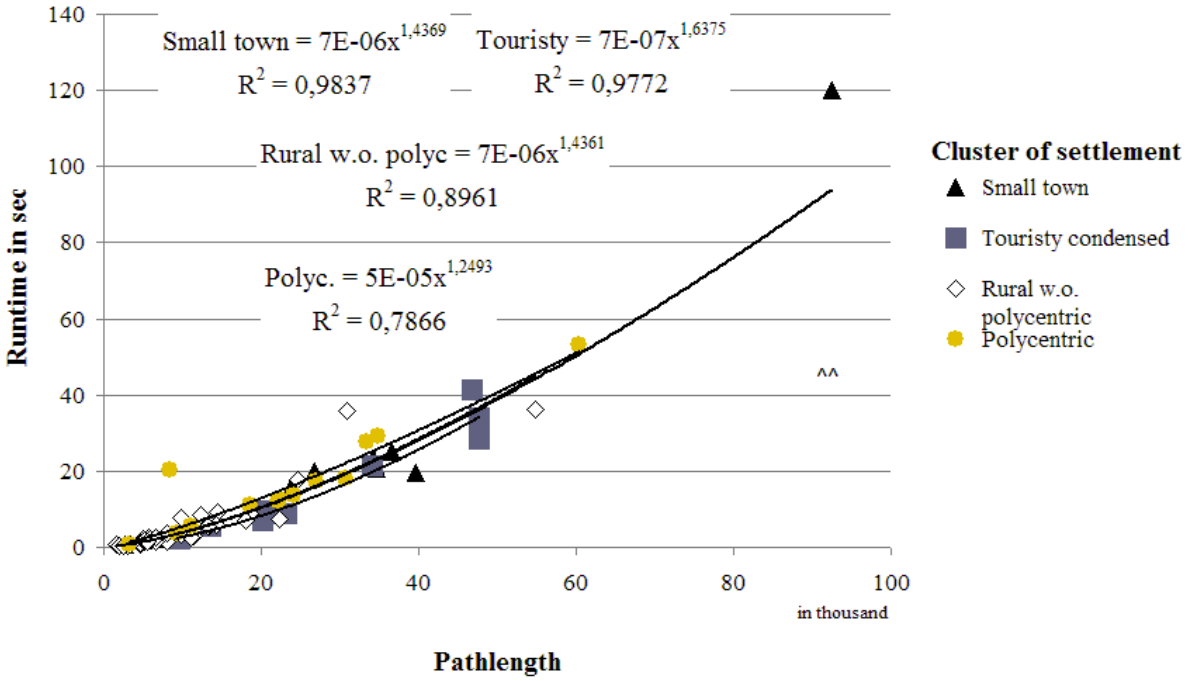


Figure 3.3: Runtime of scenario DESC_{All} for non-urban clusters

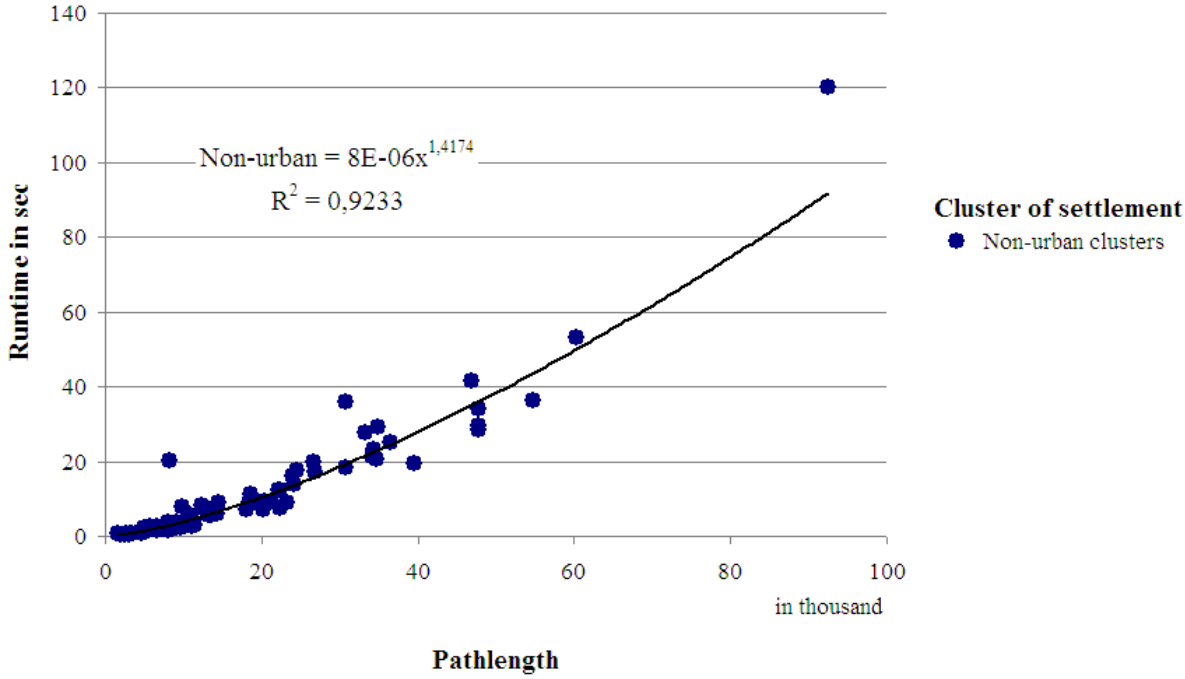


Figure 3.4: Runtime of scenario DESC_{All} for non-urban clusters II

negative, i.e. for areas with higher demand the estimation of runtime is reduced. Since runtime reduction for higher demands does not make sense and the model does not really improve the results, it is dismissed. However, a significant and positive effect of the size of demand on

runtime can be observed for urban and suburban settlements. The zero cost number adds some additional information about the structure of the problem and is able to improve the estimates significantly. This is hardly surprising, since the zero cost number determines the size of the J - Loop in Algorithm 7. The equations of the final model are

$$\widehat{\text{runtime}} = 2.21\text{E-}04 * \text{pathlength}^{0.774} * \text{ZCN}^{0.803}$$

with $R^2 = 0.968$ for non-urban clusters and

$$\widehat{\text{runtime}} = 1.92\text{E-}06 * \text{pathlength}^{1.142} * \text{ZCN}^{0.420} * \text{demand}^{0.301}$$

with $R^2 = 0.991$ for urban clusters.

To summarize, for the application to strategic questions the presented algorithm is best suited in terms of runtime. Strategic applications are concerned with principle questions of the planning task like, how should certain strategic parameters be chosen (compare Chapter 2 Section 2.2.5)? Which local loops should be selected for next year's planning program? Examples for such strategic applications are presented in Sections 3.5 *Service quality and facility utilization* and 3.6 *Effect of enforcement of the CO circle*. In general these tasks make it necessary to study several scenarios for a larger set of local loops. Despite the phrase that time is money, the time frame to solve such tasks can be expressed at least in days. Considering now, that the production of the data for the four planning scenarios DESC_1, DESC_{All}, ASC_1 and ASC_{All} for the 106 local loops of the sample did not take more than 23 hours[4], it can be concluded that a complete analysis including a report is possible within two to three days.

Operational applications, like the structural planning process described in Chapter 1 Section 1.5.3, set different requirements. Planning on paper is instantaneous and continuous. The process itself does not entail breaks or interruptions for the planner. He decides about breaks and interruptions. For a planning process which is supported by a machine to imitate such a behavior very low runtimes of algorithms are required. This is achieved by the k-median algorithms for small and maybe medium sized local loops, i.e. in most cases. For over 80% of the local loops of the sample the runtime of the DESC_{All} scenario took less than one minute. Only, for the large local loops, i.e. most of the urban settlements, 25% of the suburban and one of the small town instances, runtime exceeds one minute. In these cases the planning process has to be adopted in such a way that it accounts for the breaks resulting from increased runtime. An appropriate estimation may enhance the workflow.

3.4 Comparison of ascending and descending k-median algorithm

The following section is concerned with a comparison of the runtime of the two versions of the k-median algorithms which are presented in Chapter 2. The analysis is based on the four scenarios DESC_1, DESC_{All}, ASC_1 and ASC_{All} which are defined in the beginning of Section 3.3. Pre- and postprocessing as well as time for loading and saving data is not accounted for.

[4]This runtime estimate also includes time for loading and saving data.

From a practical point of view both versions are equally fast. On average the ascending k-median algorithm is slightly faster than the descending version.

To compare the two implementations the algorithms are applied to similar tasks. What's the 1-median problem for the ascending algorithm is the 1-reduction for the descending version. In Table 3.6 the runtime differences for the two scenarios ASC_1 and $DESC_1$ are depicted. In 17 out of the 106 instances the descending algorithm is faster. But, the total runtime advantage[5] for the ascending algorithm is 111 seconds, i.e. less than 2 minutes. The best case advantage for the ascending algorithm lies around half a minute. Clearly, such advantages are only achieved for the big examples from the urban and suburban clusters. For all other clusters the differences between the runtime of the two algorithms can be measured in seconds.

Runtime difference between scenarios ASC_1 and $DESC_1$

Cluster of settlement	n	$DESC_1$ faster	Runtime in seconds sum	min	max
Total	106	17	-111	-31,9	2,8
Urban	16	6	-40	-15,9	2,8
Suburban	15	4	-50	-31,9	0,9
Small town	11	2	-6	-2,3	0,5
Touristy condensed	12	1	-4	-2,0	0,6
Rural	52	4	-12	-1,3	0,1
Scattered settlements	10	1	-3	-1,2	0,0
Scattered settlements with center	7	1	-2	-1,0	0,0
Street village	12		-1	-0,4	0,0
Clustered settlement	10	1	-1	-0,4	0,0
Polycentric	13	1	-6	-1,3	0,1

Table 3.6: Runtime difference RT(ASC_1) – RT($DESC_1$)

The second comparison is based on scenarios ASC_{All} and $DESC_{All}$ which correspond to the tasks to situate as many facilities into the local loop as there are necessary to provide all customers with at least 20 Mbps (scenario ASC_{All}) and then to remove all these facilities again (scenario $DESC_{All}$). Table 3.7 shows the details of the analysis.

Still, the ascending version is faster. However, the number of cases where the descending algorithm outperforms increases to 48. The total runtime advantage of the ascending implementation decreases to one minute. For urban and suburban the best case advantage remains relatively high. For all other clusters the runtime difference can now be measured in tenths of a second.

Finally, the issue which is first addressed in Section 2.2.6 and which is the basic motivation for the development of the descending algorithm to solve the k-median problem is studied: Is there an advantage to solve the k-median problem in directed trees for relatively large values of k by approaching the value from above, by removing facilities one by one from the zero

[5]Total runtime advantage corresponds to the sum of all runtime differences.

Runtime difference between scenarios ASC_{All} and $DESC_{All}$

Cluster of settlement	n	$DESC_{All}$ faster	Runtime in seconds sum	min	max
Total	106	48	-60	-37,7	2,5
Urban	16	9	-20	-12,4	1,3
Suburban	15	4	-41	-37,7	2,5
Small town	11	5	0	-1,5	1,8
Touristy condensed	12	4	-1	-0,6	0,3
Rural	52	26	3	-0,5	1,0
Scattered settlements	10	5	1	-0,1	0,4
Scattered settlements with center	7	3	0	-0,1	0,4
Street village	12	3	-1	-0,3	0,1
Clustered settlement	10	6	0	-0,1	0,0
Polycentric	13	9	2	-0,5	1,0

Table 3.7: Runtime difference RT(ASC_{All}) – RT($DESC_{All}$)

cost solution, until the desired value is reached? The question is explored by comparing the ZCN-median solution for the ascending algorithm (scenario ASC_{All}) with the 1-reduction for the descending version (scenario $DESC_1$).

The descending k-median algorithm is faster than its ascending counterpart for values of k which are close to the zero cost number. However, the data contains surprises.

Table 3.8 summarizes the results. In 103 of 106 cases it is faster to compute the 1-reduction than the ZCN-median in the ascending fashion. There are three cases where the ascending algorithm is faster, even if the difference is less than a second.

This is not surprising in the case of the two rural local loops (one scattered settlement and one street village). Runtime is very low for small local loops anyway. Zero cost numbers, i.e. the maximal number of facilities necessary for full supply, is generally low for these instances. The maximal advantage which can be observed does not exceed half a minute.

For urban and suburban areas the total advantage of 10 and 7.6 minutes respectively is significant. However, recalling the work flow considerations at the end of Section 3.3 such improvements can rarely be utilized in a practical framework.

The urban cluster contains a surprise. There is one relatively large instance (the second largest number of vertices: 14,632; the forth largest pathlength: 575,676; the second largest demand: 47,593) where the ascending algorithm reaches the ZCN-median solution before the descending algorithm removes one facility. However, the zero cost number of 39 (rank 52) is incredibly low for a local loop of this size. Furthermore, in this case the runtime of all four scenarios differs at most by 2 seconds, i.e. the descending reduction of all facilities is nearly as fast as the placement of a single facility.

Runtime difference between scenarios ASC_{All} and $DESC_1$

Cluster of settlement	n	$DESC_1$ faster	Runtime in seconds sum	min	max
Total	106	103	1307	-0,8	254,8
Urban	16	15	610	-0,8	254,8
Suburban	15	15	458	0,3	222,8
Small town	11	11	60	0,0	24,1
Touristy condensed	12	12	34	0,2	11,3
Rural	52	50	28	0,0	23,2
Scattered settlements	10	9	36	0,0	13,9
Scattered settlements with center	7	7	3	0,1	23,2
Street village	12	11	10	0,0	1,2
Clustered settlement	10	10	67	0,1	3,1
Polycentric	13	13	144	0,1	14,5

Table 3.8: Runtime difference RT(ASC_{All}) – RT($DESC_1$1)

In relative numbers the advantage of the descending algorithm over the ascending is expressed as the portion of time saved in the total runtime of the ascending algorithm. In these terms the advantage looks a bit more impressive. On average 20% of runtime can be saved. The proportion strongly varies between clusters. Interestingly, it is now the small clusters which show the highest potential to save time. In rural clusters the gain is nearly 30% of the total runtime of ASC_{All}. On the other hand, the average for the urban cluster is 10%, the median even lies slightly below 5%. With 17.4% the suburban cluster is close to the overall average.

3.5 Service quality and facility utilization

The fundamental motivation for the work presented in Chapter 2 was the basic realization described in Section 2.1.6 that in order to produce efficient solutions, planners ensure that only facilities of a certain degree of utilization are built. For this purpose they are willing to waive some of the networks transmission quality. This, of course, is in contrast to the strategies of the marketing department. Different quality levels make it difficult to sell products. For a certain price customers will expect a certain standard. Not all products can be offered to every customer. Difficulties in communication are the result. A compromise between these two sides can be prepared by studying the relation between network quality and different lower bounds of the utilization of facilities. Based on scenario $DESC_{All}$ an exemplary analysis of this topic is presented.

A possible marketing scenario could be that product managers plan to design two products for the market in the local loop under consideration. One product offers a bandwidth of at least 10 Mbps for a moderate price. A second product guarantees a significantly higher bandwidth of 20 Mbps for a higher price. This second transmission rate is also the target rate for the planning scenario (see beginning of Section 3.2). So, the proportions of customers (demand) who can be provided with one or both of these products are focused on. Additionally, for

the advertising department it could be of interest to put a special focus on areas with very high levels of transmission rate directly. Therefore, the proportion of customers who can be provided with a transmission rate of at least 25 Mbps is also reported. Figure 3.5 depicts the analysis for all 106 local loops of this study.

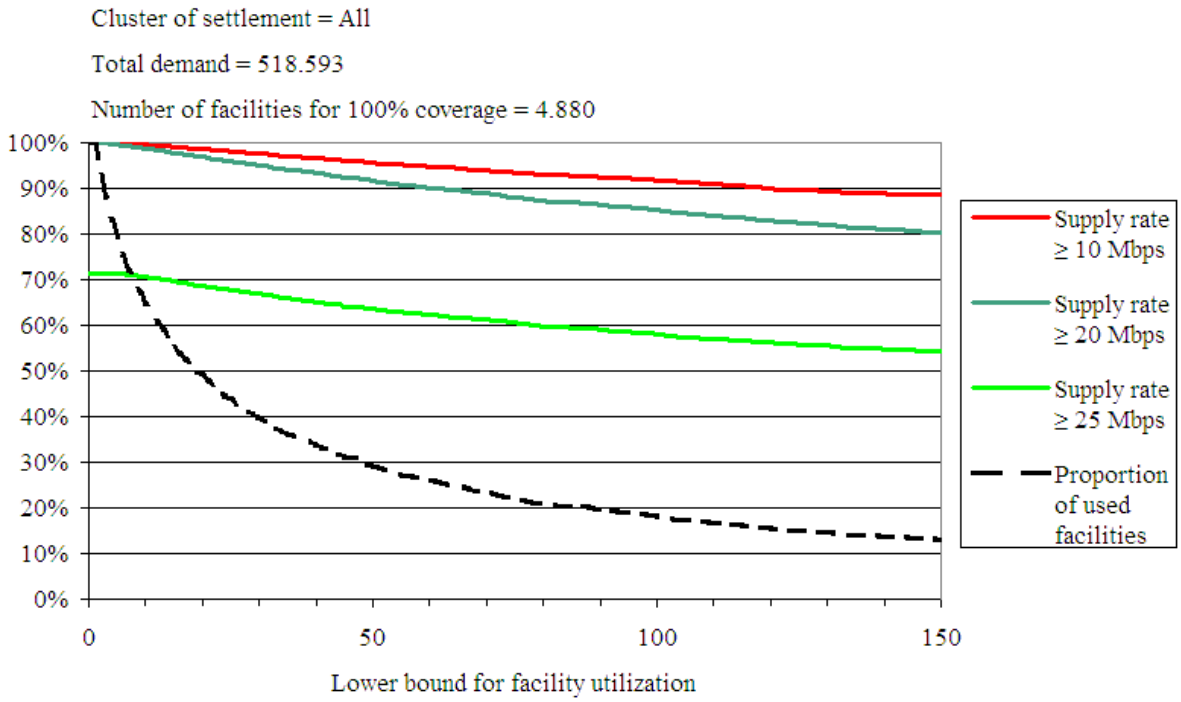


Figure 3.5: Coverage versus prevention of underutilization

The coverage with 10, 20 and 25 Mbps transmission rate is shown in relation to the lower bound which is imposed on the utilization of facilities. Clearly, as the bound increases coverage decreases. Additionally, the number of facilities which are necessary to achieve the corresponding coverage is depicted. Certainly, the number of facilities decreases with increasing lower bounds.

Over the total range of lower bounds from 0 to 150 nearly 90% of the original number of facilities can be saved, whereas at the same time 10% of the demands lose their support by both of the two products, and another 10% can be offered at least the low level product. The top coverage with 25 Mbps drops only by 15% points. The picture convincingly shows how reasonable it is to impose lower bounds on the facility utilization, since the proportion of necessary facilities reduces a lot faster than the coverage.

However, the situation depends on the type of settlement under consideration as can be shown by the following figures.

The urban cluster which covers the majority of the demand stays comparably unaffected by imposing high lower bounds on facility utilization (Figure 3.6). The availability of the low level product drops by 1.3 points and that of the high level product by 3.5 points. High quality coverage (over 25 Mbps) is reduced by at most 5.2 points. Cost cannot be reduced as strongly as in the overall picture. At most 55% of the facilities can be saved.

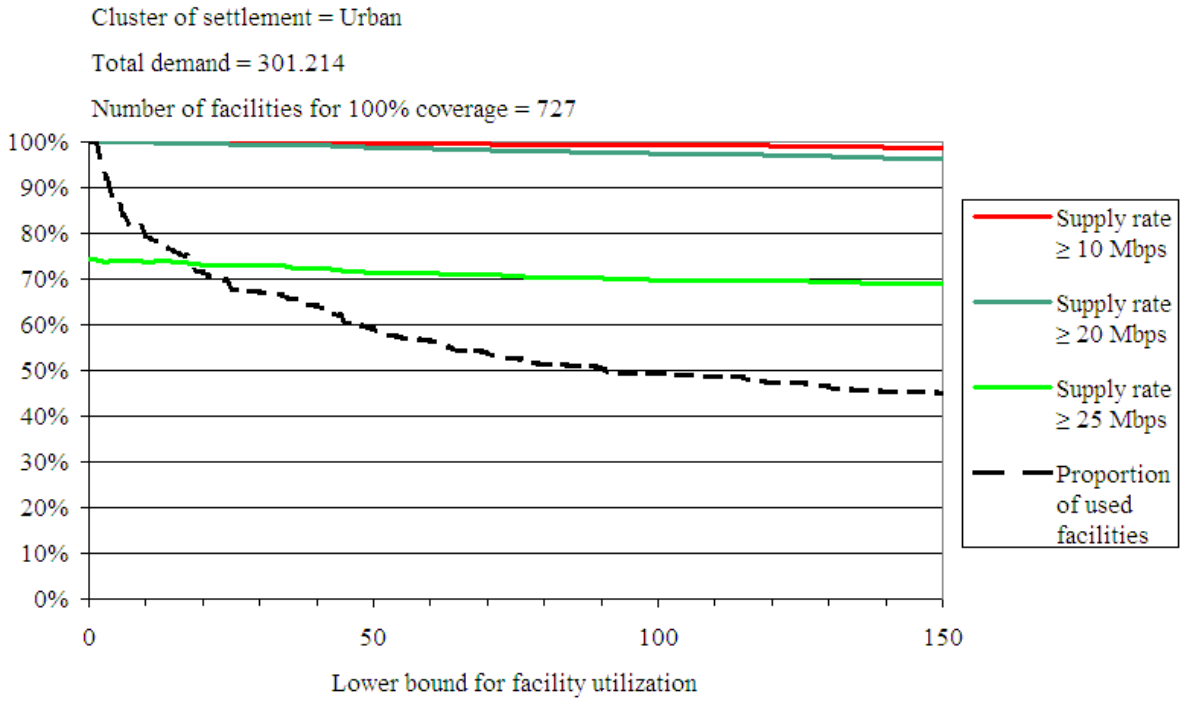


Figure 3.6: Coverage versus prevention of underutilization in urban access areas

Additionally, overutilization (see Sections 1.5.3, 2.1.9 and 4.1) which is not in the focus of the analysis of this section is an important problem for local loops where a high number of demand has to be covered. The proportion of heavily utilized facilities increases only slightly over the analyzed range of lower bounds (less than +1%). But, with over 11% of the facilities covering more than 1,000 demands and around 3% covering more than 2,000 demands the overutilization problem exceeds the average proportions of 2% and 0.3% respectively by far. Of course, the solution of overutilization accounts for additional cost.

The results for the suburban cluster (Figure 3.7) already lie below the average numbers. Top coverage (25 Mbps) decreases to 40%, coverage with the target transmission rate finally reaches 70%, and the availability of the low level product decreases to 85%. In the best case 82% of the 897 facilities which are necessary to provide full coverage can be saved. Overutilization is not a a relevant problem. Only 0.6% of the facilities have to deal with more than 1,000 demands.

An interesting detail comes to light. Initially, the increase of the lower bound for the facility utilization leads to an increase in top level bandwidth as can be seen in Figure 3.8.

Bandwidth increases although the number of facilities decreases. A contradiction? Not at all. The graph in Figure 3.9 gives an explanation why this phenomenon may occur.

There are five customers contained in this tree. Customer c_2 and c_3 have got demand 1, customer c_1 and c_4 have got demand 2 and c_5 has got demand 94. With two facilities at locations f_1 and f_2 the left picture shows the zero cost solution of the example. Because of the given demands the 1-median solution situates a facility at location f_3. Dashed lines and

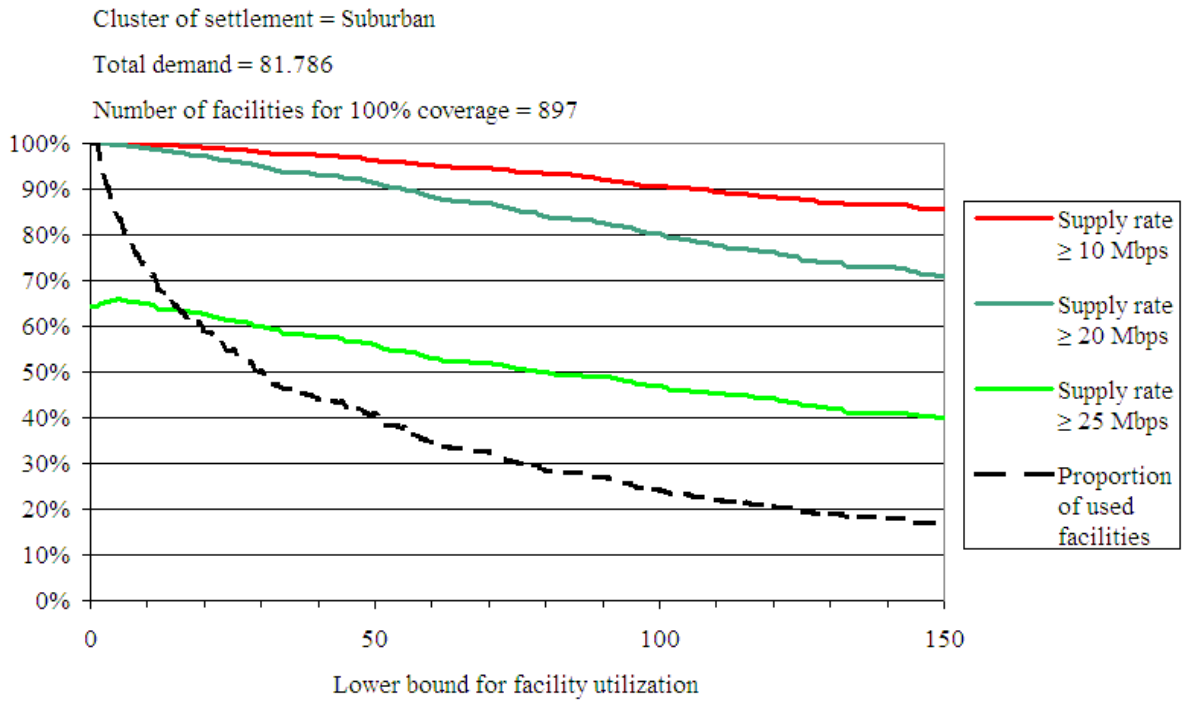


Figure 3.7: Coverage versus prevention of underutilization in suburban access areas

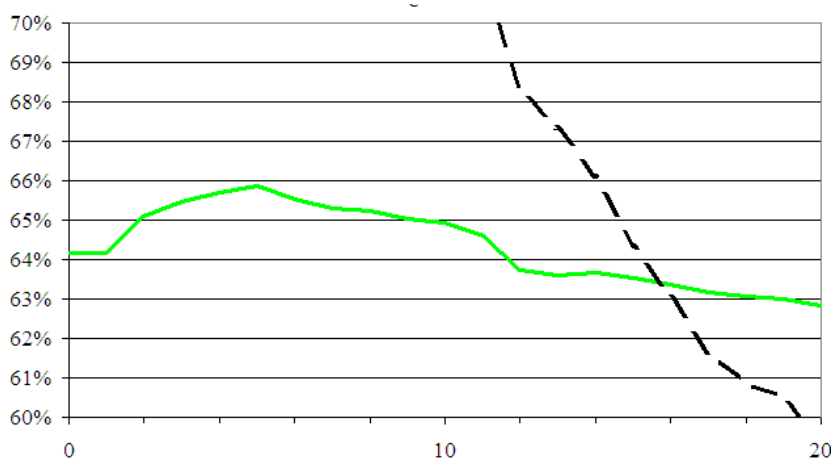


Figure 3.8: Detail from Figure 3.7

point-dashed lines indicate that the customer may be assigned to a facility at the given location. The transmission rate along dashed lines would be 25 Mbps and along pointed-dashed lines 20 Mbps. If there is no such line (e.g. from c_1 to r) than transmission rates are below 20 Mbps or even zero.

The zero cost solution (left graph) supplies 6% of the demand with 25 Mbps (all customers except c_5) and the remaining 94% of the demand at customer c_5 with 20 Mbps. The utilization of the facilities at f_1 and f_2 is weak. If a lower bound greater than three is imposed on facilities, then the 2-median solution is not permissible anymore. The 1-median solution can be used

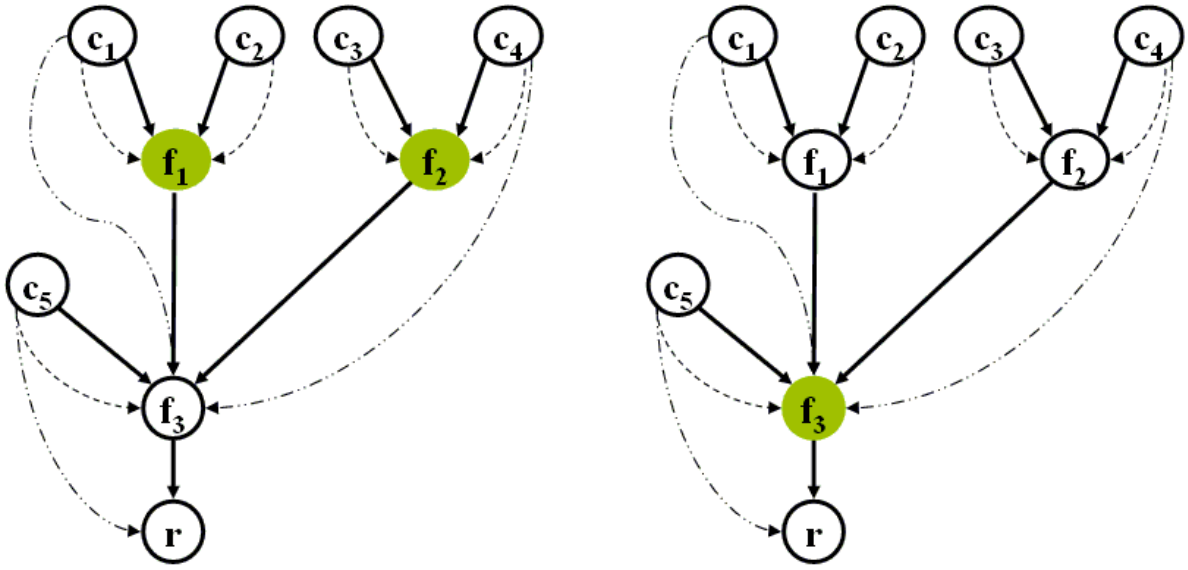


Figure 3.9 Graph with all possible k-median solutions

instead. Coverage with at least 20 Mbps drops down to 98%, since customers c_2 and c_3 are too far away from the facility at f_3. But, top coverage with 25 Mbps increases from 6% to 94%.

Finally, the union of all rural clusters is inspected (Figure 3.10). The effect of using different lower bounds for facility utilization is most problematic for access areas which are scattered and sparsely populated.

To cover the total demand of 62,000 there are nearly 2,000 facilities necessary which corresponds to 31 units of demand per facility compared to 400 per facility in urban areas. A relatively high coverage of 95% with 20 Mbps is undercut for a lower bound of utilization of as little as 15 units per facility. In urban areas the same limit is kept over the entire range of bounds which is studied. Suburban areas reach a coverage rate of 70% for a lower limit of 150. The same value is achieved for 90 units in rural areas.

In the end, with a lower bound of 150 units of demand imposed on facilities, 40% of the demand can be offered neither of the two products and 20% have to be satisfied with the low level product. Overutilization does not become a problem.

3.6 Effect of enforcement of the CO circle

The CO circle rule (R12) is described in detail in Chapter 1 Section 1.4.12. There the repeated discussions with some colleagues are mentioned who were of the opinion that it is advantageous to enforce the CO circle. This belief is already challenged by a theoretical example illustrated in Figure 1.18 which demonstrates that the number of facilities can be reduced, if the CO circle rule is not enforced. This section adds empirical arguments based on a comparison of scenario DESC_{All} which is defined at the beginning of Section 3.3 and a similar scenario which differs from DESC_{All} only by the enforcement of the CO circle.

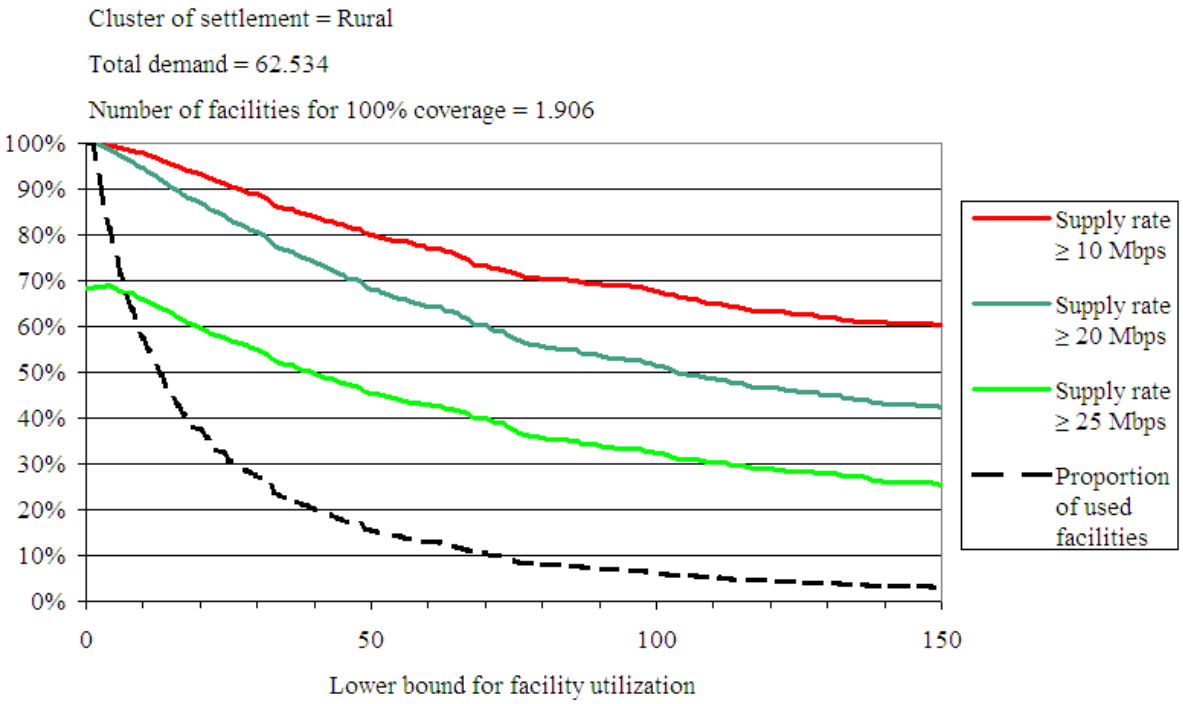


Figure 3.10: Coverage versus prevention of underutilization in rural access areas

The enforcement of the CO circle rule makes more facilities necessary and consequently, leads to more expensive solutions. Moreover, network quality decreases.

This statement is best illustrated by directly comparing the analysis from Figure 3.5 with a similar analysis where the CO circle is enforced. Figure 3.11 and 3.12 show the number of facilities and the coverage distribution depending on different lower bounds of facility utilization.

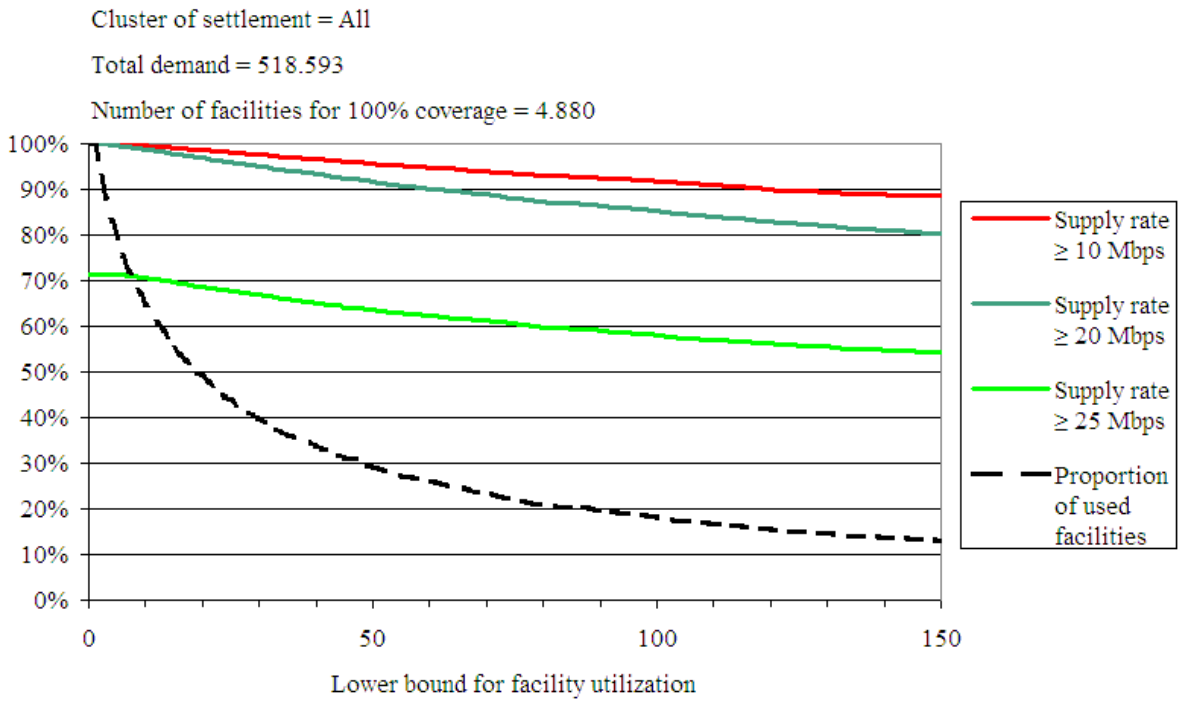


Figure 3.11: Coverage versus prevention of underutilization, CO circle is not enforced

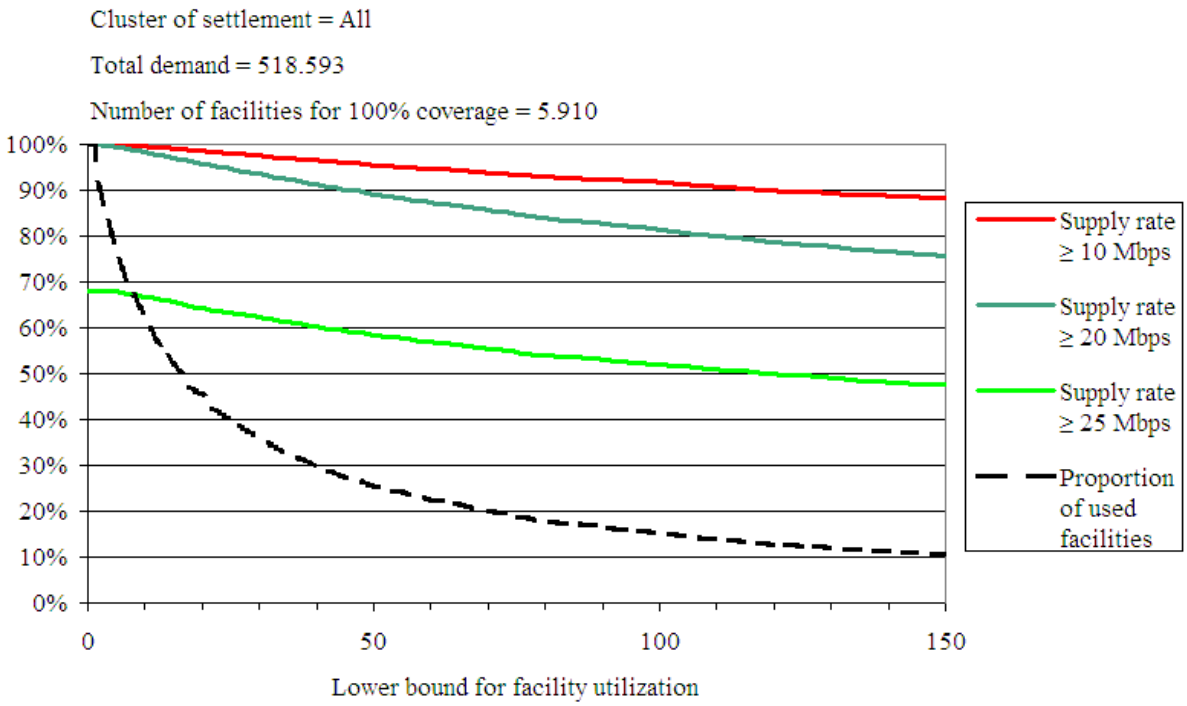


Figure 3.12: Coverage versus prevention of underutilization, CO circle is enforced

An additional 1,100 facilities are necessary to grant 100% coverage with 20 Mbps. At least for the graphs of coverage with 20 and 25 Mbps it is noticeable that the results are lower in case of enforced CO circle.

The differences become more visible for urban clusters (Figure 3.13 and 3.14). The effect on the number of facilities and hence on cost has to be put into perspective in this case. Because of overutilization there are more facilities necessary than is reported here, anyway.

However, the impact on transmission rates is clear. Whereas the target bandwidth of 20 Mbps hardly changes if the CO circle is ignored, it decreases by 10% over the range of lower bounds as soon as the circle is enforced. A similar behavior can be observed for the top transmission quality.

The pictures for all other clusters are very similar:

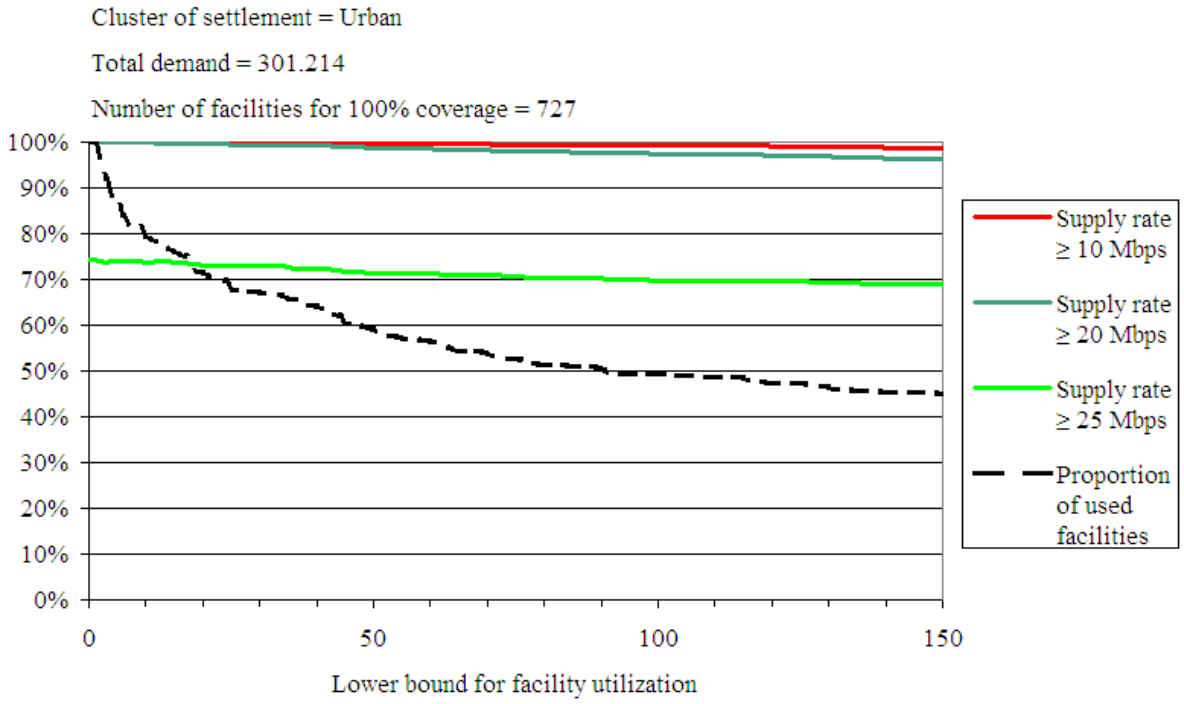


Figure 3.13: Coverage versus prevention of underutilization, CO circle is not enforced, urban cluster

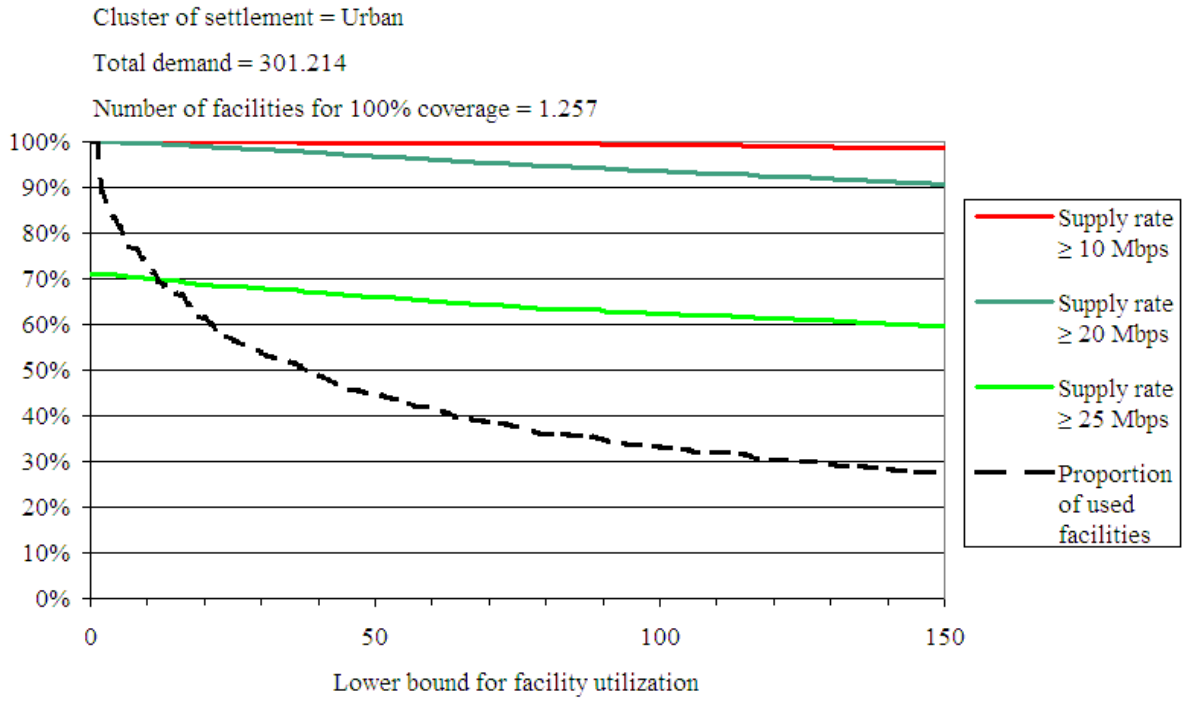


Figure 3.14: Coverage versus prevention of underutilization, CO circle is enforced, urban cluster

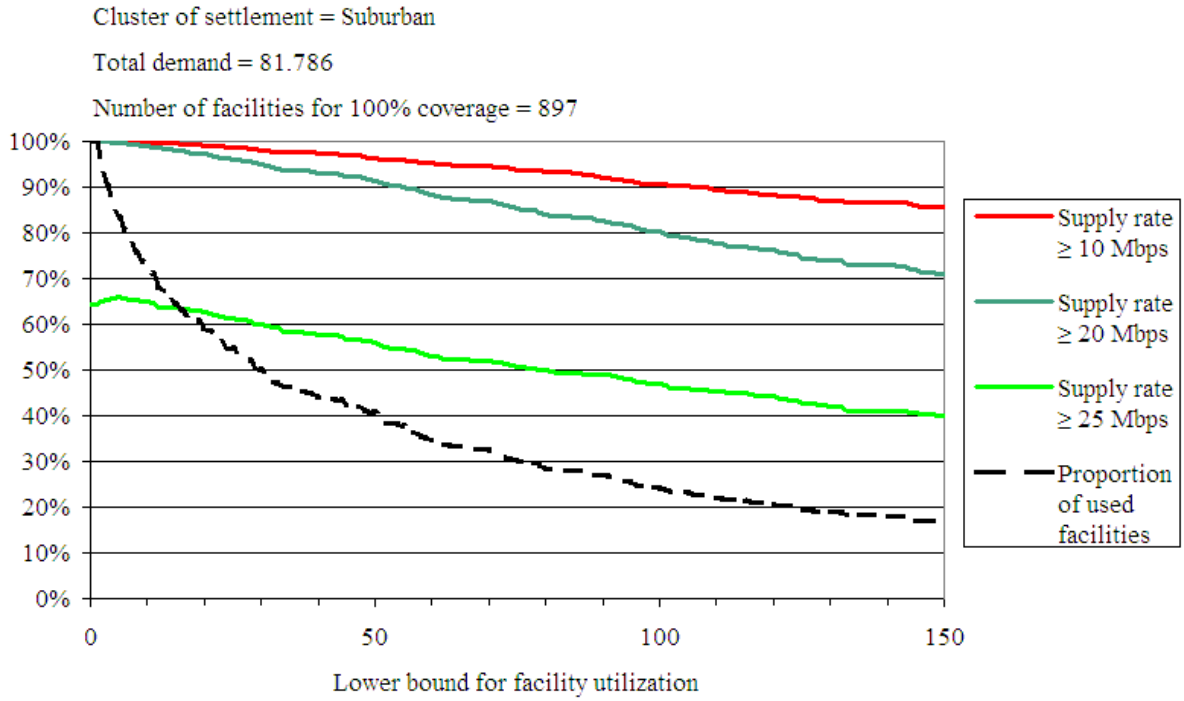


Figure 3.15: Coverage versus prevention of underutilization, CO circle is not enforced, suburban cluster

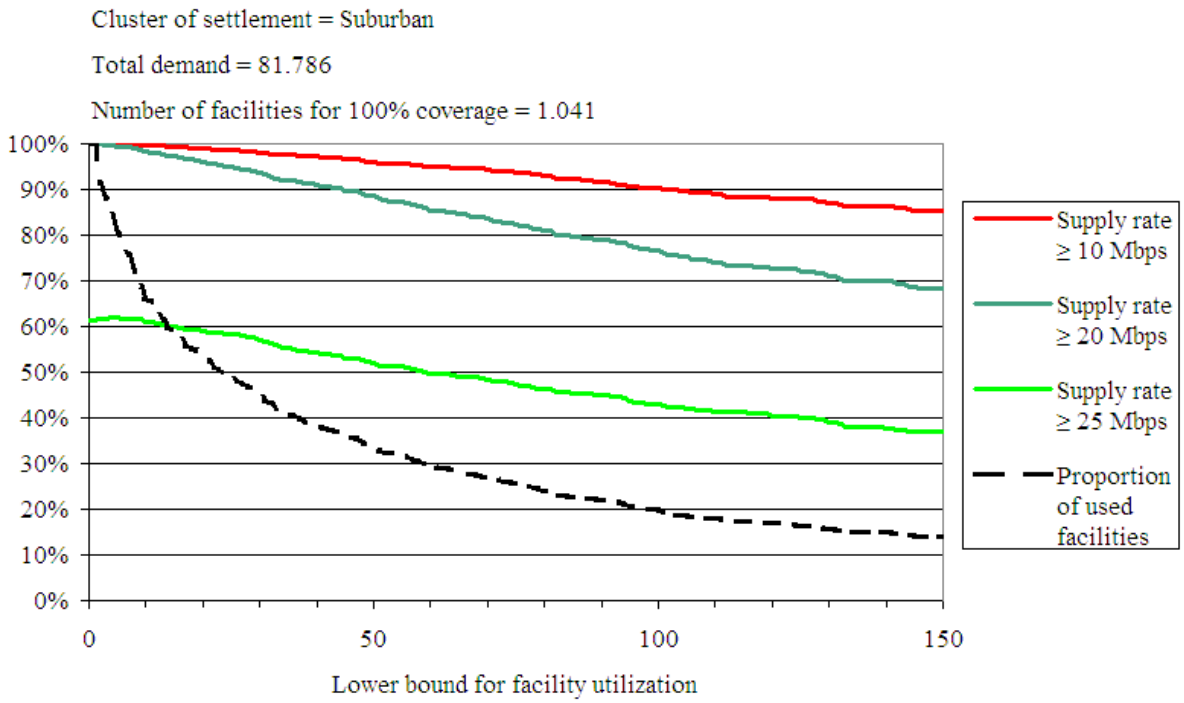


Figure 3.16: Coverage versus prevention of underutilization, CO circle is enforced, suburban cluster

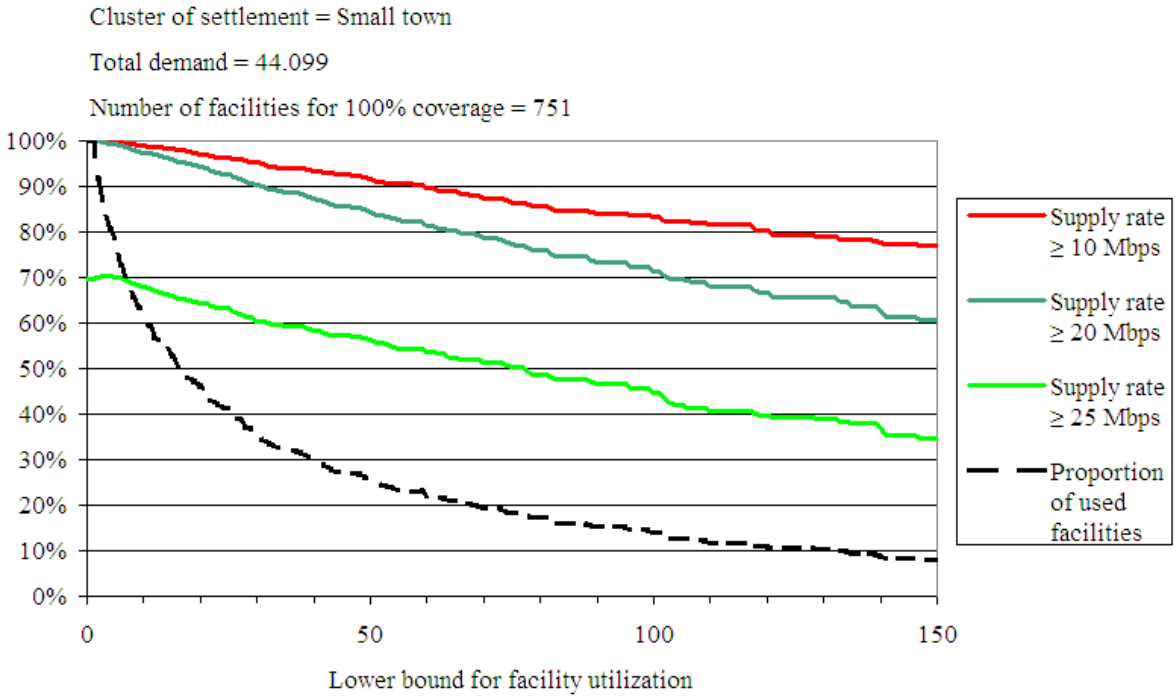


Figure 3.17: Coverage versus prevention of underutilization, CO circle is not enforced, small town cluster

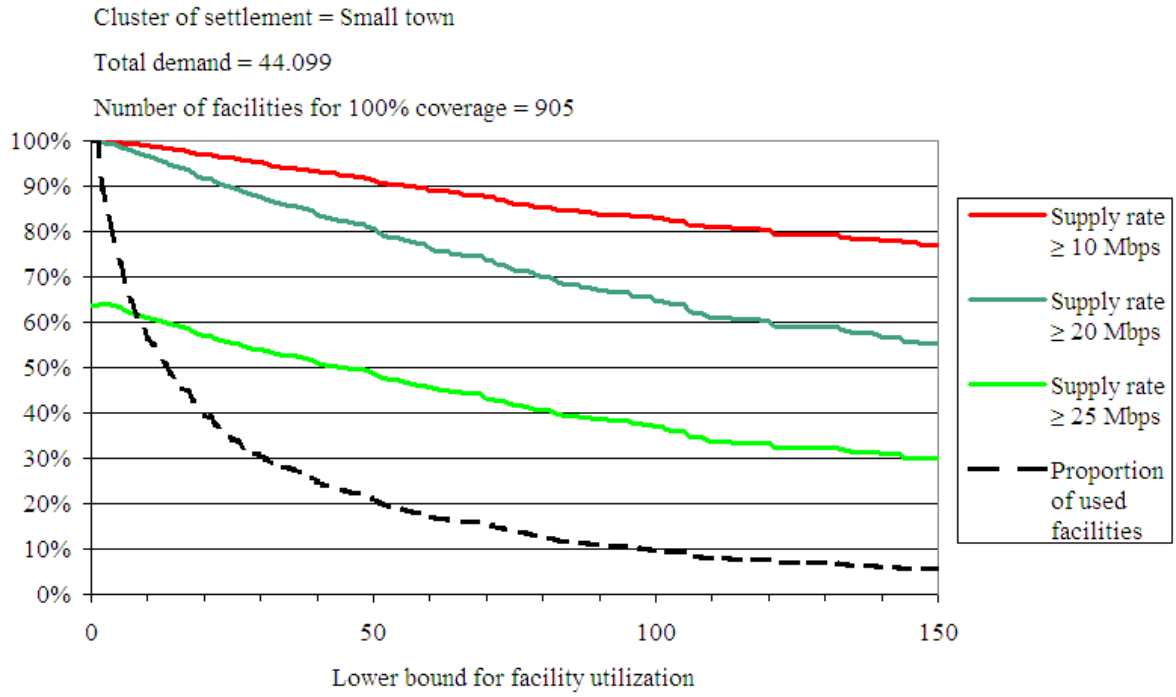


Figure 3.18: Coverage versus prevention of underutilization, CO circle is enforced, small town cluster

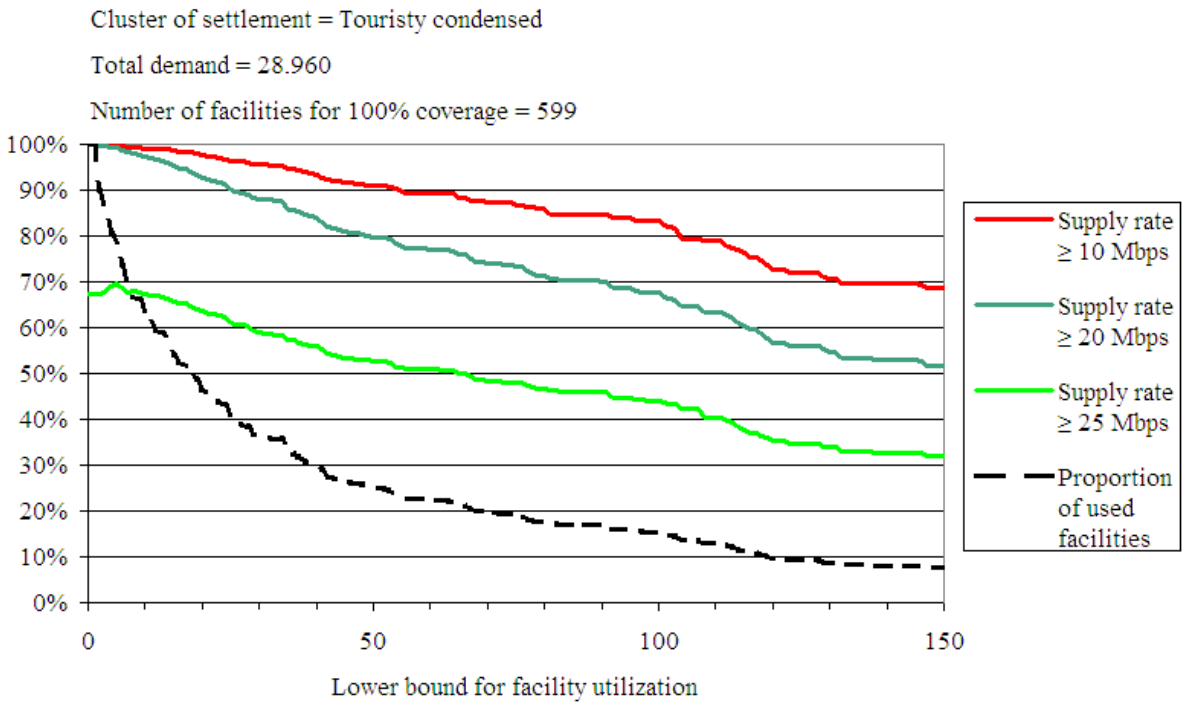


Figure 3.19: Coverage versus prevention of underutilization, CO circle is not enforced, trouristy condensed cluster

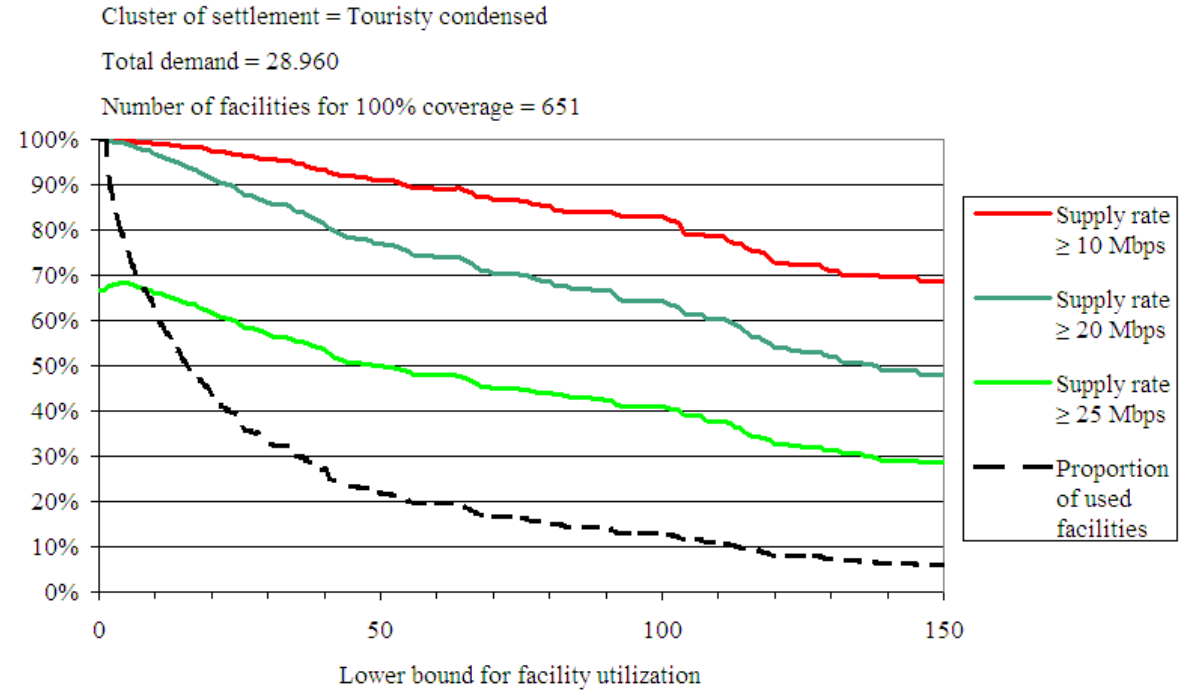


Figure 3.20: Coverage versus prevention of underutilization, CO circle is enforced, trouristy condensed cluster

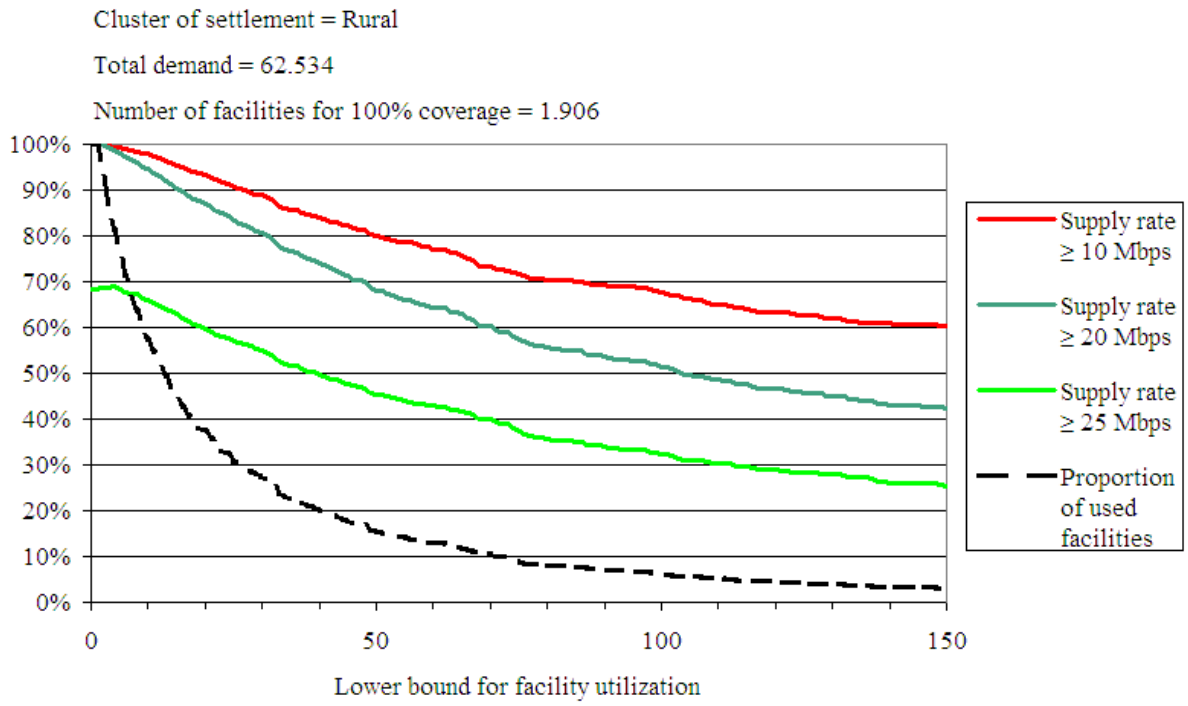


Figure 3.21: Coverage versus prevention of underutilization, CO circle is not enforced, rural cluster

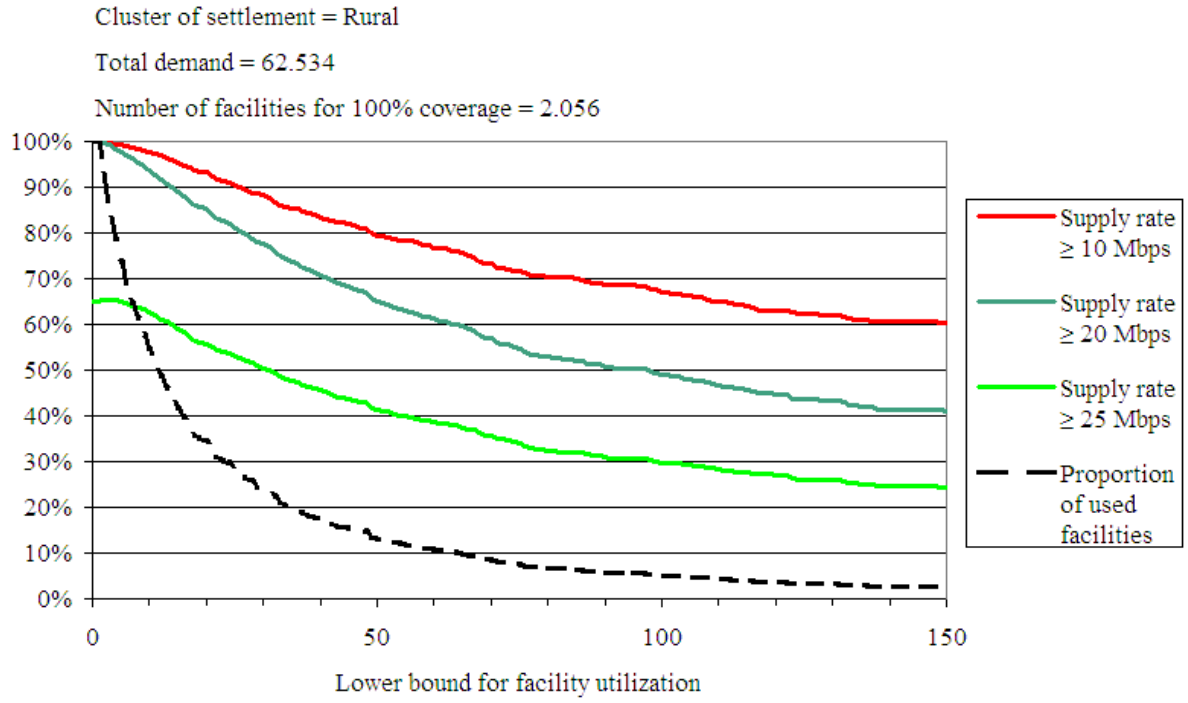


Figure 3.22: Coverage versus prevention of underutilization, CO circle is enforced, rural cluster

Chapter 4

Epilog: Further work, recommendation and caution

At the beginning of project SARU the practitioners — the planners — showed little interest in any kind of limitations with respect to the capacities of facilities (see Chapter 1 Section 1.4.8 "R8: ARU capacities"and 1.5.3 "Structural planning process"). They mainly focused on identifying the copper centers. Therefore, uncapacitated problem formulations were chosen for the modeling approaches. However, in the course of the project the issue of capacities of facilities became more pressing (compare Chapter 2 Section 2.1.6 "Capacity Utilization"and 2.1.9 "Capacity Utilization problem"). Actually, there are two different problems associated with capacity utilization of facilities during the cost optimal planning of internet access networks: **overutilization** and **underutilization**.

4.1 Utilization problems and decision support

Overutilization addresses a situation where too many customers are assigned to a single facility. The optimization literature suggests to solve this problem by introducing capacity limits for facilities. The resulting classes of problems are called capacitated concentrator location problems and capacitated concentrator location problems with multitype concentrators (Chapter 1 Section 1.5.2).

However, in practice the capacitated approach is not completely satisfactory. Providing capacity limits for every potential facility location may lead to a lot of work and cost and may be too time consuming. Therefore, a structural planning process was suggested (Section 1.5.3) which starts by solving an uncapacitated facility location problem, in order to enable the human planner to identify locations of high concentration of demands. Then all necessary information can be collected and adequate restrictions can be imposed on the capacities and types of facilities. Finally, a capacitated facility location problem is solved. If desired, the planner may reenter the information feedback step.

The second problem which is associated with the capacity utilization of facilities turned out to have a greater importance for the practitioners than overutilization. **Underutilization**

describes a solution of the facility location problem where too few customers are assigned to a single facility. To open such a facility does not seem to be economically efficient. The reason why such a facility is situated anyhow is because certain minimum quality standards have to be observed by assigning customers to facilities.

In the context of SARU the **quality** of an internet access network is expressed by transmission rates[1]: the higher the transmission rate, the better for the customer. This approach easily applies to the individual customer. But, how can the overall network quality be measured? Should the average transmission rate be kept high, or should a certain percentage of the customer's demands be covered with a given minimum transmission rate (compare Section 2.2 and Sections 2.5.3 and 2.5.4? In SARU the choice of the project was the concept of **coverage**: thresholds for a minimum percentage of customers who are provided with a minimum transmission rate.

Clearly, resolving underutilization and keeping coverage high are opposed to each other. In the context of the planning rules of SARU (Section 1.4, especially Rule 1.4.1 "Unaltered copper net") underutilization can only be resolved by reducing the number of facilities. This necessarily leads to situations where some customers cannot be provided with the desired transmission rate anymore, whereas others may even loose provisioning completely.

In SARU the number of facilities was directly controlled by means of the k-median approach which can be understood as a facility location problem with a side-constraint on the maximum number of facilities. An interesting side effect of the applied algorithms is that they do not only deliver the solution for a single value k of the number of facilities, but for a whole range of these values (Chapter 2). That way the potential design of access networks in a given local loop can be described by a listing of the number of situated facilities, the achieved percentage of coverage and the corresponding number of minimum utilization of facilities.

Alternatively, the number of facilities can be controlled by imposing a side-constraint on the minimum coverage while simultaneously minimizing cost and thereby minimizing the number of facilities. Or it can be done by demanding a certain minimum of facility utilization and maximizing coverage. In the literature this latter case is also known as lower-bounded facility location (LBFL). Z. Svitkina studied this matter in [52] for the metric facility location problem, i.e. distances between customers and potential facility locations are derived from metric space and not from an underlying graph structure. An attempt to solve the LBFL for undirected networks or directed trees is not known to the author and therefore reason for future work.

In both cases selective solutions, i.e. solutions for only one or just a few different values of the coverage or minimum utilization threshold respectively, turn out to be unsatisfactory in practice. On the one hand, a desirable high coverage rate will rarely lead to acceptable utilization numbers, whereas on the other hand, the targeted high facility utilization will lead to low coverage.

In any case, for practical purposes it is important to be able to inspect and study whole ranges of the parameter space of whatever concept is used to control quality and cost of solutions.

[1] In contrast to connectivity, for example

Such an approach increases the potential of decision support. It provides the planners and the decision makers with a wide range of solutions which can be studied and are the base for new learning. It is probably outside the scope of mathematical optimization to determine the good compromise between coverage and facility location, between service quality and solution cost. Maybe it is rather a task for practitioners.

The challenge for mathematical research lies in developing efficient and exact algorithms or at least good heuristics which are able to provide such kind of decision support.

This goal is not restricted to questions of facility location as they are exclusively considered in this thesis. The issues of overutilization and underutilization are also part of network design. In case of overutilization this is obvious. Space and resources are usually limited. But, inefficient utilization of resources also occurs, if, for example, a new trench has to be dug to provide a single customer with services. So, problems of overutilization and underutilization can be readily extended to the entire problem of network design (Chapter 1 Section 1.5.1) leading to similar considerations as those given before.

This increases the challenge for research even further. The problem of network design requires a more complex set of parameters. The study of the spaces of these parameters is therefore even more time-consuming. A comprehensive and consequently useful communication of the results of such studies to planners and decision makers is a problem in its own right.

4.2 Planning based on multiple quality standards

Another branch of research questions in this area arises from the idea to control the overall network quality not only by one, but two or even more quality thresholds: at least $x\%$ of the customer's demands have to be covered with a minimum transmission rate of M_x Mbps, and $(x + y)\%$ of the customer's demands have to be covered with at least M_y Mbps, where $M_y < M_x$.

The concept of controlling service quality by several thresholds becomes especially useful if network design is part of the modeling approaches and mixed planning strategies are considered, i.e. the planning strategy is comprised of a mixture of different access technologies.

FTTx is an example (see Chapter 1 Section 1.1.7). Fiber to the home will in general be more expensive than fiber to the curb. But, it may be desirable, e.g. because of political reasons, that a certain region is covered with a higher percentage of FTTH than might be achieved by a mere cost controlled mixed planning process. Another example is provided by a combination of the fixed net technologies FTTC and FTTH with mobile broadband access based on LTE (compare Chapter 1 Section 1.1.3). Such a strategy produces three different quality layers which have to be controlled during optimization.

One issue for further mathematical research is to develop appropriate algorithms and to study and describe their effects on solutions. There, especially the differences between iterative and simultaneous algorithms have to be considered. Of course, such multi-technology and multiple quality standard problems can be solved by iteratively applying the same algorithm to different

technologies: the first iteration determines the FTTH customers, the second iteration the FTTC customers and the last iteration covers the rest with LTE. What is the difference in the solution, if the order is changed? What would the advantage of a simultaneous determination be?

And finally, once again the question of how to provide decision support to practitioners and decision makers based on increasingly complex models arise.

4.3 Network planning based on revenue

During project SARU ideas of controlling network planning not only based on cost but also based on revenues were considered and discussed. In the literature problem formulations and models which are concerned with this issue are usually addressed as "prize collecting". For example, J. Hackner applied the prize collecting Steiner tree problem in his thesis [24] to the planning of district heating systems. His solution not only generates the plan for the cost optimal realization of the infrastructure which establishes the heating distribution, but also selects a set of customers for which this system is planned in such a manner that the difference between revenue and cost is maximized, i.e. the heat distribution infrastructure is not necessarily made available to all customers within a district. It may even happen that only a small subset of the customers within this district is connected.

Similar approaches were discussed at Telekom Austria for internet access networks and they were dismissed. Some of the arguments are listed below.

1) **Internet access is is relatively cheap.** The matter is thoroughly discussed in Chapter 1 Sections 1.1.1 and 1.1.2 for internet access products. It can be safely assumed that internet providers will continue to compete by either reducing prices or increasing bandwidth over the next few years. The necessary infrastructure for internet access — a computer — is not too expensive either and many consumers own computers for purposes which have nothing to do with internet access. All computers are compatible with all internet providers, i.e. the consumer does not have to buy a new computer if he wishes to change provider. This has to be seen in contrast to, for example, energy providers and access to district heating systems as described by Hackner in [24]. High investment costs are necessary for both, the provider and the costumer, to establish a fully functional heating system for the customer. Consequently, the customers of energy distribution systems have a strong tendency to stay with their providers over a long period of time. This is not the case in telecommunication markets.

2) **Internet access does not produce high emotional binding of customers.** If internet access is the window to the internet, then the internet provider is like the window cleaner: if you see him, you know something is wrong. A consumer of internet access does not have and actually does not want to be aware of the product he is using. This is the very nature of the product. Unlike expensive watches or cars, the newest and coolest mobile device internet access products are not very useful to brag about. Internet access products produce low emotional binding with their customers, or, in worst case, high but negative feelings. Emotional reactions may only occur in case of malfunction of the product.

Consequently, customer loyalty is very low in internet access markets. Customers are very flexible. If they see better opportunities with a different provider, product or technology, they will take it.

3) **The estimation of future revenues in internet access markets is a complex task.** Revenues depend on product prices and take rates[2]. Consequently, revenue estimation relies on realistic estimations of product prices and customer demands for bandwidth over a longer period of time.

   Recalling again especially Section 1.1.2 in Chapter 1 the description of the Austrian telecommunication market over the last few years presents the picture of a very competitive market where prices fall and transmission rates grow. Furthermore, this process is repeatedly driven by the introduction of new or improved technologies.

   There are no signs and there is no reason to believe that this development will cease during the following years. The fourth generation of mobile phone technologies — LTE — is only one of the new technologies which are going to change the market.

   So, who is going to provide a meaningful and stable estimation of the revenues over the next 10 years?

4) **Individual revenue estimates for a large number of potential customers are not available.** A network planning model which optimizes by maximizing revenue and minimizing deployment cost simultaneously needs to be fed not only with the regional cost structure of the access area under consideration, but also with revenue estimates for the individual customers of that region.

   Beside the problem of the principal estimation of revenues over a long period the problem arises, how should these estimates be made for individuals? There is at least no openly accessible database available which describes the economic situation of a large number of private customers. Should sales agents walk around a perspective area and evaluate who is willing and able enough to take internet products for which the access network is not even planned yet?

Of course, consideration of revenues plays an important role during the planning process. This mainly happens during the pre-selection phase, i.e. during the process when it is decided which local access areas are going to be reconstructed next. Typically, a business case analysis is set up where several access areas are evaluated based on different scenarios — a computation which can be very well supported by Operations Research and methods of facility and network design.

But, once a decision is made and the network is in planning, the most reasonable approach to consider revenue is to control the utilization of facilities. In general a facility of low utilization is not going to provide high revenues, even if the customers in its access area are rich and demanding. Moreover, imposing lower bounds on facility utilization already leads to an exclusion of certain regions and the customers living there, as is manifoldly demonstrated in this thesis.

[2]The fraction of potential customers in a given region who actually really take the product and pay for it.

Appendix A

Abstracts

A.1 Abstract (English)

As indicated by the title this thesis is based on an Operations Research project which was conducted at the Austrian telecommunications provider Telekom Austria between 2006 and 2009. An increasing number of internet users, new internet applications and the growing competition of mobile internet access force fixed line providers like Telekom Austria to offer higher rates for data transmission via their access networks. As a consequence access nets have to be improved which leads to investments of significant size. Therefore, minimizing such investments by a cost optimal planning of networks becomes a key issue.

The main goal of the project was to support the planning process by utilizing discrete optimization methods from the field of network design. The key results which are presented in this thesis are algorithms for facility location. However, before dealing with the theory and the solutions — in practice as well as in this thesis — a thorough analysis of the stated problem is undertaken.

To begin with the telecommunication market before 2006 and especially between 2006 and 2009 is reviewed to provide some background information. The industry had already developed different strategies to improve fixed line infrastructure. Their relevance for the stated problem is presented. Furthermore, the most important problem specifications as they were collected in cooperation with the practitioners are listed and discussed in detail. A first solution was based on a dynamic program for solving the facility location problem which was derived from the specifications. The statement of conditions for the optimality of this algorithm and their proofs conclude Chapter 1.

It turned out that this first solution did not provide the desired result. It rather fostered the discussion process between Operations Researches and practitioners. New specifications were added to the existing list. The planners dismissed these first solutions because they were not efficient enough. These solutions contained facilities which were underutilized, i.e. too few customers were assigned to such facilities. To overcome this problem facilities of low utilization had to be removed from the solutions. The remaining facilities were rearranged in a way to maximize the coverage with a certain minimum transmission rate. This strategy was realized

by adapting the concept of the k-median problem: The number of facilities is bounded whereas simultaneously the number of optimally supplied customers is maximized. Then for different bounds the minimum facility utilization is reported. That way the practitioner is enabled to find the optimal balance between efficient facility utilization and coverage of customer demands.

After sketching the events and discussions which made further development necessary and listing the additional specifications, the theory of the k-median problem is presented and a basic algorithm from the literature is cited. For the specific requirements of the given problem a variant of the algorithm is developed and described at the end of Chapter 2: The algorithm from the literature inserts facilities one by one into the solution that way approaching the bound in an ascending manner. However, since the expected number of facilities is usually large it is more advantageous to approach the bound from above in a descending manner.

Finally, an extensive empirical study of 106 different local access areas is presented. The main purpose of this demonstration is to give a concrete impression of how the adapted and developed methods can be utilized in preparation of the planning process by studying strategic questions (e.g. CO circle enforcement, balancing between facility utilization and coverage) and by providing information (runtime) which is useful to set up an appropriate working environment for the future users. Additionally, the two variants of the k-median algorithm — the ascending and the descending method — can be compared.

A.2 Abstract (German)

Wie der Titel bereits andeutet bezieht sich diese Dissertation auf ein Operations Research Projekt, dass der Österreichische Telekommunikationsanbieter Telekom Austria in den Jahren 2006 bis 2009 durchführte. Die wachsende Zahl von Internet Nutzern, neue Anwendungen im Internet und die zunehmende Konkurrenz von mobilem Internet zwingen Festnetzbetreiber wie Telekom Austria ihre Produkte für den Internet Zugang mit höheren Bandbreiten zu versehen. Zwangsläufig müssen die Zugangsnetze verbessert werden, was nur mit hohen Investitionskosten erreichbar ist. Aus diesem Grund kommt der kostenoptimalen Planung solcher Netzwerke eine zentrale Rolle zu.

Ein wesentliches Projektziel war es, den Planungsprozess mit Methoden der diskreten Optimierung aus dem Bereich Network Design zu unterstützen. Die Ergebnisse, die in dieser Dissertation beschrieben werden, beschäftigen sich mit Algorithmen aus dem Gebiet Facility Location (Bestimmung von Versorgungsstandorten). Vor der Präsentation der dazugehörigen Theorie und ihrer Anwendung auf die gestellten Problem werden zweitere gründlich analysiert.

Zunächst wird der Telekommunikationsmarkt bis 2009 mit speziellem Fokus auf den Zeitraum zwischen 2006 und 2009 beschrieben. Die Telekommunikationsindustrie hatte bereits einige Strategien zur Verbesserung der Netzwerkinfrastruktur entwickelt. Ihre Relevanz für die gestellten Probleme wird herausgearbeitet Dem folgt eine Auflistung der Problemspezifikationen, wie sie in der Evaluierungsphase des Projekts mit den beteiligten Anwendern erstellt wurde. Mit Hilfe eines dynamischen Programms wird die gestellte Fragestellung unter Berücksichtigung aller Spezifikationen gelöst. Eine Auflistung von Bedingungen, wann dieser Algorithmus die optimale Lösung liefert, und die dazugehörigen Beweise beschließen Kapitel 1.

In der Folge stellte sich allerdings heraus, dass die Praktiker mit dieser ersten Lösung nicht zufrieden waren. Die Liste der Spezifikationen war nicht vollständig. Sie musste verändert und erweitert werden. Mangelnde Effizienz machte die Lösungen für die Praxis unbrauchbar. Die Lösungen enthielten Versorgungsstandorte, die minder ausgelastet waren (underutilized), d.h. diesen Standorten waren zu wenige Kunden zugeordnet worden. Solche Lokationen mussten aus den Lösungen entfernt werden. Dann aber waren die Verbleibenden so zu repositionieren, dass die Versorgung mit einer vorgegebenen Mindestübertragungsrate für die größtmögliche Menge an Kunden sichergestellt werden konnte. Diese Strategie wurde mit Hilfe des Konzepts der k-Mediane umgesetzt: Unter der Nebenbedingung, dass die Anzahl der Standorte durch eine Konstante k beschränkt ist, wird die optimale Zuordnung von Kunden zu Versorgungsstandorten, d.h. ihre Versorgung, gesucht. Anschließend löst man dann k-Median Probleme für verschiedene Werte von k und bestimmt die Mindestauslastungen und Versorgungsraten, die diese Lösungen erzielen. Dieses Vorgehen versetzt den Anwender in die Lage unter verschiedenen Lösungen zwischen effizienter Auslastung der Versorgungsstandorten und der Höhe der Versorgungsraten balancieren zu können.

In Kapitel 2 werden zunächst die Ereignisse und Diskussionen beschrieben, die eine Änderung der Lösungsstrategie notwendig machten, und die geänderten bzw. neuen Spezifikationen werden präsentiert. Dem folgt die Vorstellung der Theorie der k-Mediane inklusive der Beschreibung eines Algorithmus aus der Literatur. Am Ende des 2. Kapitels wird eine Variante dieses

Algorithmus entwickelt, der für die spezifischen Anforderungen noch besser geeignet ist: Der Algorithmus aus der Literatur fügt Lokationen schrittweise in die Lösung ein, d.h. pro Iteration erhöht sich die Anzahl der Versorgungsstandorte um einen, bis die maximale Anzahl von Lokationen erreicht ist. Im Falle von Zugangsnetzen ist die zu erwartende Anzahl von Standorten aber eher groß. Daher ist es vorteilhafter die gewünschte Anzahl von oben, durch Reduktion der Anzahl von Versorgungsstandorten in der Lösung zu erreichen.

Kapitel 3 liefert eine extensive empirische Analyse von 106 verschiedenen Zugangsnetzen. Konkreter Zweck dieser Demonstration ist es einen Eindruck zu vermittelt, wie man die entwickelten und adaptierten Methoden bei der Vorbereitung des Planungsprozesses einsetzen kann. So ist es einerseits möglich strategischen Fragestellungen vorab zu analysieren (z.B. Effekt der Erzwingung des HV Kreises, Balance zwischen Auslastung der Versorgungsstandorte und der Versorgungsrate), und andererseits Vorschläge für passende Planungsprozesse für die Anwender zu entwickeln (z.B. durch Laufzeitanalysen). Zusätzlich werden die beiden Methoden zur Lösung des k-Median Problems, die in dieser Arbeit vorgestellt werden, noch bzgl. ihres Laufzeitverhaltens verglichen.

Appendix B

Curriculum Vitae

Personal information:

Name:	Bertram Wassermann
Title:	Mag.rer.nat
Address:	Nelkengasse 9
	Austria 3100 St. Pölten
Date of Birth:	March 11, 1966
Place of Birth:	Lienz / Osttirol
Citizenship:	Austria
Martial status:	Married with Veronika Ott since 2011
Languages:	German and English

Education:

1972 - 1976	Grade school, Volksschule Dellach
1976 - 1980	Junior high school, Hauptschule Kötschach
1980 - 1984	High school, Bundesoberstufenrealgymnasium Hermagor
June 26, 1984	School leaving examination (Matura) in the following subjects: Mathematics, Philosophy, Music, English, German and Latin
1984 - 1992	University of Klagenfurt, Mathematics with special focus on Algebra, Number Theory and Statistics
March 19, 1992	Graduation to Magister der Naturwissenschaften (Master of Science) at University of Klagenfurt
1992 - 1994	Participation in a Master program at Penn State University in State College, Pennsylvania, USA as a Fulbright Scholar
August 27, 1994	Graduation to Master of Arts at Penn State University
1994 - 1996	Further studies and teaching assistantship at Penn State University
2004 - 2011	PhD studies in Applied Statistics and Operations Research at University of Vienna

Scholarships and awards:

1992 - 1994	Fulbright Scholar at Penn State University
2005	Förderungsstipendium Universität Wien
2007	Telekom Austria Top Performer Bonus

Occupational history:

Present (2011)	A1 (Formerly Telekom Austria) and University of Vienna
Since 2006	Teaching at University of Vienna (Business statistics)
Since 2005	Member of the A1 Operations Research group responsible for • Statistics, • Discrete optimization and • Simulation
2002 - 2004	at A1 as data miner and marketing statistician responsible for • Churn prediction, • Cross- and upselling models and • Customer segmentation
1997 - 2002	Leading position at the market research company INFO Research International (Vienna) with the following fields of responsibility • CATI (Computer Aided Telephone Interviewing) and • CAPI (Computer Aided Personal Interviewing) Manager • Data Processing and • Statistical analysis (Multivariate methods like conjoint analysis, segmentations, cluster analysis, ...)
1996 - 1997	Civil service at a senior citizen home in Carinthia, Austria
1994 - 1996	Teaching assistantship at Penn State University
Until 1992	Several freelance positions and project participation at University of Klagenfurt, AMS Kärnten (Federal Employment Office) and Kuratorium für Verkehrssicherheit (KfV)

Bibliography

[1] A. V. Aho, J. E. Hopcroft, and U. D. Ullman. *Data Structures and Algorithms*. 1983.

[2] A. Arulselvan, A. Bley, S. Gollowitzer, I. Ljubic, and O. Maurer. MIP modeling of incremental connceted facility location. In *Proceedings INCO 2011*, 2011.

[3] V. Arya, N Garg, R. Khandekar, A. Meyerson, K. Munagala, and V. Pandit. Local search heuristics for k-median and facility location problems. *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, 2001.

[4] Y. Bartal. Probabilistic approximations of metric spaces and its algorithmic applications. *IEEE Symposium on Foundations of Computer Science*, pages 184–193, 1996.

[5] Y. Bartal. On approximating arbitrary metrics by tree metrics. *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, 1998.

[6] R. Benkoczi and B. Bhattacharya. A new template for solving p-median problems for trees in sub-quadratic time. *Lecture Notes in Computer Science*, 3669/2005:271–282, 2005.

[7] R. Benkoczi, B. Bhattacharya, M. Chrobak, L. Larmore, and W. Rytter. Faster algorithms for k-medians in trees. *Mathematical Foundations of Computer Science, Volume 2747/2003*, 2003.

[8] Robert Benkoczi. Cardinality constrained facility location problems in trees. *PhD thesis, School of Computing Science, Simon Fraser University, Burnaby, BC, Canada*, 2004.

[9] M. Charikar, S. Guhay, E. Tardos, and D. Shmoys. A constant-factor approximation algorithm for the k-median problem. *Journal of Computer and System Sciences. Originally appeared in Proc. 31st Annual ACM Symposium on Theory of Computing (STOC99)*, 2002.

[10] M. Chrobak, L. Larmore, and W. Rytter. The k-median problem for directed trees, extended abstract. *Mathematical Foundations of Computer Science, Volume 2136/2001*, 2001.

[11] Marek Chrobak, Claire Kenyon, and Neal Young. The reverse greedy algorithm for the metric k-median problem. *Computing and Combinatorics, Lecture Notes in Computer Science*, 3595:654–660, 2005.

[12] Cisco. Cisco visual networking index: Forecast and metholody, 2009-2014. *http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827*, 2010.

[13] G. Cornuejols, G. L. Nemhauser, and L. A. Wolsey. The uncapacitated facility location problem. *P. Mirchandani and R. Francis, editors, Discrete Location Theory*, page 119 171.

[14] Nell B. Dale and Susan C. Lilly. *Pascal Plus Data Structures*. 4 edition, 1995.

[15] Austrian newspaper Der Standard. 20 Jahre Internet.at. 7./8. August 2010.

[16] Ericsson. Ericsson consumerlab mobile broadband study 2009 - austria. 2009.

[17] D. Erlenkotter. A dual-based procedure for uncapacitated facility location. *Operations Research*, 26(6):992 – 1009, 1978.

[18] V.J.M.F. Filho and R.D. Galvão. A tabu search heuristic for the concentrator location problem. *Location Science 6 (1998)*, pages 189 – 209, 1998.

[19] Rundfunk & Telekom Regulierungs GMBH. RTR telekom monitor 3. quartal 2007. *www.rtr.ar/de/tk/ClusterBB*, 2007.

[20] Rundfunk & Telekom Regulierungs GMBH. RTR telekom monitor 2. quartal 2010. *www.rtr.ar/de/tk/ClusterBB*, 2010.

[21] Stefan Gollowitzer and Ivana Ljubic. MIP models for connected facility location: A theoretical and computational study. *Computers & Operations Research*, 38/2:435–449, 2011.

[22] Eric Gourdin, Martine Labbé, and Hande Yaman. Telecommunication and location. 2001.

[23] Sudipto Guha and Samir Khuller. Greedy strikes back: improved facility location algorithms. In *Proceedings of the ninth annual ACM-SIAM symposium on Discrete algorithms*, SODA '98, pages 649–657, Philadelphia, PA, USA, 1998. Society for Industrial and Applied Mathematics.

[24] Johannes Hackner. *Energiewirtschaftlich optimale Ausbauplanung kommunaler Fernwärmesysteme*. PhD thesis, Technical University of Vienna, 2004.

[25] S. L. Hakimi. Optimum distribution of switching centers in a communication network and some related graph-theoretic problems. *Operations Research, 13, pp. 462475*, 1965.

[26] Kaj Holmberg, Mikael Ronnqvist, and Di Yuan. An exact algorithm for the capacitated facility location problems with single sourcing. *European Journal of Operational Research*, 113:544 – 559, 1999.

[27] F.K. Hwang, D.S. Richards, and P. Winter. The steiner tree problem. *Annals of Discrete Mathematics, New York: North-Holland*, 53, 1992.

[28] Integral. AIM - austrian internet monitor april bis juni 2010. 2010.

[29] Kamal Jain and Vijay V. Vazirani. Approximation algorithms for metric facility location and k-median problems using the primal-dual schema and lagrangian relaxation. *Journal of the ACM*, (48):274–296, 2001.

[30] O. Kariv and S. L. Hakimi. An algorithmic approach to network location problems i: The p-centers. *SIAM Journal on Applied Mathematics, 37:539560*, 1979.

[31] O. Kariv and S. L. Hakimi. An algorithmic approach to network location problems ii: The p-medians. *SIAM Journal on Applied Mathematics, 37:539560*, 1979.

[32] J.G. Klinkewicz and H.Luss. A lagrangian relaxation heuristic for capacitated facility location with single source constraints. *Journal of Operational Research Society*, 37:495 – 500, 1986.

[33] M. R. Korupolu, C. G. Plaxton, and R. Rajaraman. Analysis of a local search heuristic for facility location problems. *Journal of Algorithms, 37:146188*, 2000.

[34] Andreas Kriesel. Basic idea to DB tree algorithm. Personal communication, 2006.

[35] C.Y. Lee. An algorithm for the design of multitype concentrator networks. *Journal of Operational Research Society*, 44:471 – 482, 1993.

[36] Retsef Levi and David B. Shmoys. Lp-based approximation algorithms for capacitated facility location. In *Proceedings of the 5th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA*, pages 206–218. ACM Press, 2004.

[37] B. Li, M. Golin, G. Italiano, X. Deng, and K. Sohraby. On the optimal placement of web proxies in the internet. *IEEE InfoComm'99, pages 1282-1290*, 1999.

[38] Jyh-Han Lin and Jeffrey Scott Vitter. Approximation algorithms for geometric median problems. *Information Proceeding Letters*, (44):245–249, 1992.

[39] Jyh-Han Lin and Jeffrey Scott Vitter. ϵ-Approximations with minimum packing constraint violation (extended abstract). In *Proceedings of the twenty-fourth annual ACM symposium on Theory of computing*, STOC '92, pages 771–782, New York, NY, USA, 1992. ACM.

[40] Ivana Ljubic. *Exact and Memetic Algorithms for Two Network Design Problems.* PhD thesis, Technical University of Vienna, 2004.

[41] Ivana Ljubic, Peter Putz, and Juan-Jos Salazar-Gonzlez. Exact approaches to the single-source network loading problem. *Technical Report 2009-05*, Univerity of Vienna, 2009.

[42] A. Mirzaian. Lagrangien relaxation for the star-star concentrator location problem: Approximation algorithm and bounds. *Networks*, 15:1 – 20, 1985.

[43] MR&BI. Telekom Austria internal product database.

[44] J. Reese. Solution methodes for the p-median problem: an annotated bibliography. *Networks*, 48/3:125–142, 2006.

[45] Mikael Ronnqvist, Suda Tragantalerngsak, and John Holt. A repeated matching heuristic for the single-source capacitated facility location problem. *European Journal of Operational Research*, 116(1):51 – 68, 1999.

[46] R. Schnepfleitner, Z. Daroczi, R. Kalasek, and W. Feilmayr. GIS-Einsatz bei der Regulierungsberhörde für Telekommunikation. *Business Geographics, M. Fally, J.Strobl (Hrg.)*, pages 106–121, 2001.

[47] Rahul Shah and Martin Farach-Colton. Undiscretized dynamic programming: faster algorithms for facility location and related problems on trees. In *Proceedings of the thirteenth annual ACM-SIAM symposium on Discrete algorithms*, SODA '02, pages 108–115, Philadelphia, PA, USA, 2002. Society for Industrial and Applied Mathematics.

[48] M. Shindler. Approximation algorithms for the metric k-median problem, 2008. *Internet, URL: www.cs.ucla.edu/∼shindler/shindler-kMedian-survey.pdf, Last accessed: 27.9.2010.*

[49] D. Shmoys, E. Tardos, and K. Aardal. Approximation algorithms for facility location problems. *29th ACM Symposium on Theory of Computing*, pages 265–274, 1997.

[50] R. Sridharan. A lagrangian heuristic for the capacitated plant location problem with single source constraints. *European Journal of Operational Research*, 66:305–312, 1993.

[51] R. Sridharan. The capacitated plant location problem. *European Journal of Opterional Research*, 87:203–213, 1995.

[52] Zoya Svitkina. Lower-bounded facility location. *ACM Transactions on Algorithms*, 6/4, 2010.

[53] A. Tamir. An $O(pn^2)$ algorithm for the p-median and related problems on tree graphs. *Operations Research Letters 19 (1996) 59-64*, 1996.

[54] M. Thorup. Quick k-median, k-center, and facility location for sparse graphs. *SIAM Journal of Computation*, 34(2):405–432, 2005.

[55] Alessandro Tomazic and Ivana Ljubic. A GRASP algorithm for the connected facility location problem. *Applications and the Internet, IEEE/IPSJ International Symposium on*, 0:257–260, 2008.

[56] A. Vigneron, L. Gao, M. Golin, G. Italioan, and B. Li. An algorithm for finding a k-median in a directed tree. *Information Processing Letters*, 74:81–88, 2000.

[57] B. Wassermann and I. Ljubic. Project SARU, How to situate Access Remote Units and construct a minimal cost fibre optic cable network. *Presentation at EURO XXI — Rekijavik, Iceland*, 2006.

[58] Wikipedia. http://en.wikipdia.org/wiki/fttx. 10.9.2010.

[59] GSM world. http://www.gsmworld.com/about-us/history.htm 5.8.2010. *Internet*, 2010.

[60] Hande Yaman. *Concentrator location in telecommunications networks.* Springer, 2005.