INTRODUCTION

Are you the type of person that likes waiting at a red light? Do you enjoy being stopped by uncoordinated signals? If not, you are invited to read on to find out how the coordination of traffic signals can help to reduce delays and, thus, avoid having to wait at red lights.

In urban areas there is a strong demand for transportation. Probably, the most sustainable means of transportation are the public ones, like buses, trams or the underground. Nevertheless, not all demands for transportation can be covered by the public sector. A major part of the overall transportation in cities is composed of individual drivers.

There are different ways to improve road traffic conditions in city areas. However, infrastructural arrangements like broadening streets or even building new ones to cope with increasing demands are often not an appropriate option, due to high costs or space limitations. Instead, intelligent means of traffic control are required to solve today's road traffic problems, like high delays, narrow capacities or traffic jams.

When speaking of ways to control traffic, *traffic lights* or *traffic signals* are of primary importance. A clever adjustment of the signal settings surely helps to reduce delays and increase capacities, thereby avoiding traffic jams. Today, intelligent computer-aided traffic control signals are even capable of reacting to different traffic situations. Namely, they adapt their settings to the respective demands at the junctions.

However, there are traffic scenarios where the traffic responsive signals reach their limit. For example, when there is constant and high traffic volume, the responsive signals repeatedly apply similar control strategies. Therefore, they behave comparably to fixed time traffic signals. These fixed time traffic signals repeatedly respond to a prescribed signal timing program and not to the actual traffic conditions. Hence, research into fixed time traffic signals and their control strategies is an ongoing endeavor.

Operating fixed time signals offers different means of controlling traffic. On the one hand, some signal parameters' adjustments influence the traffic flow locally at a single junction. In many situations, such a local calibration of the signals turns out to be sufficient to cope with the aforementioned problems. On the other hand, in situations with constant high traffic, another non-local control strategy for the fixed time signals becomes more significant: the coordination of the signals.

Coordinating traffic signals means the following: coupling of signals via a parameter called *offset*. This quantity specifies how green phases of different signals are shifted (or offset) to each other. Most prominent coordination objectives are so-called "green waves", where vehicles travel without being impeded by a signal showing red. Nevertheless, when considering networks of signals instead of arterials of signals, it is often not possible to adjust green waves for the whole network. Instead, the goal "green wave" has to be replaced by a more practical term like "minimum possible delay". Hereby, the item "delay" refers to waiting times of vehicles facing red at the signals.

Many approaches and models have been proposed in order to find good coordinations of signals in networks. Still, the majority of them reveal shortcomings either way, be it unrealistic modeling of real-world circumstances or the fact that they do not give guarantees for their solution quality.

To summarize, there is a need for a mathematical optimization approach for coordinating fixed time traffic signals in networks.

This discussion on new required control strategies for fixed time signals is not a theoretical one. Rather, the industry, i.e., traffic companies that plan, manage, and control traffic, demands applicable approaches for coordinating traffic signals in networks.

As an indication thereof, we briefly report on an industry project that emerged between the TU Berlin and the PTV AG, which is a traffic planning software company from Karlsruhe, Germany. In this project, the aim was to develop mathematical optimization software to coordinate fixed time traffic signals in networks. During the project, we developed a mixed-integer linear programming approach, which minimizes the delay of vehicles in a network by adjusting optimal offsets. However, several other functionalities were incorporated in the model. The outcome of the project with the PTV, though, is a concrete implementation of the optimization approach, which is about to be included in PTV software soon.

In our mixed-integer linear program (MIP) for the coordination of traffic signals, a particular physical constraint has been modeled. This constraint, which we will therefore call "Cycle Constraint", has to be formulated for all cycles $\ell \in \mathcal{C}$ of the graph G that represents the network of signalized junctions. It suffices, however, to state the cycle constraints for the elements of a cycle basis. This then implies these constraints for *all* cycles of G. Depending on the respective application, though, it has to be a cycle basis with a certain property. In our case of a MIP for coordinating fixed time signals in networks one has to define the cycle constraints for the elements of an *integral* cycle basis.

This means that any integral cycle basis can be used to define the cycle constraints for our MIP. Although any two integral cycle bases lead to MIPs with equal optimal objective value, the computational behavior of 'their' MIPs may be different. Observe that this may be of importance, since we are considering networks of large size where one may not come up with optimal solutions.

A quantity one can use to compare cycle bases is the so-called *width* of a basis. Loosely speaking, the width of a basis is defined as the product—over all cycles of a basis—of the number of possible values that the integer variable for the cycle constraint for that cycle can take. Thus, the width of a basis gives an impression of the size of the MIPs feasibility region. The hope is that the smaller the width of a basis, the better the corresponding MIP performs.

Among the class of integral cycle bases strictly fundamental cycle bases are a prominent subclass. For a graph $G = (V, E)$, a strictly fundamental cycle basis B is defined by a spanning tree T of G. In particular, the cycles of B are exactly the ones induced by non-tree edges of T with respect to the graph G. Then, in the Minimum Strictly Fundamental Cycle Bases (MSFCB) problem one seeks a spanning tree T that induces a basis of minimum length.

In 1982, Deo et al. [DKP82] proved the MSFCB problem to be NP-complete for general graphs. Since then, many heuristics for the MSFCB problem have been proposed. Nevertheless, for comparing the results of these heuristics, i.e., whenever concrete experiments were conducted, sample graph classes were considered. Besides random graphs, grid graphs are the most important such graph class.

Grid graphs are also of interest for the following two reasons. First, considering the coordination of traffic signals, many real-world networks have a grid-like structure. One only has to think of the layout of central areas in north american cities. Second, for the MSFCB problem, grid graphs turn out to be computationally tricky. This fact is probably due to an extreme amount of symmetric spanning trees on grids.

In 1995, Alon et al. [AKPW95] proved that for square grids with n vertices, the size of an optimal solution to the MSFCB problem is in $\Theta(n \log n)$. Still, we decided to investigate bounds on the optimal value of an MSFCB on a square grid having the form

$$c_1 \cdot n \cdot \log_2 n - o(n \log n) \;\leq \mathrm{OPT}_n \;\leq\; c_2 \cdot n \cdot \log_2 n + o(n \log n).$$

We could prove that the above statement is true for $c_1 = 1/12$ and $c_2 = 0.979$, respectively.

An optimization problem that is closely related to the MSFCB problem is the one of finding a tree t-spanner with minimal t, [CC95]. In this problem, one seeks a spanning tree T for a given general graph G, such that the maximum over all pairs of vertices $(u, v) \in V \times V \setminus \{(v, v) \mid v \in V\}$ of the ratio $d_T(u, v)/d_G(u, v)$ is minimal. Here, $d_G(u, v)$ refers to the length of a shortest path between u and v in G. The quantity $d_T(u, v)$ denotes the length of the path between u and v in T.

The relation between finding a minimal tree t-spanner of a graph and an MSFCB can be noticed when considering the following unified notation for tree spanner (UNTS) problems. In the UNTS, a problem is defined through a triple

$$(\texttt{goal}, \texttt{domain}, \texttt{term})\,.$$

Here, `goal` is either the maximum stretch or the average stretch. Second, as `domain`, either all non-tree edges or all edges or all pairs of vertices are considered. Finally, `term` may be one of the following four: $d_T(u, v)$ or $d_T(u, v)/d_G(u, v)$ or $d_T(u, v) + w(e)$ or $d_T(u, v)/w(e)$, with $w(e)$ denoting a weight of an edge. Although not all combinations of `goal`, `domain` and `term` are possible, there remain 20 tree spanner problems, classified by the UNTS. Interestingly, these 20 notationally different problems collapse to 12 with a general weight function w and to only five, when considering 0/1-weights on the edges.

Among these five problems that do not coincide even in the unweighted case, are—besides the MSFCB problem—prominent optimization problems like the "Minimum Average Stretch Spanning Tree Problem" [PT01], the "Shortest Total Path Length Spanning Tree Problem" [DKP82, WCT00] and the "Minimum Diameter Spanning Tree Problem" [HL02].

Generally speaking, the UNTS provides a classification of related problems, which had not been realized as such before. Hence, interconnections can be revealed and properties like complexity status or inapproximability factors can be carried forward between the problems.

The chronology of topics within this introductory part is reflected in the organization of the thesis.

Outline of the Thesis

In Chapter 1, we investigate the Network Signal Coordination (NSC) problem. After a short introduction of the most important traffic engineering terms related to traffic signals, we first examine a study of related work. In the case of the NSC problem this turns out to be of importance in order to clearly restrain the problem from other optimization tasks regarding traffic signals. Thereafter, we formally define the NSC problem and report on similarities to the related "Periodic Event Scheduling Problem" (PESP). Moreover, we take advantage of the PESP in order to prove the NSC problem to be NP-complete. Then, in the main part of the chapter, we present a model for the NSC problem. In particular, we develop in detail a mixed-integer linear programming (MIP) approach to solve the NSC problem. We conclude the chapter with a discussion of possible applications in practice of an NSC model in general and the MIP approach in particular.

In the mixed-integer linear programming formulation for the NSC problem, the subproblem of finding appropriate integral cycle bases arises. In Chapter 2, we consider the problem of finding Minimum Strictly Fundamental Cycle Bases (MSFCB) on grid graphs. In particular, we investigate lower and upper bounds for this problem. As for the lower bounds, we consider both combinatorial approaches and mixed-integer linear programming formulations of the problem that we enrich with several additional cuts. Thereafter, we consider upper bounds for the MSFCB problem on grids. In particular, we construct trees by making intensive use of recursively defined sub-structures. We conclude the chapter with an experimental section in which we provide benchmark results for the MSFCB problem on grids which help evaluating further research.

The MSFCB problem can be interpreted as a problem of finding a spanning tree that minimizes the sum of path lengths between particular pairs of vertices in a given graph. Interestingly, finding minimum average stretch tree spanners or min-max stretch tree spanners of graphs can be interpreted in a very similar way. In Chapter 3, we provide a classification of several problems that aim at finding spanning trees in a graph, which minimize the average or the maximum value of certain distances between particular pairs of vertices in a graph. We propose a unified notation for these problems, which include several prominent problems in combinatorial optimization. With this notation at hand, we identify all coincidences and anti-coincidences of these problems. Moreover, we provide a missing complexity status for one of the problems and observe that an inapproximability result of one of the problems can in fact be applied to another problem too, where it had previously been unknown.

In Chapter 4, the experimental work that is related to the Network Signal Coordination (NSC) problem is presented, thereby coming full circle back to the first chapter. The experiments conducted are threefold: First, we perform a solver comparison at some example instances for our MIP model. Here, we compare the MIP solvers CPLEX, MOPS, and SCIP with respect to their ability to find good solutions in short time. In particular, we run two series of experiments using once the default MIP solver settings and once settings that emphasize the finding of good primal solutions. Second, we report on the influence of cycle bases to the computational behavior of the MIP for the NSC problem. This experiment is of general interest, because a positive influence of short bases on computation times of mixed-integer programming formulations of practical applications is expected although very few studies actually proved it. So, in particular, we investigate the correlation between the width of a cycle basis and the lower bound obtained by a MIP computation of 10 seconds. Finally, the third series of experiments is probably the most important one: we evaluate our model by carrying out case studies. Namely, we consider the real-world inner city networks of Portland and Denver and compare the results obtained by our optimization approach with results found by other means. For these comparisons, we use the microsimulation tool VISSIM.

How to read this thesis

The thesis is chronological in structure. However, the chapters can be followed independently, too. Chapter 2 and Chapter 3 come with their own introduction and consider related, but individually presented, problems. Furthermore, these two chapters do not explicitly require the reading of the Chapters 1 and 4. On the other hand, the Chapters 1 and 4 are strongly related and we recommend that Chapter 1 is read prior to Chapter 4.

Moreover, we give a chapter outline at the beginning of each chapter. Also, conclusions are drawn and open questions are raised at the end of each chapter.

A further remark

We assume the reader of this thesis to be familiar with the basic concepts in linear and integer programming, graph theory, and complexity theory. For additional information on linear and integer programming we refer to [Sch86, NW88]. Good textbooks on graph theory are for example [Wes96] and [Die00]. Moreover, concepts in complexity theory that are necessary to follow this thesis are covered by [GJ79] and [HO02].

As for the parts that deal with traffic engineering concepts or with traffic signal terms in particular, we refer to Section 1.1.1 for short textual explanations of the most important terms. Herewith, following most parts of Chapter 1 and Chapter 4 should be unproblematic. Of course, while developing our mixed-linear integer program in Section 1.3, we give formal definitions of all relevant terms, too. However, additional information can be found in "Richtlinien für Lichtsignalanlagen RiLSA, Lichtzeichenanlagen für den Straßenverkehr" [ril92] and in the "Highway Capacity Manual" [hcm00].

1

THE NETWORK SIGNAL COORDINATION PROBLEM

In this chapter, we consider the Network Signal Coordination (NSC) problem. The NSC problem was introduced in 1975 by Gartner et al. [GLG75a], though many similar optimization tasks were known and have been defined already a lot earlier. After a brief summary of the most important terms concerning traffic signals in Section 1.1.1, we give a short overview of the most significant optimization problems on traffic signals in Sec. 1.1.2, also to be able to clearly define the NSC problem and to restrict it from other problems. Then, in Sec. 1.2.1 we formally define the NSC problem and illuminate similarities to the Periodic Event Scheduling Problem (PESP) in Section 1.2.2. We report on the complexity of the NSC problem in Section 1.2.3. Thereafter, in Sec. 1.3 we develop in detail a revised mixed-integer linear programming (MIP) formulation for the NSC problem. Finally, we explain a possible application of our MIP in practice, see Section 1.4. Parts of this chapter were published in [MNW06].

1.1 Introduction to Coordination of Signals in Networks

Before we define the Network Signal Coordination Problem, we introduce the most important terms related to traffic signals and give an overview of what kinds of optimization tasks concerning traffic signals have been considered so far. Other surveys on the topic are provided for example by [SS95, tft] or contained in [Läm07].

1.1.1 The Language of Traffic Signals

There is no unique language in the field of traffic engineering in general, and neither in topics related to traffic signals. Rather, the terms and notation depend on the respective country and language. However, since following this thesis requires only basic knowledge of traffic terms, in this section we give only a short textual description of the most important terms. Notice that we do not give formal definitions here. For those, we refer to section 1.3. Nevertheless, whenever it is possible, i.e., when we do not work with a term, but only want to give an intuition, we omit a formal definition at all. For complete information see [hcm00] and [ril92].

In traffic engineering one considers traffic, i.e., vehicles, moving through a single and isolated **junction**, an **arterial**, which is a, possibly bi-directionally traversable, series of junctions, or through a whole **network**, i.e., an arbitrary set of junctions. Then, one distinguishes different types of **signals** at the junctions. Here, the term signal refers to *all* signaling devices at a junction. Roughly speaking, the following two types are the most important ones. On the one hand, there are **traffic responsive signals**. At these signals, the signal settings react on the present traffic. On the other hand, **fixed-time (controlled) signals** do not react on the actual traffic. Here, after a prescribed time span, called **cycle length**, the pattern of red phase and green phase repeats. The particular division of a cycle length into a red and a green is referred to as **(red green) split**. At a fixed-time signal, there are usually different **signal groups** that control the traffic for particular directions. The green phases of different signal groups are shifted against each other, since they usually control competing traffic streams. In addition, the order of the signal groups at a signal is called **phase sequencing**. See Fig. 1.3 on page 19 for an example of a **signal timing plan** in which the relevant data for one fixed-time traffic signal is merged.

A very important term is the one of an offset. The **(inter node) offset** determines how different signals are operated or shifted relatively to each other. That means the following: at each signal there is a marked out reference point, which sometimes is the begin of the green phase of the first signal group. Then, the offset denotes the time span between reference points of two signals at two consecutive junctions. In this case, there is an offset for each pair of consecutive signals. However, the offset can also be defined for one single signal. Then, it determines the time span between this signal's reference point and a given network-wide zero reference point. See Figure 1.4 for an illustration of both types of offsets. A sketch of the **intra-node offset**, which determines the shifting of different signal groups at one signal, is depicted in Figure 1.5 for example.

Of course, when considering signalized junctions, arterials or networks, several optimization tasks come to mind. Generally, one is interested in optimizing signal settings in order to achieve a certain goal. Such signal settings are the red green split, the phase sequencing, the cycle length, and the offset. As for the goals to achieve, for example, minimizing the **delay** or maximizing the **bandwidth** have to be mentioned. Here, the term delay refers to the delay that is due to the signalization, i.e., delay that occurs when vehicles have to wait because of a red. On the other hand, maximizing bandwidth means that the signals along an arterial or within a network are adjusted, such that a preferably wide possible corridor through the green phases of consecutive signals exists, within which the vehicles do not have to stop at the signals at all. Such a corridor is sometimes called a **greenband**. See Figure 1.1 for a visualization of greenbands.

Whenever the offsets are included in the signal settings to be optimized, we say that we optimize the **coordination**. In the literature the term **synchronization** is sometimes used synonymously. However, we prefer the term coordination and leave the item synchronization to cases where the offsets *and* the cycle length are optimized.

When considering traffic flow one distinguishes between a **microscopic** view and a **macroscopic** view. The model is said to be microscopic if each individual vehicle is considered. On the contrary, we speak of a macroscopic model or approach, if the vehicles are aggregated in some sense. For example, it is popular to consider **platoons** of vehicles, that is, groups of consecutive vehicles close together that are treated as one quantity. However, it has to be mentioned that there are traffic models for which a classification into one of the two views is not obvious.

1.1.2 Optimizing Traffic Lights

Since the introduction of automatic traffic signals in the 1920s, much work and research has been done on modeling, analyzing, and later also on simulating and optimizing traffic signals. In this section we mention the most important modeling and optimization approaches. Notice, however, that we do not claim to provide a complete overview.

When talking about optimization in the context of traffic signals, one faces many different optimization tasks. Table 1.1 gives a glimpse of possible differentiations between them.

Table 1.1: The table provides criteria to distinguish between mathematical approaches for problems dealing with traffic signals. Observe, however, that not all combinations are reasonable.

Criteria	Possibilities
type of approach	optimization, heuristic (genetic algorithms, local search etc.)
variables	offset, red-green split, cycle length, phase sequencing, travel speed, routes, almost any combination thereof
objective	minimizing delay, number of stops, fuel consumption; maximizing greenband; combinations thereof
type of signal	fixed-time signals, traffic responsive signals, both
type of approach	theoretical, practical
application on	single junctions, arterials, networks
preconditions on traffic	public only, individual only, none
preconditions on demand	high demand only, low demand only, none
preconditions on signalization	common cycle length, none
modeling perspective	macroscopic, microscopic

One of the first important scientific publications on traffic signals was by Webster [Web58] in 1958. In this pioneering work, he prepared the ground for analyzing single traffic signals, e.g., by providing delay-estimating formulae that are, in a slightly changed form, still in use today. Using this formulae, Webster also researched on minimizing the delay by adjusting optimal green proportions at a signal.