Maritime Transportation: Between AVs and tankers, there is (not) the middle of the sea

Author(s):
Glauser, Yannik M.

Publication Date:
2021-07-13

Permanent Link:
https://doi.org/10.3929/ethz-b-000501042
Yannik Glauser

Maritime Transportation

Between AVs and tankers, there is (not) the middle of the sea

Master’s Thesis

Institute for Dynamic Systems and Control
Swiss Federal Institute of Technology Zurich

Supervision

Gioele Zardini
Dr. Andrea Censi
Prof. Dr. Emilio Frazzoli

July 2021

IDSC-proj-mt-glauser
Abstract

The maritime transportation of crude oil by tankers represents a major aspect of global trade, as crude oil is refined to all sorts of different fuel products. The crude oil shipping market is a decentralized system based on traditional concepts with multiple stakeholders involved (e.g. ship owners, charterers, ship brokers, etc.). Most recently (March 2021), the blockage of the Suez Canal by a vessel showed that this system can be strongly affected by external factors. Literature on crude oil shipping either addresses approaches for very specific problems, therefore not applicable to other scenarios, or more general aspects, mainly by focusing on analytical methods. Alternatively, we present in this Master’s thesis a general simulation framework in order to analyze, model and simulate the crude oil shipping market in its entirety. For this work, we strongly collaborated with the commercial ship management group, The Signal Group, who provided us access to their valuable knowledge, software and data. The simulation framework describes efficiency in crude oil shipping as combination of incomparable metrics addressing resources in time, emission and cost. Based on that, a concept for preferences is defined where these metrics are considered simultaneously, accounting for mixed strategies during simulation. The mathematical model and implementation of the framework cover multiple different aspects, e.g. graphs representing the transportation network, demands and vessels, etc. By running simulations over time, dynamic aspects of the system could be analyzed, such as congestion at canals. This dynamic interaction and dependence of the vessels on each other enables a new perspective for the usually static scheduling and ship routing problems. The results of the simulations based on different preferences provide first insights on the relations between the chosen metrics for efficiency. The implemented simulation framework provides a solid foundation for further refinements and extensions in the future, especially due to its modularity. Ultimately, this should lead to a better understanding of efficiency in crude oil shipping and its optimization.

Keywords: Maritime Transportation, Crude Oil Shipping, Simulation Framework, Network Design, Canal Congestion.
Acknowledgment

First of all, I would like to thank Prof. Dr. Emilio Frazzoli very much for giving me the opportunity to elaborate my Master’s Thesis in his research group at the Institute for Dynamic Systems and Control (IDSC). I especially appreciated his valuable inputs during the project.

Additionally, many thanks go to Dr. Andrea Censi for his scientific supervision.

I would also like to express my very high gratitude to Gioele Zardini for his kind and professional support as tutor during this project. Especially in this extraordinary time of the COVID-19 pandemic, I am very thankful for his patience and input during many hours of online meetings.

Special thanks go also to Marc Albert for his valuable help with the program setup in C++ and also for his appreciated inputs to code related questions in general.

Furthermore, I would like to express my appreciation towards Ioannis Martinos for enabling the promising collaboration with Signal. Many thanks also for his kind feedback and the valuable discussions.

Many thanks are also addressed to Dimitrios Tasoulis for all the inputs during our regular meetings and for the great support throughout the whole project.

Last but not least, thanks a lot to the whole Signal team for their continuous help in providing data and access to various platforms.

This project has been edited by using the \LaTeX{} template of Ritter et al. [1].
# Contents

## Nomenclature

1. Introduction
   1.1 Motivation ........................................ 1
   1.2 Objective ......................................... 2
   1.3 Related Works .................................... 3
   1.4 Statement of Contribution ........................ 6
   1.5 Structure of this Document ....................... 7

2. Approach to Problem .................................. 9

3. Theoretical Part
   3.1 Efficiency .......................................... 11
   3.2 Transportation Network ............................ 13
      3.2.1 Graphs ....................................... 13
      3.2.2 Nodes and Arcs ............................... 17
      3.2.3 Canals and Congestion ....................... 18
   3.3 Simulation Framework ............................... 20
      3.3.1 Demands and Vessels ......................... 20
      3.3.2 Simulation Procedure ......................... 22
   3.4 Preferences and Objectives ....................... 26
      3.4.1 Preferences .................................. 26
      3.4.2 Simulation Objective ......................... 27

4. Methodology .......................................... 29
   4.1 General Details on Implementation ............... 29
   4.2 Implementation of Graphs ......................... 30
      4.2.1 Voyage Data provided by Signal .............. 30
      4.2.2 Applied Nodes and Arcs ..................... 31
      4.2.3 Graph Implementation and Generation ....... 32
      4.2.4 Calculation of Arc Weights .................. 33
      4.2.5 Historical Graphs based on Voyage Data ... 34
   4.3 Implementation of the Simulation Framework .... 36
      4.3.1 General Code Structure ....................... 37
      4.3.2 Initialization ................................ 37
      4.3.3 Processing Demands (Main Part) ............. 39
      4.3.4 Results and Statistics ....................... 42
      4.3.5 Congestion Estimation ....................... 43
   4.4 Performed Simulations ............................. 44

5. Results and Discussion .............................. 47
   5.1 General Observations and Results ............... 47
## Nomenclature

### Mathematical Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>Preference parameter, addressing the metric <em>time</em></td>
</tr>
<tr>
<td>$\beta$</td>
<td>Preference parameter, addressing the metric <em>emission</em></td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Preference parameter, addressing the metric <em>cost</em></td>
</tr>
<tr>
<td>$A$</td>
<td>Set of all arcs $a$</td>
</tr>
<tr>
<td>$\mathcal{F}_R$</td>
<td>Resource space, referring to arc weights</td>
</tr>
<tr>
<td>$G$</td>
<td>Digraph (also referred as directed graph)</td>
</tr>
<tr>
<td>$P$</td>
<td>Set of all ports $p$</td>
</tr>
<tr>
<td>$\mathcal{V}$</td>
<td>Set of all nodes $v$</td>
</tr>
<tr>
<td>$\mathcal{X}$</td>
<td>Set of all different preferences $x$</td>
</tr>
<tr>
<td>$a$</td>
<td>Arc (also referred as edge)</td>
</tr>
<tr>
<td>$c$</td>
<td>Cost $\text{USD} / t_{\text{cargo}}$</td>
</tr>
<tr>
<td>$d$</td>
<td>Additional delay at canal</td>
</tr>
<tr>
<td>$d_{\text{regular}}$</td>
<td>Delay at canal during regular operation</td>
</tr>
<tr>
<td>$e$</td>
<td>SO2-Emission</td>
</tr>
<tr>
<td>$f_{\text{canal}}$</td>
<td>Canal congestion function</td>
</tr>
<tr>
<td>$n$</td>
<td>Number of vessels at/in canal or general index</td>
</tr>
<tr>
<td>$n_{\text{crit}}$</td>
<td>Critical number of vessels at canal</td>
</tr>
<tr>
<td>$p$</td>
<td>Port</td>
</tr>
<tr>
<td>$q$</td>
<td>Waiting time per additional vessel at canal</td>
</tr>
<tr>
<td>$r$</td>
<td>Path (also referred as route)</td>
</tr>
<tr>
<td>$t$</td>
<td>Time $s$</td>
</tr>
<tr>
<td>$T_{\text{end, demand}}$</td>
<td>End of period during which new demands are released</td>
</tr>
<tr>
<td>$T_{\text{end}}$</td>
<td>End time of the simulation</td>
</tr>
<tr>
<td>$T_{\text{start, demand}}$</td>
<td>Start of period during which new demands are released</td>
</tr>
<tr>
<td>$T_{\text{start}}$</td>
<td>Start time of the simulation</td>
</tr>
</tbody>
</table>
\(v\) Node (also referred as vertex)
\(w\) General weight for arcs \(a\) or paths \(r\)
\(x\) Preference

**Subscripts**
- Aframax Aframax, referring to the Aframax subgraph
- Arrival Arrival, referring to Arrival nodes
- ballast ballast, referring to ballast arcs
- canal canal, referring to the canal congestion function
- Departure Departure, referring to Departure nodes
- destination destination, referring to the demand destination
- Discharge Discharge, referring to Discharge arcs
- end end, referring to the end node of an arc or a time period
- laden laden, referring to laden arcs
- Land Land, referring to Land nodes
- Load Load, referring to Load arcs
- origin origin, referring to the demand origin
- Panama Panama, referring to the Panama Canal
- start start, referring to the start node of an arc or a time period
- Suezmax Suezmax, referring to the Suezmax subgraph
- Suez Suez, referring to the Suez Canal
- total total, referring to the total transportation graph
- Travel Travel, referring to Travel arcs
- VLCC VLCC, referring to the VLCC subgraph

**Superscripts**
- dimensionless, referring to normalized metrics \(\text{time, emission, cost}\)

**Acronyms and Abbreviations**
- (S)ECA (Sulfur) Emission Control Area
- AMoD Autonomous Mobility-on-Demand
- API Application Programming Interface
- AV Autonomous Vehicle
- BGL Boost Graph Library
BPR | Bureau of Public Roads
---|---
DWT | Deadweight Tonnage
IDE | Integrated Development Environment
IDSC | Institute for Dynamic Systems and Control
IMO | International Maritime Organization
JSON | JavaScript Object Notation
LSNDP | Liner Shipping Network Design Problem
OR | Operations Research
TCE | Time Charter Equivalent
USD | US Dollar
VLCC | Very Large Crude Carrier

**Glossary**

(S)ECA | Emission Control Areas (ECAs) and Sulfur Emission Control Areas (SECAs) describe sea areas, usually located in coastal regions, where airborne emissions caused by ships are more strictly controlled [2].

Aframax | Aframax denotes a vessel class whose ships exhibit a capacity between 82,000 to 125,000 dwt (definition by Signal).

AMoD | On-demand mobility provided by fleets of self-driving vehicles (AVs) [3].

Ballast | In case a vessel travels in ballast, it travels without cargo (empty). The description comes from the ballast tanks of the vessel which are filled with sea water in order to provide better stability while traveling without cargo (definition by Signal).

Bunkering | Bunkering is used interchangeably with refueling (definition by Signal).

Cargo | Cargo (interchangeably used for shipment) represents a set of goods shipped together from an origin to a destination. Here specifically referred to crude oil as cargo [4].

Charterer | Charterers, usually the cargo owners, represent the customers in crude oil shipping, as they require (charter) tankers in order to move oil between different locations [5].

Crude oil | Crude oil (also known as Petroleum) is a naturally occurring, yellowish-black liquid which is commonly refined into various types of fuels [6].

DWT | Deadweight tonnage (DWT) denotes the weight carrying capacity of a ship in metric tons (often abbreviated as dwt), excluding the vessel’s empty weight [7].

Industrial | Industrial denotes a mode of operation in commercial shipping where both cargo and vessels are controlled or owned by the same company [8].

Laycan | Laycan denotes the time window within which the vessel owner has to present the ship to the cargo owner (charterer) at the agreed port (definition by Signal).

Liner | Liner denotes a mode of operation in commercial shipping based on a fixed schedule for the vessels with predetermined travel paths, usually referring to container ships (or general other cargo vessels) [4].

Ship broker | Ship brokers are consulted by charterers in order to contact vessel owners, evaluate potential tankers and negotiate the price and conditions for shipping, etc. [5].
Ship owner  Ship owners own the vessels and provide them for the shipping of cargo [5].

Suezmax  Suezmax denotes a vessel class whose ships exhibit a capacity between 125'000 to 200'000 dwt (definition by Signal).

TCE  Time Charter Equivalent (TCE) represents the charter rate for vessels in terms of "dollars per day" [5].

Tramp  Tramp denotes a mode of operation in commercial shipping based on servicing individual demands, where cargo owners (charterers) specifically charter vessels from vessel owners in order to ship their cargo. It is commonly applied for tankers and dry bulk carriers [4].

Vessel  Vessel is used interchangeably for ship [4].

VLCC  Very Large Crude Carriers (VLCCs) represent a vessel class, exhibiting a capacity between 200’000 to 350’000 dwt (definition by Signal). They are the largest vessels covered in this project.

Voyage  Voyage represents a trip of a ship, starting and ending with an empty vessel. It usually comprises multiple paths and different events (Load, Discharge and Stop) for servicing a demand (definition by Signal).
Chapter 1

Introduction

This chapter serves as an introduction to this Master’s thesis on maritime transportation. It first presents the initial motivation for this project. Based on that, the objective of this work is defined. After discussing related works in form of a literature review, we evaluate the contribution of the project within this context. At the end of this chapter, the structure of this document is given.

1.1 Motivation

The supply of energy poses a major challenge in global economy. An important aspect in this broad field is represented by the maritime transport of crude oil with over 1.9 billion tons in 2019 [8]. Crude oil occurs naturally and serves as precursor for all sorts of refined fuel products. During extraction, refining and burning of this fossil fuel vast amounts of greenhouse gases are released, which strongly contribute to climate change [6, 9].

Generally, crude oil is shipped from the production sites to the refineries by crude oil tankers. Therefore, tankers play a significant role within the energy value chain. While refined oil products are usually carried by smaller “clean” or “product” tankers, crude oil is predominantly transported by larger, so-called “dirty” tankers. Only the latter ones are considered in this work [5].

Three main classes of crude oil tankers are distinguished based on their size and capacity: Very Large Crude Carriers (VLCCs), Suezmax and Aframax. They exhibit capacities of up to 2 million, 1 million and 600’000 barrels of crude oil per vessel, respectively (find the proper specifications for the different vessel classes in the glossary). The global fleet of crude oil tankers comprises about 700 VLCC, 500 Suezmax and 900 Aframax vessels [5].

A new tanker needs at least two years to be built (between purchase order and delivery) and has a life span of about 20 years. The price for a VLCC vessel lies between 80 and 160 million USD. The purchasing price of a tanker results from the general supply and demand of tankers, similarly to the oil price which represents an indicator for supply and demand of crude oil [5, 10].

Both, the tanker as well as the crude oil markets strongly interact with the shipping market for crude oil. As this project fully focuses on the crude oil shipping market, dynamic aspects of the other two markets, such as fleet management or oil price fluctuations, are not specifically considered here. Further related discussions are found in [11, 12, 13, 14, 15].

The crude oil shipping market is a decentralized transportation system with multiple different stakeholders. The customers, commonly denoted as “charterers” of the vessels, require tankers in order to move oil between different locations. These include National and International Oil Com-
panies (e.g. Saudi Aramco, Total, Shell and Chevron) as well as trading companies (e.g. Trafigura and Glencore) and large refineries. On the other side, vessel owners provide the tankers for crude oil shipping. In many cases, charterers additionally consult ship brokers who contact vessel owners, evaluate potential tankers and negotiate the price and conditions for shipping, etc. [5].

The crude oil shipping market is strongly based on experience and traditional concepts. Due to the longer transport times in crude oil shipping (weeks to months) compared to public transportation as an example (minutes to hours), planning and scheduling multiple voyages ahead for different demands is quite challenging.

Moreover, very recently (March 2021) we have seen that the maritime transportation system is strongly affected by external factors, e.g. when a vessel blocked the Suez Canal for several days [16] [17]. Similarly, the current COVID-19 pandemic already showed some of its consequences, e.g. in lower freight rates for oil tankers due to a lower economic activity caused by multiple lockdowns [18].

Consequently, to better predict and anticipate such challenging scenarios, the maritime transportation system for crude oil has to be further studied and analyzed. This Master’s thesis tries to contribute to a better understanding and regulation of global crude oil shipping.

1.2 Objective

This project focuses on analyzing, modeling and simulating the crude oil shipping market to optimize it. Therefore, the main goal of this thesis is to build a general simulation framework for crude oil shipping which is based on real data.

For this task, we were supported by the Greek ship management group, The Signal Group, here commonly referred as Signal. The entire Signal Group covers multiple business segments, referring to Signal Maritime, Signal Ocean and Signal Ventures. Signal Maritime is a commercial ship management company with an own fleet of tankers, predominantly consisting of Aframax vessels. Signal Ocean provides a platform, accessible by web or mobile application, which facilitates the planning, organization and decision making aspects in commercial shipping. Signal Ventures represents the investment branch of the group, leveraging and supporting startups and entrepreneurs in the field of maritime shipping.

As part of the collaboration with Signal for this thesis, they provided us access to their voyage data, the Signal Ocean platform as well as their API platform. Additionally, we could benefit from their valuable knowledge to better understand and address the real challenges in crude oil shipping.

Based on this collaboration with Signal the following objectives were set for this thesis:

1. Find adequate metrics to define efficiency in maritime transportation.
2. Create a theoretical model for crude oil shipping in form of a general framework based on the chosen efficiency metrics. The model should consider multiple preferences and dynamic aspects of the system, while being consistent with Signal’s voyage data.
3. Implement the simulation framework for the theoretical model, applying the voyage data provided by Signal.
4. Apply the simulation framework by running simulations to get insights on the relations between the chosen efficiency metrics.

In order to provide a broad context in the field of maritime transportation, previous works are discussed in a literature review in the following section.
1.3 Related Works

Today, there exists a large variety of different literature about maritime transportation. Retrospectively, this was not always the case. In the early days of Operations Research (OR) related works in maritime shipping, dating back to the 1950s, there was a lack of attention dedicated to this field, leading to a scarcity of corresponding literature. According to [4, 19], this was caused by multiple reasons:

- **Low Visibility**
  People usually do not get into contact with large container ships or tankers in their daily life compared to other transportation modes, e.g. trucks, trains or aircrafts.

- **Problem Structure**
  Due to the large variety of problems related to maritime transportation in many different operational environments, they are very difficult to structure and organize.

- **High Uncertainty**
  The large time scales for voyages (weeks to months), in combination with the influences of weather, mechanical problems and strikes, lead to a high uncertainty in planning and scheduling.

- **Long Tradition**
  Maritime transportation by ships is one of the oldest forms of transportation. Therefore, the system seems to be conservative, strongly based on experience and traditional concepts. Moreover, the maritime shipping industry is fragmented and decentralized, i.e., there are many different participants involved (e.g. small family owned or large industrial companies).

However, a positive trend in research regarding maritime transportation is observed, as the number of related works almost doubled every decade. The corresponding literature published over the past decades was thoroughly summarized and updated each decade by [19, 20, 21, 22]. Other more specific taxonomy and survey papers (as secondary literature) focus on weather routing [23], containership routing [24], vessel speed models [25], decarbonization of shipping [26] or environmental sustainability [27].

A general introduction and overview of maritime transportation is given by different resources, see [4, 5, 8, 28]. Based on [4], planning problems in maritime transportation are classified into strategic, tactical and operational problems according to their planning horizon (different time scales).

- **Strategic Problems (long-term)**
  These problems refer to a long-term planning horizon, including tasks such as market and trade selection, ship design, network design, fleet management (fleet size and mix) as well as port design, etc. They address very fundamental aspects of maritime transportation.

- **Tactical Problems (medium-term)**
  Tactical problems have a shorter planning horizon than strategic ones. Among many other tasks, they include ship routing and scheduling, ship management and fleet adjustments. Most works in literature focus on such problems.

- **Operational Problems (short-term)**
  They represent short-term problems in maritime shipping, e.g. speed selection, ship loading, environmental routing (weather), etc. They usually tackle the question how specific processes or smaller tasks within the system should be performed.

Moreover, three basic modes of operation are distinguished in commercial shipping: liner, tramp (charter) and industrial operation. Liner operation is based on a fixed schedule for the vessels with predetermined travel paths. This operation mode usually refers to container ships (or general other
In contrast to that, tramp operation is based on servicing individual demands. In this case, cargo owners (charterers) specifically charter vessels from vessel owners in order to ship their cargo. This operation mode is commonly applied for tankers and dry bulk carriers, therefore particularly relevant for the crude oil shipping market. In addition to that, the industrial operation represents the mode where both cargo and vessels are controlled or owned by the same company. Maritime transportation of crude oil can also be represented by this mode [4].

Generally, most of the papers in maritime transportation consider a ship routing or scheduling problem for a specific operation mode. Since specific papers about the tramp (charter) or industrial operation in crude oil shipping are quite rare, also literature about container ships in liner operation is discussed below.

Previous works in the field of maritime transportation are often structured similarly, especially when addressing a ship routing or scheduling problem. First, the maritime transportation system is modeled in a formal way. Second, the model is solved/optimized analytically or numerically by using specific solution methods and algorithms. Finally, the system is tested and the simulated data analyzed.

The existing models in literature are difficult to summarize due to their extensive and complex mathematical definitions and constraints. There is no general nomenclature for the many different parameters and variables. However, the general approach of these models is usually structured in a similar way:

- **Setup and Definition**
  In this part of the mathematical model, the system is framed, the general idea is explained, the variables are defined and the different aspects of the model are put together. Despite the different nomenclatures, similar concepts can be observed in various papers, e.g. definition of a graph $G(V,A)$ based on nodes $V$ and arcs $A$, etc.

- **Cost/Objective Function**
  This part usually consists of a specific function which has to be optimized (e.g. min for cost, max for profit). Depending on the type of model, these expressions are stated differently. Among others, the most common models are so-called 'arc flow' and 'path flow' models. The main difference lies in the optimization formulation. Path flow models assign complete paths (also referred as routes) to different vessels. Arc flow models select single arcs (pairs of ports or cargo) for different vessels. The two different models can often be converted into each other, because the basic concept of the transportation system represented by nodes and arcs usually remains the same for both. As discussed later in Section 3.3, a so-called “particle flow” model is applied for this project, where each vessel is considered as individual particle within a transportation graph.

- **Constraints**
  The cost function (discussed above) is subject to a variety of different constraints and restrictions, which can consider time, capacity, flow conservation, etc. These constraints contain important details, leading to potential differences in the model solution. Usually, each model applies its different constraints.

It should be mentioned that the model structure depicted above is greatly simplified and therefore just provides a basic idea when commenting some specific examples below.

Liner shipping network design problems (LSNDPs) for container ships are addressed in [29, 30, 31, 32, 33]. All these papers strongly focus on the optimization process of a specific objective function, predominantly in path flow formulation. A detailed problem description of LSNDPs is provided in [32], including a mathematical model as well as a benchmark suite for LSNDPs. [31] presents an optimization algorithm for a LSNDP (so-called "matheuristic"), which contains a detailed model...
for the transportation network, represented by graphs (shows similarities to the graphs developed in this project, see Section 3.2). A similar model is applied in [29], addressing a time constrained LSNDP. In [30] a multi-commodity network flow problem with time constraints is covered, referring to its application for liner shipping. Deadlines for a LSNDP are applied in [33].

In [34] a model is presented for scheduling and routing vessels in the North Sea for a major Norwegian oil and gas company. The operation mode resembles the liner operation, following a fixed weekly schedule. Another work which describes liner shipping along the Norwegian coast is [35]. Other studies regarding the planning of fleet size and mix in liner operation are [36, 37]. Specific models for the routing and scheduling of container ships are discussed in [38, 39].

For tramp (charter) and industrial operation, many models apply so-called 'set partitioning' formulations. Such problem descriptions mainly focus on an optimization process by which specific sets are partitioned into smaller subsets. For ship routing and scheduling, these models usually apply algorithms that partition the set of all feasible schedules for vessels in order to optimize a specific objective function. In these cases, the system modeling is often less important than the analytical solution/optimization process, as seen in [40-49].

Regarding crude oil tankers, a scheduling problem for a major oil company is addressed in form of a set partitioning model in [40]. A similar concept is applied in [41] for bulk cargo, considering the fleet of the Military Sealift Command of the US Navy. Other papers which discuss scheduling problems for bulk cargo represent [42, 43, 44]. In [45] another scheduling problem for tankers is investigated, specifically related to the Chevron Shipping Company. An oil tanker scheduling problem for a Japanese oil transportation company is given in [46]. Similarly, [47, 48] consider the routing and scheduling problem for oil tankers with regard to the Kuwait Petroleum Corporation. A ship routing and scheduling problem for tramp (charter) operation and flexible cargo sizes is described by [49].

A ship routing model which considers also the vessel speed based on time, cost and environmental objectives is developed in [50]. The consideration of multiple preferences is an important concept within this project as well, as introduced in Section 3.4. Another work about maritime routing that specifically addresses the vessel speed is [51]. Additionally, this paper also focuses on the environmental aspects of shipping by considering so-called Emission Control Areas (ECAs) and Sulfur Emission Control Areas (SECAs). These denote sea areas, usually located in coastal regions, where airborne emissions caused by ships are more strictly controlled [2].

Generally, the literature in maritime shipping of the past years is very broad, as seen by all examples mentioned above. The different works range from general theoretical models, mainly focusing on the analytical tools of the solution approach, to very specific applications of real problems. However, it seems that the major part of the papers focuses more on theory than on having an actual practical value. In these cases, the solution approach is more relevant than the actual problem definition and the system modeling. This discrepancy between literature and reality is specifically addressed by [52].

Especially the literature on crude oil shipping is rather scarce and does not appear to be really balanced. Corresponding papers are either very specific, considering concrete examples or data, and therefore not applicable to other scenarios, or they cover very general theoretical aspects without solving actual real world problems.

Moreover, most of the works only discuss static aspects of maritime transportation, e.g. by describing specific ship routing or scheduling problems. The focus is often set on methodological tools of a specific solution approach, for which the crude oil shipping market serves as an interesting application field. In many of these works, the system modeling aspect falls behind. In addition to that, transport efficiency in the crude oil shipping market seems not to be studied in a general
way. Many papers make statements on efficiency based on their specific objective function only, which does not cover all aspects of efficiency in regard of potential preferences different stakeholder may have.

Taking these insights into consideration, a different approach for this project was applied, strongly basing on co-design frameworks for future mobility systems, as seen in Chapter 2.

The theory of co-design, based on category theory, describes general design problems as functionalities, implementations and resources that are related to each other by so-called 'feasibility relations'. Based on that, problems of type 'find minimal resources required to implement given functionalities' can be structured and solved [53].

These general concepts were applied for the design of autonomous vehicles (AVs) as well as the design of AV-enabled mobility systems. Both design problems are strongly coupled. In [3, 54] an intermodal framework is presented that jointly connects AVs with public transportation. AVs are assumed to provide a so-called Autonomous Mobility-on-Demand (AMoD) service. This framework was extended in [55], by including micromobility as well (e.g. shared bikes and e-scooters). It describes incomparable resources (time, CO₂-emission and cost) that are required to provide a specific functionality. The framework was applied in a real case study for Washington D.C., USA. Furthermore, a great review on AmoD in general is given in [56].

The fact that the framework for simulations in maritime shipping is similarly approached as the one for AV-enabled mobility systems led to the title of this thesis. Despite of all the differences between tankers and AVs, developing frameworks for their transportation systems represent similar problems that can be tackled analogously.

Based on the given context regarding literature in the field of maritime transportation as well as on mobility in general, the statement of contribution of this thesis is formulated below.

1.4 Statement of Contribution

This Master’s thesis presents a general simulation framework for analyzing, modeling and simulating the crude oil shipping market. The framework is strongly based on the voyage data of Signal, addressing real challenges in crude oil shipping.

With reference to the defined objectives for this project (see Section 1.2), the details on the contributions for each objective are given here.

1. At first, different incomparable metrics are introduced in order to define the efficiency of maritime transportation in a general way. These metrics address different aspects of efficiency, covering resources of time, emission and cost. Inspiration for the selection of these metrics has been taken from works on AV-enabled mobility systems [3, 54, 55].

2. Based on the efficiency definition, we developed theoretical concepts for preferences and simulation objectives, in which multiple efficiency metrics are considered simultaneously. This allows us to account for mixed strategies during simulation which has not been specifically addressed in literature so far. Moreover, we provide the theoretical model for all other aspects of the framework, e.g. graphs representing the transportation network, demands and vessels, simulation procedures, etc. In order to achieve this, the voyage data provided by Signal was evaluated and applied.

3. Based on the theoretical model, the entire simulation framework was implemented in C++. The implemented code of the framework represents the main work of this project.
As application of the framework, we run different simulations over time. With these simulations, dynamic aspects of the system could be analyzed, e.g. congestion at canals. To account for dynamic congestion effects at canals, a congestion estimator was implemented and tested. The dynamic interaction and dependence of the vessels on each other enable a completely new perspective on ship routing and scheduling which usually represent static problems only. Furthermore, we give first insights on the relations between the chosen efficiency metrics based on the results of simulations for different preferences. With the implemented simulation framework arbitrary scenarios in crude oil shipping can be simulated and their potential consequences predicted.

In its current iteration, the framework provides a solid foundation for further refinements and extensions in the future, mainly due to its very general and modular structure. Hence, this work tries to contribute to a better understanding and regulation of the crude oil shipping market.

1.5 Structure of this Document

In Chapter 2 the general approach for analyzing, modeling and simulating the crude oil shipping market is discussed in more detail. The chapter introduces the different aspects of the chosen approach as a guidance to better understand the structure of the project.

The theoretical aspects of this project, including the model for the whole simulation framework, are given in Chapter 3 e.g. efficiency, graphs, demands, vessels, general procedures, preferences, etc. This model builds the foundation for all following chapters.

The methodology in Chapter 4 describes how the theoretical model for the simulation framework was applied and implemented. This also includes the consideration of voyage data provided by Signal for this project. Additionally, further details on running simulations are given.

Results obtained by running different simulations are summarized in Chapter 5. Multiple plots and figures are shown to visualize and facilitate the discussion on observations and findings.

Conclusions are given in Chapter 6. Additional suggestions for further investigations are discussed as outlook in Chapter 7.

The separate code description (README file) and the code itself are given in [57].
Chapter 2

Approach to Problem

This chapter gives an overview of the different aspects that have been covered in order to tackle the problem of analyzing, modeling and simulating the crude oil shipping market. It describes the overall concept of this work, serving as a general guidance.

The main objective of this project is to analyze, model and simulate the crude oil shipping market in general, based on the voyage data provided by Signal, as mentioned in Section 1.2. This way, we intended to find ways to ideally optimize the current market. Due to the extensive task to represent the complete maritime transportation system for crude oil in one general framework, the project comprises many different aspects that interact with each other. Figure 2.1 depicts these aspects schematically.

At first, metrics are characterized based on which we can determine the efficiency of a system. Furthermore, the transportation network is considered as foundation for the simulation framework. It is modeled by graphs, as will be shown in Section 3.2. The framework also includes demands and vessels, which have to be further specified.

All these aspects had to be considered for defining and implementing the actual procedures of the transportation system. These procedures represent the main part of the simulation framework as they describe how demands and vessels are processed during simulation.

In a next step, simulation objectives can be expressed based on different preferences. After running different simulations, their results are analyzed in order to find specific trends and other interesting phenomena.

Figure 2.1: Individual aspects of the approach for a simulation framework in crude oil shipping.
It is very important to mention that all these different tasks were strongly based on the voyage data provided by Signal. Therefore, theory, implementation and application of the simulation framework focuses on real challenges of crude oil shipping based on actual data.

Following the discussion in section Section 1.3 it can be seen that the general approach of this simulation framework for crude oil shipping is very similar to the one for AV-enabled mobility systems [3, 54, 55]. However, the individual aspects of the general approach can differ a lot from each other, as both systems consider completely different modes of transportation (large tankers compared to AVs, trains, bikes, etc.).

During the subsequent chapters, all mentioned aspects are discussed first from a theoretical point of view, followed by the description of their implementation and application.
Chapter 3

Theoretical Part

This chapter discusses the theoretical aspects of the simulation framework as model for the crude oil shipping market. At first, metrics for efficiency are characterized. Subsequently, the transportation network as foundation of the simulation framework is modeled in form of graphs. After elaborating the details about demands and vessels, the actual simulation procedures are described which represent the main part of the simulation framework. At the end of this chapter, the concept of preferences is introduced, including the objectives for different simulations.

3.1 Efficiency

Efficiency can have different interpretations depending on the context of application within the crude oil shipping market. In order to analyze and compare different simulations, specific metrics are used to express transport efficiency.

The current practice in the shipping industry for crude oil usually derives efficiency based on the total distance that vessels traveled without cargo (also known as traveling "in ballast"). According to that, the system is more efficient the smaller this distance gets. But there is also a variety of other quantities that provide information on how efficiently the system operates, e.g. total travel distance, total travel time, total time traveled in ballast, total cost, fuel consumption, total emission, vessel speed, etc.

Based on these considerations, we characterized different fundamental metrics for indicating transport efficiency. These metrics are incomparable with each other and cover different aspects of efficiency. Among them, primary metrics are distinguished from a secondary metric.

The primary metrics are all incomparable quantities with different units representing different resources. The smaller these required resources are, the more efficient is the system.

The main applications for these primary metrics include the determination of the arc weights (see Section 4.2.4) and the analysis of the simulation results (see Section 5.2.2). Depending on these applications, different units are used for the metrics (as mentioned below).

The primary metrics include:

- **Time**
  The metric time considers the aspect of utility, i.e., the faster the shipping, the better the service. For the application of time as arc weight, we use the unit seconds [s], while the total required time as part of the simulation result is indicated in hours [h].
3.1. Efficiency

- **SO₂-Emission**
  The SO₂-emission as metric represents the environmental aspect of efficiency, i.e., the less sulfur is emitted in form of SO₂, the more efficient is the system in this respect. As already mentioned, different units are used for different applications. In order to compare different vessel classes with each other, the unit applied for arc weights is tons of sulfur per tons of transported cargo \( \frac{t_{sulfur}}{t_{cargo}} \) (normalized per cargo ton). However, the total emission of a simulation as part of the simulation outcome is given in an absolute term, in tons of sulfur \( t_{sulfur} \).

- **Cost**
  Cost indicates the economical aspect of efficiency, i.e., the lower the cost, the higher the efficiency in this regard. Similarly to emission, the units for cost are dollar per ton of cargo \( \frac{USD}{t_{cargo}} \) or only dollars \( USD \) depending on the application.

Why cost instead of profit?
Instead of considering cost, one could have chosen the profit as a metric considering the economical aspect of efficiency. The main reason why we selected cost over profit lies in the determination of charter rates for vessels (e.g. time charter equivalent, TCE), representing the main source of revenue for vessel owners. These charter rates are crucial in order to determine the profitability of a vessel for servicing a demand. But these rates are very difficult to determine as they are strongly dependent on supply and demand of vessels in a specific region. Therefore, these rates are dynamically changing and would result from the simulation itself. In order to consider profits, an additional financial structure that dynamically interacts with the simulation framework would have to be considered. For simplicity, we therefore only considered cost in this first iteration of the framework.

The primary metrics are chosen very similar to the applied metrics in the AV-enabled mobility systems in [3, 54, 55], where CO₂ was considered instead of SO₂.

In addition to the primary metrics, there is also a secondary metric that can be characterized:

- **Certainty for Planning**
  The secondary metric considers the aspect of predictability referring to the probability of a specific event to happen and its impact. The higher this probability is, the better we can plan events and the better we can manage and avoid risks. The probability for an event to happen can be indicated as percentage \( \% \), while the impact of such an event can be indicated by the primary metrics (time, emission, cost) representing the potential consequences.

During this project, we only focused on the primary metrics as the secondary metric is not considered so far by the first iteration of the framework.

Additionally, it is worthwhile to mention that the chosen primary metrics can have different correlations with each other, strongly dependent on the current situation. Considering the example of a vessel traveling on a path through a canal compared to traveling on an alternative path without canal, these metrics behave differently with respect to each other according to the congestion situation at the canal. In case there is no/little congestion at the canal with a short waiting time, then travel time and emission usually decrease (due to shorter distance), while the cost increases (canal costs are high). In contrast to that, for high canal congestion with long waiting times, the travel time increases (due to long delays), while the emission still decreases and the cost still increases.

These trends and correlations will be very important when we analyze and discuss the simulation results in more detail (see Chapter 5).
Chapter 3. Theoretical Part

3.2 Transportation Network

The model of the maritime transportation network builds the foundation for the simulation framework, as it defines how the vessels can move within the system.

3.2.1 Graphs

The entire transportation network is represented by a digraph $G_{\text{total}} = (V, A)$, consisting of a set of nodes $V$ and a set of arcs $A \subseteq V \times V$, shown in Figure 3.1 at the end of this section. In order to better understand the ideas behind this model, it will be derived step by step starting with the concept for ports.

The set of ports $P \subseteq V_{\text{Land}} \times V_{\text{Departure}} \times V_{\text{Arrival}} \subseteq V \times V \times V$ is based on different subsets of nodes. A port $p \in P$ includes three different nodes that lie geographically close to each other. These three nodes exhibit different functions within the port. According to that, different subsets of nodes are distinguished, indicating three different node types:

- **Land nodes**
  The set of Land nodes $V_{\text{Land}} \subseteq V$ represents the maritime facilities where vessels dock for loading and discharging cargo or other related port processes. Therefore, Land nodes describe the origins and destinations of demands, as well as the initial positions of vessels (see Section 3.3.1).

- **Departure nodes**
  The exits of ports where vessels depart, after having done all required port processes, are described by the set of Departure nodes $V_{\text{Departure}} \subseteq V$.

- **Arrival nodes**
  The set of Arrival nodes $V_{\text{Arrival}} \subseteq V$ indicates the entrances to ports where vessels arrive after their travels.

The nodes are connected by different types of arcs dependent on the processes they represent, leading to multiple subsets of arcs:

- **Travel arcs**
  Travel arcs of the set $A_{\text{Travel}} \subseteq V_{\text{Departure}} \times V_{\text{Arrival}}$ connect Departure and Arrival nodes of different ports with each other. These arcs represent the actual traveling of vessels within the system, as vessels move along that arc from one port to another.

- **Load arcs**
  Load arcs $A_{\text{Load}} \subseteq V_{\text{Land}} \times V_{\text{Departure}}$ describe the loading process of vessels and other related departure procedures at a port. They connect Land and Departure nodes of the same port.

- **Discharge arcs**
  For indicating discharge processes and other related arriving procedures at a port, Discharge arcs of the set $A_{\text{Discharge}} \subseteq V_{\text{Arrival}} \times V_{\text{Land}}$ are applied. They direct from Arrival to Land nodes within the same port.

- **Stop arcs**
  The set $A_{\text{Stop}} \subseteq V_{\text{Arrival}} \times V_{\text{Departure}}$ describes all Stop arcs. These arcs represent general stops of vessels at ports during which the vessel does not have to dock at a specific facility in order to load or discharge cargo (no visit of corresponding Land node). These stops can have multiple purposes, e.g. bunkering or other voyage related waits within a path. Stop arcs directly connect Arrival and Departure nodes of the same port.

Arcs of type Load, Discharge or Stop are considered as port arcs, since they only connect nodes within the same port. In contrast to that, Travel arcs always connect Departure and Arrival
nodes of different ports (no loops at ports). Based on the loading status of the vessels that move along the arcs (ballast, laden), each subset of arcs can be further subdivided into corresponding ballast arcs and laden arcs. For example, the set of Travel arcs is composed of ballast Travel arcs $\mathcal{A}_{\text{Travel,ballast}}$ and laden Travel arcs $\mathcal{A}_{\text{Travel,laden}}$. It holds that $\mathcal{A}_{\text{Travel}} = \mathcal{A}_{\text{Travel,ballast}} \cup \mathcal{A}_{\text{Travel,laden}}$ with $\mathcal{A}_{\text{Travel,ballast}} \cap \mathcal{A}_{\text{Travel,laden}} = \emptyset$.

By analyzing the voyage data of Signal, it can be observed that the major part of events happening at ports includes Load, Discharge and Stop events. Each of these different events can be modeled by applying the introduced node and arc types, see Figures 3.1a to c.

Figure 3.1a shows the model of a Load event. An empty vessel travels along a ballast Travel arc and reaches the port at the Arrival node. The ballast Discharge arc then represents the port processes of the arriving vessel, connecting the Arrival node with the corresponding Land node. Once the vessel is at the Land node, the loading process of the vessel is denoted by the laden Load arc, connecting the Land node with the Departure node. Finally, the laden vessel leaves the port on a laden Travel arc.

The model for the Discharge event, shown in Figure 3.1b is similar to the one for the Load event, just reversed. Laden vessels enter and empty vessels leave the port. The idea remains the same. In order to avoid visiting the Land node during Stop events, Stop arcs directly connect the Arrival node with the Departure node. Depending on the loading status of the vessel, a ballast and laden case can be distinguished, as shown in Figure 3.1c.

By combining the models for the different events, the general model for ports $p \in \mathcal{P}$ is derived. Figure 3.1d shows a port in its most general form consisting of three nodes connected by six port arcs.

![Figure 3.1: Schematic models for (a) the Load event, (b) the Discharge event and (c) the Stop event. (d) General model for a port $p \in \mathcal{P}$ as combination of all three events.](image-url)
By connecting multiple ports via their Travel arcs, we get a digraph $G_{\text{vessel class}}$, depicted in Figure 3.2. It represents the transportation network for a specific vessel class in form of a subgraph of the total transportation graph $G_{\text{total}}$.

![Figure 3.2](image)

Figure 3.2: General model for the transportation subgraph of a specific vessel class $G_{\text{vessel class}}$, resulting from the connection of ports.

Two different layers are considered, the Sea layer and the Land layer. As the Sea layer contains all Departure and Arrival nodes as well as all Travel and Stop arcs, it indicates where the actual traveling of the vessels within the digraph happens. The Land layer is composed of the set of all Land nodes and is connected with the Sea layer via Load and Discharge arcs. The Land layer represents the origins and destinations of demands as well as the initial positions of the vessels (more on that in Section 3.3.1).

For each vessel class (in this project VLCC, Suezmax and Aframax) we can model such a subgraph, as done in Figure 3.3. These subgraphs can be different according to the applied sets of nodes and arcs (see Section 4.2.2).

![Figure 3.3](image)

Figure 3.3: Schematic illustration of the transportation subgraphs for the different vessel classes, $G_{\text{VLCC}}$, $G_{\text{Suezmax}}$ and $G_{\text{Aframax}}$. They may differ according to the applied sets of nodes and arcs.

The vessels of a specific class are confined in their corresponding subgraph, i.e., VLCC vessels only move within the VLCC subgraph $G_{\text{VLCC}}$. The same holds for Suezmax and Aframax vessels accordingly with $G_{\text{Suezmax}}$ and $G_{\text{Aframax}}$. 


In contrast to vessels, cargo (here crude oil) that is transported by the vessels is not confined in a specific subgraph of a vessel class. Therefore, the different subgraphs can be combined to the total transportation graph $G_{\text{total}}$, illustrated in Figure 3.4.

![Total Transportation Graph](image)

Figure 3.4: Total transportation graph $G_{\text{total}}$, representing the full transportation network in crude oil shipping (for all vessel classes combined).

The subgraphs of the different vessel classes are connected by the Land nodes. As a result, only one Land node exists per port while Departure and Arrival nodes can occur multiple times for the same port, on the Sea layers of the different subgraphs.

Due to its structure, this general model for the maritime transportation network offers great modularity and flexibility.

The structure of the graph is theoretically able to model transshipment of cargo between vessels of the same or different vessel class at the Land nodes. In this context, transshipment represents the process of transferring cargo from one vessel to another. The new vessel carrying the cargo then completes the remaining path to the demand destination. Due to simplification, transshipment of cargo is so far not considered in the first iteration of the simulation framework. But for future iterations, this aspect provides an interesting opportunity to expand the current framework (more on that in the outlook, see Chapter 7).

In case an additional vessel class needs to be included within the model, the total transportation graph can be extended modularly by an additional Sea layer for the corresponding vessel class. Depending on the requirements the model has to meet (general or specific ones), different sets of nodes and arcs can be applied for generating the graph. This provides a large flexibility during implementation, as discussed in greater detail in Section 4.2.2.
3.2.2 Nodes and Arcs

Nodes and arcs describe the fundamental components of the total transportation graph $G_{\text{total}} = (V, A)$. Specific properties and details about these components are described in this section.

Each node $v \in V$ can be uniquely identified by determining:
- the location, i.e., the corresponding port $p(v) \in P$ which it is part of,
- the node type, i.e., Land, Departure or Arrival, and
- the vessel class whose subgraph it is part of, e.g., VLCC, Suezmax or Aframax (only necessary for Departure and Arrival nodes).

Arcs $a \in A$ are uniquely described by:
- the pair $(i, j)$, consisting of start and end node $i, j \in V$,
- the loading condition of the vessels that move along arc $a$, i.e., ballast or laden.

The model for graphs in its current iteration does not exhibit multiple parallel arcs with the same loading condition, i.e., there is at most a ballast arc and a laden arc between two nodes. Therefore, we do not have to consider additional parameters for the identification of the arcs.

Furthermore, each arc $a \in A$ has specific weights assigned to it which represent the required resources for vessels to transport cargo along that arc. These weights are indicated by the chosen primary metrics for efficiency, i.e., time, SO$_2$-emission and cost.

Assuming that these required resources are non-negative, the arc weights are defined as function $f : A \rightarrow \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+ (\text{in } \text{time/ton, } \text{SO}_2/\text{ton, } \text{USD/ton})$ with $a \rightarrow f(a) = (t, e, c)$. The function $f$ maps the set of arcs $A$ to the resource space $\mathcal{F}_R$ with $(t, e, c) \in \mathcal{F}_R$, representing specific arc weights as vectors of time $t$, emission $e$ and cost $c$.

For the application as arc weights, the metrics of emission and cost are normalized by tons of transported cargo, as already mentioned in Section 3.1. This allows us to compare arcs of different vessel classes. More specifically, the required resources for transporting a specific amount of cargo can be compared for multiple vessel classes as alternative shipping options.

However, in the current implementation of the framework, we directly imply the vessel class for a specific demand (more details in Section 3.3.1). Therefore, the consideration of multiple vessel classes as alternative shipping options.

Based on the definition of arcs, we can define a path $r$ as a sequence of connected arcs, with $r = \{a_1, a_2, ..., a_i, ..., a_n\} \in \bigcup_{n \geq 1} A^n$. The end node of an arc $a_i$ has to coincide with the start node of its subsequent arc $a_{i+1}$, so that $v_{\text{end}}(a_i) = v_{\text{start}}(a_{i+1})$, $\forall i \in \{1, 2, ..., n - 1\}$. A path conceptually describes a specific route within the transportation network that can be taken by a vessel.

Similar to the definition of the arc weights, the total weight $f(r)$ of a path $r$ can be calculated as the sum of the weights of the individual arcs covered by the path, $f(r) = (t_r, e_r, c_r) = \sum_i (t_{a_i}, e_{a_i}, c_{a_i})$. The path weight describes the total resources that a vessel requires in order to transport cargo along this path, applying the chosen primary metrics.

Depending on the choice of paths, different amounts of resources are required. This concept will be essential for describing preferences and simulation objectives in Section 3.4. Moreover, this leads to the consideration of canals and congestion in the next section below.
3.2.3 Canals and Congestion

Canals are very important for the maritime transportation, as they pose a main reason for alternative paths between two ports.

So far, the two most important canals are considered, namely the Suez Canal and the Panama Canal. While the Suez Canal can be passed by all considered vessel classes (VLCC, Suezmax and Aframax), the Panama Canal cannot be passed by VLCC vessels due to their large size.

As an example, Figure 3.5 shows two alternative paths from Rotterdam (A) to Singapore (B). The longer path around Cape Hope avoiding the canal is highlighted in red. The shorter path via the Suez Canal is depicted in green.

![Figure 3.5](image)

In the model of the transportation network, Travel arcs between two ports (from Departure to Arrival node) represent the direct path between these ports avoiding the canal. In order to model a canal, an additional Travel arc is therefore added between the entry ports of the corresponding canal. In case of the Suez Canal, these entry ports are Port Said (Mediterranean Sea) and Suez (Red Sea), in case of the Panama Canal these ports are Cristobal (Atlantic Ocean) and Balboa (Pacific Ocean). By this construction, paths through canals always contain Stop arcs at the beginning and end of the canal (entry ports) plus a Travel arc representing the actual canal.

Based on this graph structure, canal related costs or waiting times (due to congestion) can be specifically implemented within the arc weights of the Travel arcs representing the canals.

An interesting dynamic aspect of the crude oil shipping market is the congestion at canals. Due to the limited capacity of vessels passing a canal at the same time, congestion at canals can occur dynamically, causing additional waiting times and delays for the vessels.

Based on previous models regarding road congestion with vehicles [58, 59], we introduced a congestion function with \(d = f_{\text{canal}}(n)\). This function describes the additional waiting time at a canal \(d\) as a function of the number of vessels currently waiting at the canal or passing the canal (vessels on Travel arc of canal, \(n\)).
There is a variety of options how this function could be defined. The function for our model, representing congestion at canals, is derived from a road congestion model which uses so-called 'BPR functions' (by the Bureau of Public Roads) and their approximations [59].

Figure 3.5b schematically illustrates the applied congestion function. It is fully defined by three parameters: delay at regular operation \( d_{\text{regular}} \), critical number of vessels at and in the canal \( n_{\text{crit}} \) and slope \( q \).

- The regular delay \( d_{\text{regular}} \) indicates the waiting time at the canal at regular operation without congestion.
- The delay at the canal \( d \) is assumed to be constant up to a critical number of vessels at and in the canal \( n \leq n_{\text{crit}} \), representing the case of no congestion.
- For vessel numbers that exceed the critical number \( (n > n_{\text{crit}}) \), the delay \( d \) linearly increases by slope \( q \) for each additional vessel at and in the canal. This is the case of congestion.

For the sake of completeness, the formal definition of the applied canal congestion function is given by

\[
 f_{\text{canal}}(n) = \begin{cases} 
 d_{\text{regular}}, & \text{if } n \leq n_{\text{crit}} \\
 d_{\text{regular}} + q(n - n_{\text{crit}}), & \text{if } n > n_{\text{crit}}
\end{cases}
\]

Additional details regarding the implementation of this congestion model are presented in Sections 4.3 and 4.4 of the methodology.
3.3 Simulation Framework

The simulation framework as a whole describes the crude oil shipping market in general. This includes multiple aspects, as mentioned in Chapter 2. The definition of the dynamics between supply and demand represents the central part of the simulation framework, as it covers the core mechanisms, procedures and dependencies. This section discusses the dynamics of the system, starting with the concepts for demands and vessels.

3.3.1 Demands and Vessels

The total demand of the system is formed by the aggregate of all individual demands, representing cargo owners or charterers who want to ship their cargo from one place to another. More precisely, each demand includes multiple properties:

- **Identification number**
  The identification number distinguishes demands from each other. Therefore, each number is different.

- **Demand origin / destination**
  The demand origin and destination indicate from where to where the cargo has to be transported. Both locations are represented by Land nodes $v_{\text{origin}}, v_{\text{destination}} \in V_{\text{Land}}$. The cargo is loaded at $v_{\text{origin}}$ and discharged to $v_{\text{destination}}$. Due to simplification reasons, only demands with single Load and Discharge events are considered in the current iteration of the framework.

- **Cargo size**
  The cargo size represents the quantity of the specific good that has to be transported, here crude oil. The cargo size is indicated in metric tons $|t_{\text{cargo}}|$.

  Because we have not distinguished between different types of crude oil so far, the model only considers a single cargo type in its current iteration. Moreover, since only demands with single Load and Discharge events are addressed, the demands are assumed to represent full shiploads. For this reason, the demands directly imply the vessel class of the vessel that has to be assigned for servicing a specific demand. Potential refinements regarding the definition of demands for future iterations of the framework are further discussed in the outlook (see Chapter 7).

- **Fixed dates for planning (timeline)**
  As further discussed in Section 3.3.2 below, there are specific dates in the timeline of a demand that are important for planning. These include the dates of release, vessel and path assignment, laycan start and laycan end, etc.

- **Assigned vessel with paths**
  In order to service a demand, a suitable vessel has to be evaluated. In addition to that, a path for the specific vessel to the demand origin (if vessel not already there) as well as a path from the demand origin to the demand destination has to be selected. Both, vessel and paths are then assigned to the corresponding demand.

  As a simplification in this iteration of the framework, the vessel class of the assigned vessel is predetermined for each specific demand. However, as already mentioned in Section 3.2.2, the arc weights are defined in a way that allows a comparison between different vessel classes as shipping options. Therefore, the current model for demands serves as basis for further refinements and extensions, as addressed in the outlook (see Chapter 7).
The total supply of the system is represented by all the vessels. They are provided by ship owners for servicing demands, i.e., for the transport of crude oil from origin to destination. A vessel is characterized by different properties:

- **IMO number and vessel name (identification)**
  Similar to the purpose of a name, each vessel owns a unique number for identification. This so-called International Maritime Organization (IMO) number was introduced with the intention to increase security and safety in maritime transportation [60].

- **Vessel class**
  The vessel class of a ship refers to its type, dependent on size and capacity of the vessel. In the scope of this project, addressing the crude oil shipping market, the three considered vessel classes are VLCC, Suezmax and Aframax.

- **Cargo capacity**
  The cargo capacity describes how much cargo a specific vessel can carry. There are many different quantities for the capacity of a vessel. One of the most common quantities, also applied in this framework, is the deadweight tonnage (DWT). It describes the weight carrying capacity of the ship in metric tons, excluding the vessel’s empty weight [7].

- **Vessel status**
  The vessel status indicates whether a vessel is assigned to a demand or not. Vessels that are not assigned (available), usually reallocate themselves in order to increase their probability of being assigned to a new demand.

- **Loading status**
  The loading status of the vessel denotes how the vessel is traveling, i.e., either in ballast or laden. This dictates on which specific arcs the vessel can move (ballast or laden arcs). In this iteration of the framework, vessels can either be empty (in ballast) or completely laden, meaning only full shiploads are considered.

- **Initial / current position**
  Vessels have a specific position within the transportation network which is represented by graphs. Generally, the initial and current position of a vessel are described by specific nodes. In case the vessel is currently moving on an arc between two nodes, the corresponding arc serves as an indicator for the current vessel location.

- **Assigned paths**
  Before vessels travel along a certain path, it has to be evaluated and assigned first. This includes the paths of the vessels to the demand origin, from the demand origin to the demand destination as well as paths for reallocation. Which paths we select, based on specific preferences and simulation objectives, is very important for the outcome of a simulation. More details on this are given in Section 3.4.

Vessels are the "moving" particles of the simulation framework, as they move within the transportation network (represented by the digraph $G_{\text{total}}$ in Section 3.2.1). Because all vessels within the framework are described individually, we can consider it as a particle flow model. This approach is mainly possible due to the relatively small number of vessels that has to be considered by the model. Compared to other transportation systems, e.g. road or public transportation, the numbers of vehicles or trains are too large to describe each of them individually. In those cases, arc flow or path flow models are used more commonly (see details in the literature review in Section 1.3).
3.3.2 Simulation Procedure

The simulation framework essentially represents a model for processing demands and vessels. Based on this idea, there are two different perspectives on the framework (one for demands, another one for vessels).

It should be mentioned that all the discussed schematic procedures below only represent a simplified and generalized version of the procedures happening in reality. The real procedures can vary a lot due to the lack of structure in the maritime transportation system, mentioned in Section 1.3.

General Timeline

![Diagram showing the general timeline for the entire simulation procedure.](image)

Demand Timeline

![Diagram showing the timeline for demands.](image)

Vessel Timeline

![Diagram showing the timeline for vessels.](image)

Figure 3.6: General timeline for the entire simulation procedure, with additional perspectives on the specific timelines for demands and vessels.

In Figure 3.6, the general timeline of the simulation framework is shown. The simulation starts at a specific time \( T_{\text{start,demands}} \) and ends at time \( T_{\text{end,demands}} \). In between, there is a time interval when new demands are released and processed, indicated by \( T_{\text{start,demands}} \) to \( T_{\text{end,demands}} \).

Based on the perspective for demands, Figure 3.6 depicts the timeline for demands in detail. As already mentioned in Section 3.3.1 regarding the properties of demands, there are four important dates for generating and processing demands:

- **Release**
  The release date denotes the point in time when the demand is released, i.e., it is presented to the vessel owners. Usually, the release lies about 10 to 20 days before the laycan start. Once the demand is announced, potential vessels can be evaluated.

- **Vessel and path assignment**
  At some point, a vessel for the demand has to be assigned together with the corresponding paths it takes to transport the cargo. The choice of vessel and paths is highly dependent on the underlying preference for a specific simulation, strongly influencing the simulation outcome (more details on that in Section 3.4). As a simplified assumption, this date represents the deadline when charterer and vessel owner agree by contract on prize, terms and conditions for the transportation of the cargo.
• **Laycan start/end**
Laycan denotes the time window during which the vessel owner has to present the ship to the cargo owner for loading. If a ship arrives before the laycan start, the charterer can refuse to load the vessel until the laycan begins. In case the ship arrives after the laycan end (canceling date), the charterer can cancel the contract and potentially look for another vessel to transport the cargo.

While the four dates above are assumed to be predefined (known at demand generation), the date of completion (when the cargo arrives at destination) represents an outcome of the dynamic simulation.

Additionally, Figure 3.6 illustrates the detailed timeline for a vessel. Depending on the location of the vessel, different cases have to be considered:

• **Vessel arrives before laycan start (early)**
As mentioned above, charterers can refuse to load the cargo when the vessel arrives before the laycan start. Therefore, we assume that the vessel will wait at the demand origin until the laycan begins in order to load the cargo (strongly simplified procedure compared to reality). After the wait, the vessel is loaded (described by a laden Load arc) and starts its travel along the assigned path (via laden Travel arcs).

• **Vessel arrives during laycan (on time)**
In case the vessel arrives on time, i.e., within the time window represented by the laycan, the vessel is loaded and then moves on to the demand destination.

• **Vessel arrives after laycan end (too late)**
If a vessel arrives after the laycan end at the demand origin, the charterer can cancel the contract and look for another potential vessel. However, in this iteration of the framework, we assume that the charterer continues to ship its cargo with the delayed vessel. In many cases, assigning a new vessel and waiting for its arrival at the demand origin would take more time than just moving on with the delayed vessel.

Regarding the scheme in Figure 3.6, the loading and discharging processes at the demand origin and demand destination are included within the travel to the demand destination, because the corresponding laden Load and laden Discharge arcs represent the first and last arcs of the assigned path to the demand destination.

Once the current demand is completed with the arrival at the demand destination (including discharging), again multiple cases can be distinguished depending on the vessel status:

• **Vessel already assigned to new demand**
In case the vessel has already been assigned to a new demand, the same procedure for the vessel starts again, i.e., the vessel travels to the new demand origin (if it is not already there), etc.

• **Vessel not assigned, looking for new demand**
If the vessel has not already been assigned to a new demand, we assume that the vessel waits first (based on the simplified and generic scheme). Ideally, a new demand comes in after a certain time and the vessel procedure recurs from there. In case the vessel does not find a new demand at its current location during a specific period of time, then the vessel will reallocate itself in order to increase the probability of getting assigned. At the new location after reallocation, the vessel then waits for a new demand in order to get assigned, and then for starting the procedure all over again.
Based on the structure of the timeline, it is also possible to describe these procedures for processing demands and vessels in a more schematic way, as illustrated in Figures 3.7a and b.

Figure 3.7: (a) Schematic procedure for processing demands, distinguishing demands according to their state of completion. (b) Schematic procedure for processing vessels, distinguishing states of traveling and waiting.

Figure 3.7a schematically shows how demands are processed step by step. Demands are subdivided according to their state of completion.

At first, all demands are considered as unreleased until they are announced to the vessel owners. At this point, they are considered as released. As a next step, a vessel and corresponding paths are assigned to the demand. Potential vessels for assignment include not only available vessels (currently unassigned) but also vessels which are currently completing another demand (currently already assigned). Depending on the status and current position of the assigned vessel, different scenarios can be distinguished. In case the assigned vessel has to finish another demand first, the demand is denoted as assigned, but on hold. As soon as the vessel has completed the current demand by arriving at its destination, the vessel travels to the origin of the new demand (now denoted as assigned, with vessel on the way to origin). In the other case, when the vessel is directly available, the vessel immediately starts to travel to the demand origin after assignment. Once the vessel arrives at the demand origin, or in case it was already there by coincidence, the demand is indicated as assigned, with vessel on the way to destination. This state of completion describes the travel of the vessel to the demand destination, including the loading and discharging processes at the demand origin and the demand destination. Once the vessel arrives at the demand destination, the demand is assumed to be completed. In case no potential vessels for assignment exist over a specific time range, a demand is assumed to be unfeasible. This would mean that the demand for maritime transportation is significantly higher than the corresponding supply of vessels over the specific range in time.

The schematic procedure builds the foundation for the implementation of the simulation framework, as the main part of the implementation is identically structured (perspective of demands, more details are discussed in Section 4.3.3).
Similarly, we also present a detailed scheme for processing vessels, shown in Figure 3.7). Within this procedure, vessels are assumed to either travel or wait.

As shown in Section 3.2, Load, Discharge and Stop events are represented by Load, Discharge and Stop arcs between different nodes. Additionally, these port arcs are included in the paths covered by the vessels (paths consist not only of Travel arcs). Therefore, the port events are considered here as part of the traveling.

Waiting specifically describes the process of a vessel waiting at the same node (equivalent to a loop within the graph), i.e., it does not include port processes.

As already discussed for demands, vessels can go through different scenarios depending on the location of the vessel and its status regarding demands (assigned to new demand or not).

Once a vessel is assigned to a demand, it will travel to the corresponding demand origin as soon as possible. In some cases, the vessel has to complete another demand first. Depending on the arrival date of the vessel at the demand origin, it may wait until the laycan start, before traveling to the demand destination. During the travel to the demand destination, the vessel can be already assigned to a new demand. Depending on the vessel status after the demand completion, different scenarios are distinguished. If the vessel is already assigned to a new demand, the entire process will be repeated by traveling to the new demand origin (in case it is not already there). On the other hand, in case the vessel has not been assigned to a new demand, it will wait at the current location first. If the vessel is still unassigned after a certain time, it will reallocate to a place where the probability for a new assignment is higher, e.g. at frequently used demand origins (find more details in Section 4.3). After the arrival of the vessel at the new destination after reallocation, it will wait there for a new demand. During the entire reallocation process of the vessel, including the waitings as well, a new demand can be assigned in order to start the entire vessel procedure again.

While the structure of the implementation is mainly based on the procedure for demands, vessels are processed in between the demand processing. Further details on that are explained in Section 4.3.

The general structure of the simulation framework is fixed based on the discussed procedures above. However, the decisions made within this structure heavily influence the outcome of a simulation. These decisions are based on specific preferences and simulation objectives which lead to the next section.
3.4 Preferences and Objectives

The theoretical model for the crude oil shipping market considers multiple different aspects, as discussed in Sections 3.1 to 3.3. First, we characterized metrics for transport efficiency in order to make statements on how efficiently the system operates. Furthermore, the transportation network is modeled by the digraph $G_{\text{total}}$ which builds the foundation for the simulation framework. The main part of the simulation framework then describes the detailed procedure for processing demands and vessels.

As graphs, demands, vessels and the simulation processes seem generally given, one might ask how the simulation outcome can actually be affected.

The answer to that lies in the decisions made during the simulation. They affect the simulation outcome. In the current iteration of the framework, these decisions are basically:

- Vessel assignment
- Paths selection (path to demand origin, path from demand origin to demand destination)
- Vessel reallocation

The specific decisions we take, usually depend on the strategy we pursue. Based on that, the concept of preferences was introduced.

3.4.1 Preferences

The weights for arcs and therefore also the weights for paths are determined by the primary metrics for efficiency, as discussed in Section 3.2.2.

As shown later in Section 3.2.2, the outcome of a simulation includes the total resources required for servicing all demands in the corresponding simulation. These resources (as accumulation of weights) are accordingly evaluated in the same metrics, i.e., in time, emission and cost.

As already mentioned, these metrics represent incomparable quantities with different units. Due to the fact that they generally exhibit inverse or contrary relations with respect to each other, it is usually unfeasible to minimize all of them simultaneously in one simulation. Therefore, a specific preference has to be taken before running a simulation.

Based on the choice of preference, the objective of the simulation changes as well as the applied strategy in form of the decisions that are taken. Each preference tries to be as efficient as possible by reducing a specific metric or combinations of them in the resulting simulation data.

Simple examples for preferences are the ones that specifically minimize a single metric, e.g. preference time tries to specifically minimize the total required time for servicing demands during simulation. The same holds for the preferences emission and cost correspondingly.

In order to consider also mixed preferences, we introduced a more general concept with a more formal definition of preferences. The general weight $w(a)$ for an arc $a \in \mathcal{A}$ is defined as linear combination of the normalized metrics time $\bar{t}$, emission $\bar{e}$ and cost $\bar{c}$. Applying parameters $\alpha$, $\beta$ and $\gamma$ as prefactors results in $w(a) = \alpha \cdot \bar{t} + \beta \cdot \bar{e} + \gamma \cdot \bar{c}$.

It is worthwhile to emphasize that the metrics in the expression above are normalized (therefore dimensionless metrics) in order to add them up. Ideally, the normalization constants are chosen such that the normalized metrics exhibit a similar order of magnitude. Therefore, the influence of the parameters $\alpha$, $\beta$ and $\gamma$ on the general weight $w(a)$ highly depends on these normalization con-
Based on the definition above, a general preference \( x \) is then described by the three parameters \( (\alpha, \beta, \gamma) \).

Negative arc weights for graphs are unreasonable, therefore \( \alpha, \beta, \gamma \geq 0 \). Because it only matters how the general weights of different arcs relate to each other, an additional constraint can be imposed without losing generality, so that \( \alpha + \beta + \gamma = 1 \).

Consequently, the set of all different preferences \( \mathcal{X} \) represents a subset of the total preference space \( \mathbb{R}^3_+ \), so that \( x \in \mathcal{X} \subseteq \mathbb{R}^3_+ \). Figure 3.8 illustrates the set \( \mathcal{X} \) as triangular surface in the preference space.

**Figure 3.8: Preference space \( \mathbb{R}^3_+ \) with the set of all different preferences \( \mathcal{X} \subseteq \mathbb{R}^3_+ \) indicated in blue, containing all \( x \in \mathcal{X} \) for which \( \alpha + \beta + \gamma = 1 \).**

### 3.4.2 Simulation Objective

In this framework iteration, the simulation objective based on a specific preference describes the minimization of the general weights for the individual paths taken by the vessels for servicing demands. This means that during each process of path assignment, the current path of smallest general weight is determined and assigned to the corresponding vessel.

The optimization problem is thus described by finding the path of smallest general weight as

\[
\arg\min_r w(r) = \arg\min_r \sum_{a \in r} w(a),
\]

with \( r = \{a_1, a_2, \ldots, a_i, \ldots, a_n\} \), \( v_{\text{start}}(a_1) = v_{\text{origin}} \), \( v_{\text{end}}(a_n) = v_{\text{destination}} \), \( n \in \mathbb{N}_{>0} \),

where a potential path \( r \) between demand origin \( v_{\text{origin}} \) and demand destination \( v_{\text{destination}} \) has a general weight \( w(r) \). Similarly, the problem could also be formulated for the assigned path of the vessel to the demand origin (not necessary if vessel already there).
A small remark regarding the assignment of the path to the demand origin: In this case, the path of minimal general weight is assigned to the vessel, which still enables the vessel to arrive at the demand origin before the laycan end (time constraint). If no path can fulfill this time constraint, the path of minimal travel time is assigned. This way, the vessel counteracts a potential late arrival at the demand origin. More details on that are given in Section 4.3.3 of the methodology.

The specific "shortest" path (in this more general sense) can be found by different algorithms. In the current implementation, we applied Dijkstra’s algorithm for shortest paths, as will be further discussed in Section 4.3.3. More details on Dijkstra’s algorithm are given in the corresponding literature [61, 62].

Generally, the entire concept of preferences and objectives provides an approach in order to account for mixed preferences in simulations.

It should be mentioned that this approach is not directly meant for making absolute statements about specific preferences because they highly depend on the applied units and normalization constants. And the metrics time, emission and cost remain incomparable quantities, even though we apply them in the same mathematical expression.

The main idea behind this approach is to understand the relations between different preferences qualitatively based on the defined efficiency metrics. Each preference leads to a different strategy that results in a different simulation outcome. By that, we ideally find different efficient simulation outcomes in the sense of Pareto efficiency.

In this regard, a simulation outcome is described as Pareto efficient, if there is no other outcome (based on same simulation conditions but different preference) that requires less resources for at least one single metric without worsening the remaining other metrics. In contrast to that, simulation results that are not Pareto efficient are described as Pareto dominated [63, 64].

The specific implementation of this entire concept as well as its application within different simulations are further discussed in the subsequent Chapters 4 and 5 on methodology as well as results and discussion.
Chapter 4

Methodology

In this chapter, the methodology applied for this project is described. This mainly includes the implementation and application of the theoretical aspects of the simulation framework, introduced in the previous chapter. At first, general details about the implementation are considered. Afterwards, we look at the implementation and generation of the graphs, representing the maritime transportation network. Subsequently, the general structure of the simulation framework is addressed, including all the different subcomponents. One of these subcomponents represents the congestion estimator for canals, which is separately discussed. At the end of the chapter, the simulation setup and analysis are further explained, i.e., how the specific simulations were run and evaluated.

4.1 General Details on Implementation

The total simulation framework based on the mathematical model, as discussed in Chapter 3, is implemented in the programming language C++. We applied CLion as corresponding IDE (version 2020.3.2). The implementation of the framework represents the main work of this project with over 10'000 lines of code. There are multiple reasons why we chose C++ over other programming languages:

- C++ is an object-oriented programming language. Especially, the concept of classes facilitates the implementation of different objects like demands, vessels, etc.
- It is widely used, running on many different platforms, e.g. Windows, Linux, Mac, etc.
- There are many external libraries that provide predefined methods and structures. For this project, we also applied external libraries.
- C++ can be difficult to initially start with because variables and objects have to be defined more rigorously compared to other programming languages (e.g. Python). But due to this fact, C++ is very fast and efficient, especially for processing a large amount of data.

For the implementation of the entire framework, we used two external libraries: JsonCpp (version 1.9.4) and Boost Graph Library (BGL, version 1.71.0). In order to facilitate the setup of these external libraries, the Package manager Conan was applied (version 1.2.11). Furthermore, to set up and run our program on different platforms (Windows, Linux, Mac, etc.) as consistently as possible, we made use of Docker (version 20.10.5). This software packs the program into containers, which facilitates the setup of all the dependencies and packages of the program on a new computer.

Signal provided us access to their voyage data, to their online platform as well as to their API. In order to apply the predefined methods of the API, Python (version 3.7.4) was used with Spyder as corresponding IDE (version 3.3.6).
4.2 Implementation of Graphs

The implementation of the graphs, which were introduced in Section 3.2.1, represents the first step of implementing the complete simulation framework. This step is important as it affects all the others that follow during the remaining implementation.

The graphs build the model for the transportation network where the individual vessels move. The first and central question that arises during the implementation of these graphs refers to the applied set of nodes and arcs. Depending on the applied nodes and arcs, the graphs will have different sizes and will be more or less general. To better understand this problem, we present some statistics of the voyage data by Signal, covering the last 7 years.

4.2.1 Voyage Data provided by Signal

The voyage data provided by Signal contains detailed information about the individual voyages that the vessels have done during the last 7 years (January 2014 to January 2021). The data is split into three different files, one for each vessel class (VLCC, Suezmax and Aframax). Each file is given either in DAT- or JSON-format (with corresponding file extensions .dat or .json). It is worthwhile to mention that Signal’s data includes information about all vessels of a specific vessel class globally, i.e., it covers more than just the voyages for fleets of specific companies.

Signal distinguishes in total 2,364 different ports and 9,399 different geoassets, i.e., each port represents on average approximately four geoassets. Interestingly, from all potential ports only 993 different ports were actually approached in the voyage data, considering all three vessel classes.

More precisely, only 502 different ports were used by VLCC vessels, 659 ports by Suezmax vessels and 842 ports by Aframax vessels. This information and other details are shown in Table 4.1.

| Table 4.1: Statistics based on Signal’s voyage data regarding the three vessel classes VLCC, Suezmax and Aframax. |
|---|---|---|
| **Number of vessels** | 953 | 702 | 1’221 |
| **Number of different ports visited** | 502 | 659 | 842 |
| **Number of Load ports** | 186 | 348 | 568 |
| (ports where Load events happened) | | | |
| **Number of Discharge ports** | 202 | 331 | 523 |
| (ports where Discharge events happened) | | | |
| **Number of voyages/trips** | 27’070 | 34’383 | 89’299 |
| **Number of events** | 121’657 | 132’380 | 301’119 |
| **Number of Load events** | 39’151 | 38’302 | 102’069 |
| **Number of Discharge events** | 35’105 | 40’137 | 102’036 |
| **Number of different Travel arcs covered in the voyage data** | 10’399 | 16’436 | 31’216 |
| **Number of different Travel arcs applied in the graphs (see Section 4.2.2)** | 503’004 | 867’244 | 1’416’244 |
Furthermore, less than 1% of all potential Travel arcs were actually taken by the vessels. This holds when different loading conditions (ballast and laden) as well as all possible port combinations (from Departure to Arrival node) would be considered (\(2 \cdot 2^{364} \cdot 2^{363} \approx 10^9 000\) Travel arcs).

### 4.2.2 Applied Nodes and Arcs

Based on the considerations above, there is a conflict between generality and usability of the implemented graphs, strongly dependent on the applied nodes and arcs:

- By applying all potential nodes and arcs, the graphs would be very general, enabling all potential paths within the system. However, due to their large size, they would be very time-consuming to generate and difficult to handle.

- In contrast to that, implemented graphs which are specifically based on the voyage data would be significantly smaller in size and very data specific. But as a downside, the framework based on these graphs would be limited by past habits. A loss of generality would be observed.

Therefore, a tradeoff between generality and usability was made during the implementation of the graphs:

- All used ports by a specific vessel class are applied (based on Signal’s voyage data). Moreover, all potential Travel arcs between these ports are considered (more general than Signal’s voyage data, see Table 4.1).

- Each port is represented in its most general way, consisting of three nodes connected by six different port arcs. See model for ports illustrated in Figure 3.1d of Section 3.2.1.

Within this tradeoff, the nodes are data specific, while the arcs are as general as possible. By this approach, we account for the fact that not every vessel class uses the same set of ports (due to size restrictions, specific docks, etc.). This means that the sets of nodes \(\mathcal{V}_{VLCC}\), \(\mathcal{V}_{Suezmax}\) and \(\mathcal{V}_{Aframax}\) applied for the corresponding vessel class subgraphs \(\mathcal{G}_{VLCC}\), \(\mathcal{G}_{Suezmax}\) and \(\mathcal{G}_{Aframax}\) are different based on their specific voyage data.

Additionally, ports that are accessible by vessels of the same vessel class are always directly connected with each other. Therefore, based on a given set of ports, all possible Travel arcs between such ports are considered (Travel arcs from Departure to Arrival node represent the direct path without canal).

More intuitively, the implemented subgraph for a specific vessel class should always be fully connected, i.e., between two nodes there should be no arc missing. This idea is schematically visualized in Figure 4.1 of the following section.

As an example, the case for the vessel class VLCC is depicted. The total VLCC subgraph \(\mathcal{G}_{VLCC}\) includes 502 ports, represented by \(3 \cdot 502\) nodes, plus \(6 \cdot 502\) port arcs and \(2 \cdot 502 \cdot 501\) Travel arcs. We have to keep in mind that there is no Travel arc pointing from the Departure node to the Arrival node of the same port (therefore only \(2 \cdot 502 \cdot 501\) possible Travel arcs to be considered).

It should be mentioned that small inconsistencies were found in the voyage data of Signal. The nomenclature for ports and geoassets was incorrect in a few cases. Referring to the outlook in Chapter 7, the graphs need to be updated as soon as the corrected voyage data will be available.
4.2.3 Graph Implementation and Generation

The total transportation graph $G_{total}$ represents the entire maritime transportation network for crude oil shipping, discussed in Section 3.2.1. Due to the fact that this graph is very large in size and therefore difficult to generate and unpractical to handle, we implemented several smaller subgraphs instead of just this one large graph.

By applying multiple smaller subgraphs for different applications, the implementation is significantly facilitated:

- Subgraphs for different vessel classes are implemented separately, leading to the total subgraphs for each vessel class.

  The total subgraphs for each vessel class ($G_{VLCC}$, $G_{Suezmax}$ and $G_{Aframax}$) are applied for moving vessels within the transportation network from one node to another (see Figure 4.1a for the case of VLCCs).

- For path finding problems as part of the path assignment, the smaller ballast or laden Sea subgraphs are applied for each vessel class. These subgraphs only cover the Sea layer, including uniquely Departure and Arrival nodes which are connected by Stop and Travel arcs for a specific loading condition (ballast or laden). Figure 4.1b schematically illustrates these even smaller subgraphs for the example of VLCC.

Because we only consider demands with one Load and Discharge event in the current iteration of the framework (see Section 3.3.1), the first Load and last Discharge arc of an assigned path are predetermined. This means that the vessel will not leave the Sea layer during the actual travel on the path between the Departure node of the origin port and the Arrival node of the destination port. This enables the application of the smaller ballast or laden Sea subgraphs for path finding problems during path assignment.

The general idea behind the implementation concept is to always apply the smallest possible subgraph for a specific application.

Figure 4.1: (a) Total VLCC subgraph $G_{VLCC}$ with all possible Travel arcs and ports in their most general form. (b) Applied smaller ballast and laden Sea subgraphs for VLCCs.
In order to implement and generate these multiple subgraphs, we went through multiple steps:

- First of all, we extracted all the used ports from the voyage data for each vessel class. Initially, a predefined method of the external JsonCpp library was applied to read in the data from the corresponding files (in JSON-format). But due to the large data size, this read-in process was very time-consuming and unpractical. Therefore, we implemented a faster and more efficient method which reads in the data from the files in DAT-format.

- As a next step, the graphs were actually implemented by applying the external Boost Graph Library (BGL). This library provides many different predefined structures and methods for implementing and processing graphs. In our case, the graphs were implemented as adjacency lists (see specific details in the BGL documentation [65]). The predefined graph structure allowed to subsequently add the selected nodes and arcs to the graphs.

- Furthermore, the specific arc weights for the graphs were determined. This specific process represents an important aspect in the implementation of the graphs and is therefore separately discussed in the subsequent Section 4.2.4.

- Lastly, the different graphs were generated and read out as DOT-files (with extensions .dot) by applying a predefined read-out method of the BGL. This step had to be done only once.

Once the last step had been completed, the graphs were directly read in from the DOT-files via a predefined BGL read-in method. This way, we avoided the time-consuming generation process for the graphs.

### 4.2.4 Calculation of Arc Weights

Arc weights are very important for the process of vessel and path assignment, as they determine the required resources for vessels to move along the corresponding arcs. Therefore, they also strongly affect the weights for specific paths which are defined as sum of individual arc weights.

Due to the fact that many more arcs are considered for the graph implementation than mentioned in the voyage data of Signal (see previous Section 4.2.2), we decided to calculate the corresponding arc weights, instead of deriving them from the data.

This means that the formal function \( f(p) = (t, e, c) \), defining the arc weights (see Section 3.2.2), is independent from the voyage data.

The arc weights for the Travel arcs are based on the travel distances between the ports. In order to determine these distances for all combinations of ports, Signal provided us access to their API platform. The predefined methods of the API not only calculated the specific Travel distance between two ports but also the corresponding distance covered in (S)ECAs.

As introduced in Section 1.3, Emission Control Areas (ECAs) or Sulfur Emission Control Areas (SECAs) describe sea regions in which airborne emissions caused by ships are regulated more strictly. Within these areas cleaner, more expensive fuel has to be consumed in case the vessels do not have an approved exhaust gas cleaning system. Based on the newest regulations since 2020, the sulfur limit for fuel in SECAs is 0.1% (as mass percentage), while the general sulfur limit in other sea areas is 0.5% [2].

Regarding the process of calculating the different distances by using the API of Signal, multiple steps had to be considered:

- After evaluation of the ports and all their potential combinations, their read-out was established in form of a TXT-file (with extension .txt). This way, the information could then be applied in Python.
• In Python, the API requests were performed to determine the distances for each individual port combination. This step, as sum of all individual requests, represents one of the main factors that limits the size of the implemented graphs. A single request for an arc may only take few hundreds of a second, but millions of them can easily require days or even weeks to be processed. This is one reason why we went for a tradeoff in the application of nodes and arcs (mentioned in Section 4.2.2).

• The distances for all port combinations were stored in a JSON-file. Finally, the data from this file was read into the framework in C++ to determine the arc weights for the graphs.

In addition to the distances from the API (total distances and distances covering (S)ECAs), many other parameters were applied for the calculation of the arc weights. They include:

• **Vessel speed**
  In the current iteration of the framework, we assumed a constant vessel speed of 13 kn (1 kn = 0.514 m/s). Applying the distances from the API and the vessel speed, the corresponding travel times were calculated. Additionally, the framework accounts for the times traveled in (S)ECAs with clean fuel as well the ones traveled on open sea with dirty fuel.

As later discussed in the outlook (see Chapter 7), applying a non-constant vessel speed in dependence of the local supply and demand of the system would pose an interesting refinement for a future iteration of the simulation framework.

• **Fuel consumption**
  Based on the fuel consumption of a vessel per time traveling in ballast or laden, in combination with the evaluated travel times, the required amount of clean and dirty fuel for each Travel arc was determined. For Port arcs, the idle fuel consumption per time as well as the fuel consumptions for loading and discharging processes were considered.

• **Sulfur content of fuels**
  Once the fuel consumption of a vessel was determined for a specific arc, the sulfur emission was calculated by applying the specific sulfur contents of the clean and dirty fuel.

• **Fuel prices**
  Similarly to the emission, the fuel costs were evaluated by deploying the specific fuel prices for clean and dirty fuel. The fuel costs usually represent the major part of the operational costs in maritime shipping.

• **Other costs**
  By considering other costs in addition to the fuel costs, such as costs for loading and discharging as well as port and canal costs, the total cost of an arc was determined.

The proper definition of the arc weights based on all parameters would exceed the scope of this report. Thus, for more details regarding the calculation of the arc weights, we directly refer to the separate code documentation (README file) and to the code itself [57].

4.2.5 **Historical Graphs based on Voyage Data**

Regarding the generation of the graphs, the voyage data provided by Signal only affects the applied nodes. The arcs and their weights are general and independent from this data.

As we already knew the process how to implement and generate graphs, we also built very data-specific graphs as a nice supplement. These so-called “historical” graphs only contain corresponding nodes and arcs from Signal’s voyage data, one graph for each vessel class. Compared to the general graphs of the framework, these graphs do not contain all potential arcs, but only the ones which appeared in the voyage data.
The historical graphs were not applied in the simulation framework. Their real purpose is helping us to understand and analyze the voyage data, representing a special form of database.

These graphs contain all relevant information of Signal’s voyage data in an already extracted form. As an example, the number of Load, Discharge or Stop events that happened at a specific port were directly stored as attributes of the corresponding nodes and arcs. Similarly, the average travel time or the travel frequency of a specific Travel arc could be directly attributed to the corresponding Travel arc.

Moreover, while the historical graphs remain unchanged, the applied transportation graphs, as discussed in Section 4.2.2 may be adapted due to implementation changes.

Lastly, these graphs are sparse (weakly connected) and therefore relatively small in size (in total 20 MB). This makes them easier to handle than the larger voyage data files (in total 550 MB). For better intuition, Figure 4.2 schematically illustrates the historical graph for the VLCC vessel class, which appears rather sparse in contrast to the applied total VLCC subgraph, depicted in Figure 4.1a (see previous Section 4.2.3).

For the sake of completeness, the information of these “historical” graphs was added to the nodes and arcs of the applied transportation graphs which contain this way as much information as possible.

![Figure 4.2: Schematic illustration of the historical graph for VLCCs, exhibiting a sparse appearance (arcs are missing).](image-url)
4.3 Implementation of the Simulation Framework

The implementation of the simulation framework with all corresponding procedures, mechanisms and dependencies represents a major part of this project, in addition to the implementation and generation of the graphs. Generally, the implemented simulation framework processes demands by moving vessels around within the corresponding subgraphs.

The framework is structured modularly. The main components of the framework are discussed first, before looking into their corresponding subcomponents. For further details regarding the implementation of the framework, we directly refer to the separate code documentation (README file) and to the code itself [57].

![Diagram of Simulation Framework]

Figure 4.3: General structure of the entire implemented simulation framework. Three main parts are further divided into smaller subcomponents. The subcomponents marked in red represent the processing of the demands according to their state of completion.
4.3.1 General Code Structure

The implemented simulation framework is composed of three main parts:

- **Initialization**
  Within this first part, objects of different classes are initialized, defined, set or generated. It represents the preparation of the main simulation.

- **Demand Processing**
  This step represents the main part of the framework whose demands are processed by specifically moving vessels around within the corresponding subgraphs. All previously initialized objects interact with each other based on defined procedures.

- **Results and Statistics**
  The last part of the framework focuses on the results and statistics of the performed simulation. During this part, the simulation results are read out into JSON-files in order to access the simulation data.

The main part is contained in a loop over time, i.e., the demands are processed in discrete time steps. Currently, steps of 1 h are implemented. The simulation runs through the loop until a specific end time for the simulation is reached. The general structure of the framework is illustrated in Figure 4.3.

Based on the modular structure of the simulation framework, each part is further divided into subcomponents which were developed almost individually.

4.3.2 Initialization

The individual subcomponents of the initialization are subsequently discussed in further detail. Generally, most of the individual steps consider objects of different implemented classes, e.g. preferences, demands, vessels, etc.

Define Graphs

The generated graphs were read in from the corresponding DOT-files. For each vessel class, two sets of three subgraphs were defined in the framework. Each one of the two sets includes the same three subgraphs, namely the total subgraph of the vessel class plus the laden Sea and ballast Sea subgraph (see Section 4.2.3).

The first set of three graphs per vessel class represents the current situation of the transportation network by applying the current canal waiting times (updated once per discrete time step). It is used for generally moving vessels around and for the vessel assignment.

The other set of graphs is used for the congestion estimation at canals during path assignment. In this case, the waiting times at canals are constantly estimated and updated (multiple times per discrete time step).

Set Preferences

In this step, the preference for the simulation is set by defining the corresponding parameters \((\alpha, \beta, \gamma)\). Based on that, the general weights of all arcs are calculated for every graph, as defined in Section 3.4. The preference for a specific simulation always remains unchanged during the entire simulation.

Generally, preferences are implemented as objects of the class `Preference`, including the parameters \((\alpha, \beta, \gamma)\).
Set Time Variables

The time related variables are set during this step. This includes the start time and end time of the simulation, as well as the applied discrete time step. Moreover, the time interval for the release of new demands is set.

Generate Demands

The demand origins and destinations are generated by a probability distribution which is based on Signal’s voyage data.

Demand origins can only be represented by the Land nodes of ports at which already Load events happened in the past. The same holds for demand destinations and Discharge events accordingly. Based on the number of Load and Discharge events that happened at a specific port in the past, the corresponding Land node is going more or less likely to be a demand origin or destination. This results in the mentioned probability distribution.

As already mentioned in Section 3.3.2 during the demand generation four dates in the timeline of the demands are set, namely the dates of release, deadline for vessel and path assignment, laycan start and laycan end. The laycan starts of the demands are determined by a uniform distribution over the time interval during which new demands are released (set in step before). The other three dates are deterministically set based on the corresponding laycan start.

Since the demand directly implies the vessel class for an assigned vessel, the specific numbers of generated VLCC, Suezmax and Aframax demands are given as input to the framework.

Including all mentioned properties and quantities, demands are generally implemented as objects of the class Demand. Therefore, all generated demands are implemented as vector of Demand objects.

Since demands are distinguished according to their state of completion, all demands are initially assumed to be unreleased.

Generate Vessels

Vessels are generated generally, meaning vessels of the same class are identical besides of their identification number (IMO number). The number of VLCC, Suezmax and Aframax vessels that have to be generated is given as input to the framework, as done similarly for demands.

The initial position of the vessels is determined by the same probability distribution as the demand origins. Initially, a vessel is assumed to be available (not assigned) and the current position of the vessel coincides with its initial one.

Vessels with all their properties are implemented as objects of the class Simulation_Ship. Therefore, the generated vessels are listed within a vector of Simulation_Ship objects.

Generate Reallocation Data

In case a vessel completes a demand but is not already assigned to a new demand, it will wait at the current node for a specific time. If still no new demand is found, the vessel reallocates itself to a position where the probability for a new demand is higher.

The reallocation destinations are determined by the same probability distribution as the demand origins. This way, the vessels are redistributed to potential demand origins, i.e., the vessels are more likely assigned to new demands. Moreover, the reallocated vessels, after completing all the
demands, are similarly distributed as they were initially (due to the same applied probability distribution for initial positions and reallocation destinations).

The waiting times before reallocation are uniformly distributed between a minimal and maximal waiting time, e.g. between 2 to 7 days.

**Set Canals**

The canals are already implemented in all applied graphs (six graphs per vessel class). In order to facilitate their access for an update of the corresponding waiting times due to congestion, the Travel arcs of canals were specifically denoted.

Additionally, the congestion estimator is set up by creating vectors in which the estimated arrival times of the vessels are recorded (see more details in Section 4.3.5). This initialization step represents the last one before the actual simulation procedure is started by entering the for-loop over time.

With respect to the entire initialization, the voyage data is very important for demand generation as well as the vessel generation and reallocation. In order to compare multiple simulations for different preferences, the initialization should be similar for all simulations. Therefore, all probabilistic procedures were placed into the initialization at the beginning of the framework. This is the reason why demands, vessels as well as the reallocation data for vessels are not generated stepwise during the simulation.

In order to run multiple simulations, all generated demand, vessel and reallocation data is initially read out to separate JSON-files. These files serve then as input data sets for all subsequent simulations.

**4.3.3 Processing Demands (Main Part)**

As already mentioned in Section 3.3.2, the general simulation procedure is structured based on the perspective of demands. According to that, demands are distinguished depending on their state of completion. While demands are strictly processed based on the schematic concept discussed in Section 3.3.2, vessels are processed within this procedure for demands. For example, whenever a vessel is assigned to a new demand in combination with a specific travel path, the vessel is processed and subsequently updated when required. The individual steps for processing demands (and simultaneously also vessels) are further discussed below.

**Consider Congestion**

The current vessels at the canals (Suez and Panama Canal) are counted in order to determine the current corresponding waiting times. The arc weights of the corresponding arcs are then updated for the set of graphs that represent the current situation of the maritime transportation network. This procedure is performed for each discrete time step.

The set of graphs, representing the current congestion situation, is always updated at the beginning of a new iteration over time (once per discrete time step).

On the other hand, the set of graphs, representing an estimated congestion situation in the future, is applied and therefore updated whenever a specific “shortest” path for a vessel has to be evaluated. 'Shortest' in this regard means the path of smallest general weight. As a result, for this specific set of graphs (applied for congestion estimation and path assignment), the arc weights for the corresponding Travel arcs of the canals are usually updated multiple times per discrete time step.
4.3. Implementation of the Simulation Framework

Process Demands unreleased

This represents the first step of actually processing demands according to their state of completion. Unreleased demands, that have not been published up to this moment, are presented to the vessel owners. Once this step has been done for a demand, it is considered as released.

Different states of completion for demands are implemented as individual vectors of Demand pointers. These Demand pointers refer to the actual generated demands which are stored in a separate Demand vector (mentioned during initialization). These pointers are then moved between the different vectors, representing demand processing from one state of completion to the next one. This way, the demands can be processed and modified without changing the structure or size of the original Demand vector.

The same concept was implemented for vessels, where two different states are distinguished: available and assigned vessels. Based on that, there are two vectors of pointers to Simulation_Ship objects stored in the original Simulation_Ship vector.

Process Demands released

This step includes major tasks such as vessel and path assignment. At first, all potential vessels are evaluated which lie within a certain range around the demand origin. These include available vessels as well as already assigned vessels which are currently completing a demand (on their way to a demand destination). Only vessels are considered that could potentially reach the demand origin on time, i.e., before the laycan end.

If there are no potential vessels around, also vessels are considered that potentially arrive after the laycan end (too late).

Whether potential vessels could be found or not, leads to different further scenarios. In case, absolutely no potential vessels exist at a specific point in time, the procedure is repeated for the following time steps during a certain period of time (e.g. for 1 week). If still no potential vessel could be found, the demand is assumed to be unfeasible. Otherwise, as soon as potential vessels are located, the procedure continues.

Among all potentially assignable vessels, the closest one is selected, i.e., the vessel which has the path of smallest general weight to the demand origin. In order to evaluate the shortest path for each potential vessel to the demand origin (shortest referring to smallest in general weight), Dijkstra’s algorithm is applied. This specific algorithm is already predefined by a method of the BGL.

It should be mentioned that the vessel selection bases on the current congestion situations at the canals. In this first iteration of the framework, the congestion estimation is only considered for path assignments.

Once a specific vessel has been selected, it is assigned to the corresponding demand. As a next step, the path to the demand origin (if not already there) as well as the path from the demand origin to the demand destination have to be determined.

The selection of a specific path is strongly dependent on the applied preference of the simulation. As discussed in Section 3.4.2, it is the path between two specific nodes with the lowest general weight which is selected and assigned. This holds for all paths (path to demand origin, path to demand destination and path for reallocation). Additionally for the path to the demand origin, it has to be made sure that the vessel arrives before the laycan end (see remark in Section 3.4.2). For different preferences, the values of the general weights differ and therefore the selected paths may change. In this regard, selecting the "shortest" path has a more general meaning.
Chapter 4. Methodology

The path of minimal general weight was found by Dijkstra’s algorithm. As a remark, we additionally applied a method based on the A* algorithm which also worked. However, due to the calculation of a required heuristic and due to a more complex predefined method structure in the BGL, the method based on Dijkstra’s algorithm was preferred (simpler implementation at same run time).

During the process of path selection and assignment, the current congestion at canals is considered as well as the potential congestion at any future point in time. This was achieved by an implemented congestion estimator. The estimator is further explained in Section 4.3.5. Furthermore, we will see the effect of the estimator on simulation results in Section 5.3.

After assigning vessel and paths to a demand, different scenarios are distinguished depending on the status and location of the assigned vessel, as seen in Section 3.3.2. Based on the scenario, the pointer to the corresponding Demand is moved from the current vector (representing unassigned demands) to another one (representing the subsequent state of completion).

**Process Demands assigned, on hold**

Assigned demands are on hold when their assigned vessels are currently finishing another demand. In this state of completion, the demands are not specifically processed, they basically just wait until the vessel is available for service. Once the vessel has completed the previous demand, it either moves to the origin or, in case it is already there, directly to the destination of the new demand. The pointer to the corresponding Demand is updated accordingly.

**Process Demands assigned, to origin**

During this state of completion, the assigned vessel to a demand is traveling to the corresponding demand origin. The simulation subsequently goes through each arc of the assigned path until the demand origin is reached. In case the vessel arrives before the laycan start, it waits, otherwise it will start its path to the corresponding demand destination.

**Process Demands assigned, to destination**

In this state, the assigned vessel to a demand is traveling from the demand origin to the demand destination. The implementation of moving the vessels within the graphs, based on given paths, is similar to the case of vessels traveling to the demand origin.

**Process Available Vessels**

In case vessels are not assigned to a new demand after completion of a previous demand, they reallocate after a certain time. The reallocation destinations were already generated during initialization (see Section 4.3.2). Given the current node of the vessel and the planned reallocation destination, the path with the smallest total general weight is assigned for reallocation. This process is similar to the other path assignments.

Once the path for reallocation is assigned, the vessel travels stepwise along the individual arcs of the path until it reaches the new destination. The implementation is similar to the other travel procedures, i.e., traveling to a demand origin or to a demand destination.

**Get Status Update**

After a certain time during simulation (e.g. all six months), a status update on the demands and vessels is read out. As an example, it is denoted how many demands are already completed at this specific point in time.
4.3.4 Results and Statistics

Once the simulation over time is completed, all the required data regarding vessels, demands and canals, etc. is read out. Generally, the simulation data is written into multiple JSON-files.

Get Demand Data

For each individual demand the general properties are noted. In addition, the overall resources required for servicing the specific demand are indicated. More precisely, it is considered how many resources are required for the empty vessel to arrive at the demand origin as well as the required resources for shipping the cargo from the demand origin to the demand destination. Furthermore, the demand data also contains the pool of all potential vessels from which the assigned vessel has been selected.

Get Vessel Data

For each individual vessel, the total schedule is listed. This includes the total travel path as sequence of all visited nodes in the graph in combination with the corresponding arrival times. Moreover, the total resources required by each vessel during simulation are indicated in metrics of time, emission and cost. Moreover, all the vessel properties are given, as described in Section 3.3.1.

Get Canal Data

For both, the Suez and Panama Canal, the canal congestion over time is denoted, i.e., the number of vessels at and in the canals as well as the resulting waiting times are indicated for each discrete time step (for every hour). The applied values for the parameters $n_{crit}$, $d_{regular}$ and $q$ of the congestion function are given as well for each canal.

Get General Simulation Data

The general simulation results are summarized in this data. This includes the general inputs of the simulation (time variables, congestion parameters, number of demands and vessels, etc.). Moreover, all demands are listed with their final state of completion. Similarly, all vessels are included with their final status (assigned or not). Ideally, all demands should be completed by the end of the simulation, while all vessels should be waiting after having completed their reallocation.

Based on the required resources for each individual demand, the required resources are calculated for servicing all demands of the simulation. This includes the required resources for all paths to the demand origins as well as the corresponding contribution of all paths from the demand origins to the demand destinations.

Moreover, the total resources required by the vessels for all their travels (also including vessel reallocation) are also given, calculated based on the required resources for the individual vessels.

All these total resources are measured in the metrics of time, emission and cost. Since we consider here absolute resources which were required for transporting the total amount of cargo, they are indicated in absolute units, i.e., in $\text{h}$, $\text{t}_{\text{sulfur}}$ and [USD] correspondingly.

Analyze Data

The analysis and visualization of the simulation results is done in Python. The simulation data was read in from the corresponding JSON-files and then analyzed. The final plots and figures are illustrated in Chapter 5 on results and discussion.
4.3.5 Congestion Estimation

Congestion at canals over time is addressed by changing the arc weights of the corresponding Travel arcs of the canals. The additional delay at a canal depends on the number of vessels which are currently passing the canal or waiting in front of it. This relation is defined by the congestion function $d = f_{\text{canal}}(n)$, introduced in Section 3.2.3.

This congestion function includes three parameters, namely $n_{\text{crit}}$, $d_{\text{regular}}$ and $q$. The applied values of these parameters for the Suez and Panama Canal are denoted in Table 4.2 below, for all performed simulations.

Because dynamic congestion at canals strongly influences the selection of paths, as shown later in Chapter 5, a congestion estimator was implemented as part of the framework. This estimator predicts the congestion at the canals at any future time during simulation as accurately as possible, in order to provide more information for the path selection. The estimator was applied in several simulations.

The main structure of the estimator consists of an int vector for each canal that contains all potential arrival times of vessels at the corresponding canal. These vectors are always updated when new paths are fixed and assigned. Based on the information stored in these vectors and the number of the vessels currently in the canals, the delays are estimated for each future time step. The weights of the arcs representing the canals are updated accordingly. These updates are performed for the second set of graphs, which has specifically been implemented for this purpose (as mentioned in Section 4.3.2). Whenever a new path has to be selected, these previously updated graphs are applied. This allows to take path decisions based on more information.

However, the complete implementation of the estimator turned out to be quite complex. Above, only a very simplified explanation of the estimator is given. In order to update the vectors with the arrival times as correctly as possible, all different scenarios within the general simulation procedure have to be considered. Due to this fact, the task of implementing a decent congestion estimator quickly becomes a difficult challenge. In this framework, the estimator was iteratively improved until its function could be validated by simulations (see validation results in Section 5.3).
4.4 Performed Simulations

In order to apply and validate the implemented simulation framework, different simulations were performed. The settings for these simulations are subsequently discussed.

For all performed simulations the simulation inputs remained the same, except for the applied preference. This way, we ensured that the different performed simulations were as comparable as possible. The most important simulation parameters, serving as simulation inputs, are listed in Table 4.2 below together with their applied values. They represent only a small part of the constants and variables of the framework which had to be defined first before running the simulations.

Table 4.2: Applied values for multiple simulation parameters, representing only a small part of all implemented variables.

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of VLCC vessels</td>
<td>715</td>
<td></td>
</tr>
<tr>
<td>Number of Suezmax vessels</td>
<td>527</td>
<td></td>
</tr>
<tr>
<td>Number of Aframax vessels</td>
<td>916</td>
<td></td>
</tr>
<tr>
<td>Number of VLCC demands</td>
<td>4'290</td>
<td></td>
</tr>
<tr>
<td>Number of Suezmax demands</td>
<td>3'162</td>
<td></td>
</tr>
<tr>
<td>Number of Aframax demands</td>
<td>5'496</td>
<td></td>
</tr>
<tr>
<td>$T_{\text{start}}$</td>
<td>0</td>
<td>d</td>
</tr>
<tr>
<td>$T_{\text{end}}$</td>
<td>730</td>
<td>d</td>
</tr>
<tr>
<td>$T_{\text{start} \text{ demands}}$</td>
<td>35</td>
<td>d</td>
</tr>
<tr>
<td>$T_{\text{end} \text{ demands}}$</td>
<td>400</td>
<td>d</td>
</tr>
<tr>
<td>time step</td>
<td>1</td>
<td>h</td>
</tr>
<tr>
<td>$n_{\text{crit,Suez}}$</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>$n_{\text{crit,Panama}}$</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>$d_{\text{regular,Suez}}$</td>
<td>0</td>
<td>h</td>
</tr>
<tr>
<td>$d_{\text{regular,Panama}}$</td>
<td>0</td>
<td>h</td>
</tr>
<tr>
<td>$q_{\text{Suez}}$</td>
<td>4</td>
<td>h</td>
</tr>
<tr>
<td>$q_{\text{Panama}}$</td>
<td>6</td>
<td>h</td>
</tr>
</tbody>
</table>

| **Normalization constant for time** (for general arc weight) | 3'600 | s |
| **Normalization constant for emission** (for general arc weight) | $10^7$ | $t_{\text{timem}}/t_{\text{cargo}}$ |
| **Normalization constant for cost** (for general arc weight) | $10^3$ | USD/$t_{\text{cargo}}$ |

The detailed numbers of applied demands and vessels are given in Table 4.2. The numbers of considered vessels are based on Signal’s voyage data (from last 7 years), assuming that the average life span of a vessel is 21 years [5][10]. Based on literature and Signal’s voyage data, we estimated...
that each vessel services on average 3 demands per 6 month, i.e., needing on average 2 months per demand.

As first application of the simulation framework, different simulations based on different preferences were run. For that, the set of all different preferences $\mathcal{X}$, introduced in Figure 3.8 of section Section 3.4 as triangular surface within the total preference space $\mathbb{R}_+^3$, was sampled in regular steps of 0.1. This led to 66 evenly distributed preferences $x = (\alpha, \beta, \gamma)$, as depicted in Figure 4.4. For each of these different preferences, an individual simulation over time was run with the same simulation inputs, as mentioned above. Based on the resulting simulation data as output, different aspects in crude oil shipping were analyzed.

![Preference Space](image)

Figure 4.4: Preference space with all applied preferences which were evenly sampled in steps of 0.1 (indicated in blue).

As a first aspect, the dynamic congestion at the canals was studied by looking at the corresponding waiting times over time. For each performed simulation based on a different preference, the waiting times at the canals can be plotted over time (see Figure 5.1 in Section 5.2.1).

Another interesting consideration refers to the transport efficiency of the system for different preferences. The total resources required for servicing all demands during a simulation are evaluated for every simulation based on a different preference. We focused on the total resources required for all vessel paths from the demand origins to the corresponding destinations. This way, the influence of the preference on the simulation outcome is most noticeable as demand origins and demand destinations were the same for all performed simulations. Therefore, the paths only change because of different applied preferences.

Absolute resources were considered for transporting the total amount of total cargo. Therefore, they were indicated in metrics of time, emission and cost with absolute units, i.e., in [h], [t\text{sulfur}] and [USD] correspondingly.

Consequently, the relation between applied preference and required resources represents a map from the preference space to the resource space. In contrast to the resource space for arc weights $F_R$, introduced in Section 3.2.2, the resource space applied here is not normalized by tons of transported cargo. While the concept is similar to the one for arc weights, the units of these spaces are different. A more intuitive illustration of this consideration is given later in Figure 5.2 of Section 5.2.2.
Based on the evaluation of the simulation data, specific trends between different efficiency metrics are observed and statements about the efficiency of different preferences made. More details on that are given in the results and discussion (see Chapter 5).

As a final application of the framework, the congestion estimator was validated by performing simulations with different implementations of the estimator. The corresponding results are shown in Section 5.3.
Chapter 5

Results and Discussion

In this chapter, the focus is set on the application of the simulation framework. The results of the performed simulations are presented and discussed. At first, general observations from the simulations are given. Afterwards, the influence of the applied preference on canal congestion is studied. With help of these insights about canal congestion, the dependence of the simulation results on the applied preference is analyzed. The trends between the total required resources for servicing all demands are discussed, referring to the efficiency of different simulation strategies. At the end of this chapter, a validation of the applied congestion estimator is given.

5.1 General Observations and Results

For the application and validation of the simulation framework, multiple dynamic simulations for different scenarios have been performed. It should be mentioned in advance that the presented and discussed results in this thesis only represent a small fraction of the potential applications that could be covered by the framework. The main workload of this project focused on the model and implementation aspect of the framework, in order to enable first simulations. Therefore, the performed simulations and the analysis of the resulting data only represent a smaller part of this work.

Despite the many applied simulation parameters, an initial setting has been found that works for all simulations based on the different applied preferences. As also mentioned in the outlook (see Chapter 7), the tuning of all the framework parameters and the evaluation of their influences on the simulation results represent an interesting task for future works.

Especially the amount of applied demands and vessels per vessel class highly affect the entire simulation outcome. While the fleet of vessels represents the total supply of the system, the sum of all individual demands denotes to the total demand of the simulated crude oil shipping market.

When we consider far too many demands with respect to the applied vessels, the number of unassigned and later unfeasible demands rapidly increases over the duration of the simulation. Due to the lack of assignable vessels, all uncompleted demands accumulate and overload the system. If a considerably too low number of demands is applied with respect to the number of vessels, all demands will be completed, but many vessels remain unemployed during long periods of the simulation. Moreover, the dynamic interactions among vessels do not come into play in this case, as most of them are just waiting for a new demand.

Referring to the run time of the simulations, it took on average 22 min to perform a single simulation (over totally 2 years with 1 year time period of releasing new demands, see Table 12 of Section 4.4). The main simulation process required about 5 min only, while the remaining time
was used to set up the graphs and to read out the resulting simulation data. Certainly, these run times could be further reduced, as for example the read-out method could specifically be refined according to the required data.

Observations and results of the specific simulations are analyzed and discussed in the subsequent sections below.

5.2 Simulations for DifferentPreferences

In order to study the influence of the applied preference on the simulation outcome, simulations for different preferences were performed, as discussed in Section 4.4 on methodology.

Before we provide considerations on transport efficiency, the dynamic congestion situation at the canals is discussed first. By analyzing the effects of the applied preference on the dynamic congestion at canals in Section 5.2.1, the observed trends and simulation results on transport efficiency can be better explained, as seen later in Section 5.2.2.

5.2.1 Congestion at Canals

Based on the simulation results, we observe that the Panama Canal is not prone to congestion for any preference. Irrespective of the applied preference, the number of vessels at the Panama Canal always remains smaller than the critical number $n_{\text{crit}}$. This results in a regular operation of the canal with a constant delay $d_{\text{regular}}$ at any time during simulation (see applied values of the parameters in Table 4.2 of Section 4.4).

The main reason for this observation can be explained by the geographical location of the demand origins and destinations. The direct paths between demand origins and destinations rarely lead through the Panama Canal. Only occasionally, vessels will pass the Panama Canal in case they have to, but their frequency is not high enough to result in a congestion. Another aspect to consider is that the Panama Canal is not passable by VLCC vessels. This leads to an even smaller number of vessels which could potentially accumulate in order to cause congestion at the canal.

Taking these considerations into account, the subsequent results and discussion about canal congestion refer only to the Suez Canal. For each of the sampled preferences an individual simulation was run (in total 66 times), leading to different congestion phenomena over time. Representatively, the congestion dynamics at the Suez Canal are illustrated in Figure 5.1 for six specific preferences (highlighted in red in Figure 5.1a). By going through each of these representative cases below, the different congestion scenarios at the Suez Canal are analyzed and discussed.

The waiting time at the Suez Canal over time for the preference $\text{time, } x = (1, 0, 0)$, is depicted in Figure 5.1b. We observe a waiting time which little fluctuates around an average value of about 20 days during the time period where new demands are released and processed. This result is consistent with the applied model and implementation.

In case only the metric $\text{time}$ is considered, vessels just travel through a canal when they save time, otherwise they avoid it. This should lead to an equilibrium for the waiting time, representing the average travel time saved overall paths. The travel time, that a vessel could potentially save by taking a path through the Suez canal, is strongly dependent on the corresponding travel origin and destination and can reach values up to over 30 days. As an average overall paths, the time saved by the canal lies around 20 days which is in line with our simulated waiting time. It is worthwhile to mention that the simulated waiting time which dynamically balances at about 20 days is a real output of the simulation. While the number of vessels, waiting at a canal, can be tuned by the parameters of the congestion function, the resulting waiting time cannot. This result therefore conceptually validates the framework.
Figure 5.1 shows the canal congestion over time for the preference \( x = (0, 1, 0) \). As soon as new demands are released and processed, the waiting time at the Suez Canal continuously increases up to 180 days, as more and more vessels arrive at the canal. Only after no new demands are released, the congestion slowly dissolves.

In case only the metric \( \text{emission} \) is addressed, vessels will almost always prefer the path of shorter travel distance through the canal. Because the idle consumption of a vessel is significantly smaller than its consumption while moving, the emission for waiting at the canal is smaller in most cases than the potential emission for traveling the additional distance when avoiding the canal. Therefore, vessels prefer the path through the canal irrespective of the corresponding waiting time, leading to a large congestion. Conceptually based on the applied model and implementation of the framework, the specific simulation result is reasonable.

![Preference Space](image)

Figure 5.1: (a) Preference space is depicted with all sampled preferences in steps of 0.1 for which an individual simulation was performed. Six representative preferences are highlighted in red. The waiting time at the Suez Canal over the simulation time is given for (b) the pure preference \( \text{time}, x = (1, 0, 0) \), (c) the mixed preference of \( \text{time} \) and \( \text{emission} \), \( x = (0.5, 0.5, 0) \), (d) the pure preference \( \text{emission} \), \( x = (0.1, 0, 0) \), (e) the mixed preference of \( \text{time} \) and \( \text{cost} \), \( x = (0.9, 0, 0.1) \), (f) the mixed preference of \( \text{emission} \) and \( \text{cost} \), \( x = (0, 0.7, 0.3) \) and (g) the pure preference \( \text{cost} \), \( x = (0, 0, 1) \).
In Figure 5.1c, the delay at the Suez Canal is plotted over time for a mixed preference of time and emission, \( x \sim (0.5, 0.5, 0) \). In this case, the congestion dynamics look like the mixture of the ones for the pure preferences of time or emission. The waiting time initially increases, but then saturates at a value of about 80 days.

As a remark, the observed waiting times at the Suez Canal for simulations with preferences time and emission (including mixtures of them) are significantly longer than the ones experienced in reality. The main reason for that is the fact that in crude oil shipping the financial aspect, represented by the preference cost, is never really neglected. As soon as the metric cost is also addressed, the waiting times drop to realistic values, as seen for the next three examples.

Figure 5.1g illustrates the congestion situation for the preference cost, \( x \sim (0, 0, 1) \). We see that in this case there is no congestion at the Suez Canal at all. This seems reasonable by considering the fact that canals are very expensive to pass, i.e., the vessels will more likely take the path without any canal. The depicted plot does not necessarily imply that there is absolutely no vessel passing the Suez Canal, it just shows that the number of vessels is smaller than the critical number \( n_{crit} \) which describes the case for regular operation without congestion.

A mixed preference of time and cost, with \( x \sim (0.9, 0.1) \), is depicted in Figure 5.1h. Very little congestion is observed with a maximal waiting time of 16 h. Figure 5.1i similarly illustrates a mixed preference of emission and cost, with \( x \sim (0, 0.7, 0.3) \). In this case the congestion is still relatively low with a maximal delay of 32 h.

Considering the chosen parameters \( p, \alpha, \beta, \gamma \) for the mixed preferences and their corresponding simulation results, the influence of a specific metric on the general weight can be interpreted. As an example, it can be seen that the influence of the metric cost is stronger compared to the one of the metric time. However, these influences of the different metrics are not absolute, as they heavily depend on the applied normalization constants in the calculation of the general weights (as discussed in Section 3.4). Therefore, these trends have mainly the purpose to show relative trends between metrics, also simply because of the fact that the metrics remain incomparable quantities with different units.

All these considerations on canal congestion lead to the discussion of the results on transport efficiency in the following section.

### 5.2.2 Transport Efficiency

As already mentioned in Section 4.4, plotting the total required resources for servicing all demands for each of the sampled preferences represents a map between the preference space and the resource space. Figure 5.2 shows the corresponding results in the resource space in form of a scatter plot. While the three-dimensional illustration is not very explanatory at a first glance (see Figure 5.2a), the corresponding two-dimensional projections along the different coordinate axes provide better views on the trends between the metrics. The projections are presented in the Figures 5.2b to d.

In this case, we consider the total required resources for all paths from the demand origins to the corresponding demand destinations. As exactly the same demand origins and destinations were applied for all performed simulations, the differences in required resources can directly be traced back to the different applied preferences.

By looking at the corresponding two-dimensional projections, we generally distinguish trends for two different scenarios. On one hand, there is the case of high canal congestion which exhibits large waiting times at the Suez Canal. Based on the discussion in the previous Section 5.2.1, this scenario relates to preferences which do not or very little cover the metric cost only (small \( \gamma \)). For better visualization, some examples for this scenario are highlighted in red (\( \gamma = 0 \)), as seen in
Figures 5.2a to d. On the other hand, we describe the case of low canal congestion with regular canal waiting times. This scenario results from applied preferences which also considerably include the metric \textit{cost}. Similarly to the case for high congestion, a few corresponding examples for low congestion are indicated in orange (\(\gamma = 0.5\)) in Figures 5.2a to d.

Figure 5.2b shows the projection along the \textit{time} axis, illustrating the trends between \textit{emission} and \textit{cost}. We observe that the metrics \textit{emission} and \textit{cost} exhibit an inverse relation, i.e., lowering the cost leads to higher emission. This finding is consistent with the theoretical model, assuming that vessels either travel through a canal, that saves emission but increases cost, or avoid it, that increases emission but reduces cost.

The applied preference, defined by \((\alpha, \beta, \gamma)\), determines how many vessels will pass the canals. The higher the influence of \textit{emission} on the general arc weight relative to the one of \textit{cost} (represented by parameters \(\beta\) and \(\gamma\)), the more vessels will travel through canals and save emission, while cost is rising. The same holds vice versa for increasing the influence of \textit{cost} relative to the one of \textit{emission}.

Additionally, we see that the results for cases of high congestion lead to outliers in the inverse relation between \textit{emission} and \textit{cost}, while the results for all other cases almost perfectly align on a curve. The reason for that is the fact that high congestion increases the required resources for passing the corresponding canal (Suez Canal). Congestion at canals results in more required time for waiting as well as larger emission and cost due to the idle operation of the vessels during the wait. In this regard, canal congestion can lead to inefficient simulation outcomes (find more details below, when we further discuss Pareto efficiency).

![Preference Space and Resource Space](image)

Figure 5.2: (a) Total resources required for servicing all demands (from demand origins to destinations) for each sampled preference, representing a map from the preference space to the resource space. Two-dimensional projections along the axes of (b) \textit{time}, (c) \textit{emission} and (d) \textit{cost}. Representative examples are highlighted in red (high congestion, \(\gamma = 0\)) and orange (low congestion, \(\gamma = 0.5\)) with regard to congestion at the Suez Canal.
Simulations for Different Preferences

Figure 5.2c depicts the projection along the emission axis, showing the relation between time and cost. Again, different trends are observed for the cases of low and high canal congestion, respectively.

For all samples which represent the low congestion scenario, there is an inverse relation between time and cost (see representative examples highlighted in orange). In this scenario, vessels save time when passing a canal, while the cost increases (costs for passing a canal are high). Therefore, the more vessels take a path through a canal instead of avoiding it, the lower gets the total time required for servicing all demands, while the total cost increases. The specific number of vessels which pass the canal is dependent on the applied preference.

In contrast, the case for high canal congestion shows a completely different trend (see representative examples depicted in red). With increasing cost, the total required time rises as well. This observation accounts for the fact that vessels do not save time when taking a path through a canal at high congestion. The more vessels pass the canal in a high congestion situation, the higher are the totally required time and cost.

The projection along the cost axis is shown in Figure 5.2d, depicting the relation between time and emission. Similar to the previous two projections, we distinguish the cases for low and high congestion at the Suez Canal.

In the low congestion scenario, vessels save time and emission by taking a path through the canal. Therefore, we observe a direct relation between time and emission, i.e., by decreasing the total required time, the total emission is falling as well (see representative examples highlighted in orange). The more vessels travel through canals, the lower the total time and emission. The corresponding number of vessels passing a canal depends on the applied preference.

For the scenario of high congestion at the Suez Canal, we observe a different relation between time and emission. By decreasing the total required time, the total emission increases, i.e., an inverse relation is exhibited. Vessels do not save time when passing a canal at high congestion, while they are still saving emission. Therefore, the more vessels take paths through canals, the higher gets the total required time for servicing all demands, while the total emission decreases. As already mentioned in the discussion above, the applied preference implies how many vessels will take paths through canals (referring to the previous Section 5.2.1 about congestion for more details).

By going through the results of all 66 performed simulations for different preferences, we evaluated in total 10 different simulation outcomes which are not Pareto efficient. For each of these so-called Pareto dominated outcomes, there exists another outcome that requires less resources for at least one single metric without worsening the remaining other metrics. These Pareto dominated outcomes and their corresponding preferences are indicated in green in Figure 5.3.

We observe that the corresponding preferences of the Pareto dominated outcomes strongly cover the metric time, while the metric cost is not or only little considered (small $\gamma$ relative to $\alpha$). They all represent scenarios for high canal congestion which is the main reason for their inefficiency. The simulation objective of minimizing the total general weight of each path leads to locally optimal strategies for each individual vessel. Especially in the case of high canal congestion, however, this could result in globally inefficient strategies. Therefore, the simulation outcomes in form of total required resources are Pareto dominated for preferences, which predominantly focus on the metric time, while neglecting the metric cost. A redefinition of the simulation objective could potentially lead to more efficient solutions in this regard. More generally, other game theoretical considerations also pose interesting challenges for further related works (see Chapter 7 for a more detailed discussion).
Figure 5.3: Pareto dominated preferences and their corresponding simulation outcomes (as total resources required for servicing all demands from demand origins to destinations) highlighted in green.

The observed trends and discussed results provide qualitative insights about the efficiency of maritime transportation. Due to the many applied simulation parameters which can be further tuned and analyzed (see outlook), specific quantitative statements about transport efficiency are saved for future iterations of the framework.

In the following Section 5.3 as last part of results and discussion, the effect of the congestion estimator on the dynamic simulation results is analyzed and its function validated.
5.3 Validation of the Congestion Estimator for Canals

The function of the congestion estimator, whose implementation is discussed in Section 4.3.5, has been validated by comparing the results of simulations for different estimator implementations.

All corresponding simulations were performed based on the same simulation inputs. We applied the preference \( t_p \) for all simulations, \( \chi = (1, 0, 0) \). This preference may not directly represent reality as the waiting times at canals are too long, but it really serves the best to highlight the differences between specific estimator implementations.

As already discussed in Section 5.2.1 for the preference \( t_p \) we expect that vessels pass canals in case the additional waiting time at the canal is smaller than the travel time for the longer path around the canal. This results in an almost constant waiting time at the canal, representing the average time saved by passing the corresponding canal.

Figure 5.4: Dynamic waiting times at the Suez Canal resulting from simulations for preference \( t_p \) (a) with the regular congestion estimator, (b) without congestion estimator, (c) with a conservative estimator (waiting time estimated twice as high as regular), (d) with a very conservative estimator (waiting time estimated three times higher).
Exactly this has been observed in Figure 5.1b of Section 5.2.1, here depicted again as Figure 5.4a. It shows the dynamic waiting time at the Suez Canal for the regular implementation of the congestion estimator. The average waiting time during the interval when new demands were released lies at about 20 days which refers to the average time saved by passing the Suez Canal.

In contrast to that, Figure 5.4b shows the simulation result for the case when no congestion estimator was implemented. All paths were selected based on the congestion situation at the corresponding time of the path assignment. In this case, we observe a periodic behavior where the canal waiting time strongly fluctuates between 0 and up to 40 days. This dynamic behavior is caused by the fact that vessels need a specific amount of time in order to arrive at a canal. Therefore, there is always a certain delay between the path selection (based on the corresponding congestion situation at that specific time) and the later arrival of the vessel at the canal (presenting a different congestion situation). This delay can reach weeks or even months due to the relatively long process and travel times in maritime shipping compared to other transportation modes, e.g. vehicles or trains. Thus, the system without estimator always overcompensates the occurring trends with a delay in time, leading to the observed periodic behavior.

The comparison between these simulation results for different implementations clearly illustrates the importance of the implemented congestion estimator and also validates its function. Moreover, we also performed two additional simulations with more conservative congestion estimators.

Figure 5.4c depicts the waiting time at the Suez Canal over time for an implemented estimator which estimates the canal waiting times twice as high as they actually are. More precisely, for the final calculation of the estimated waiting time at the canal during path selection, a value twice as high as the one of the congestion function \( d = f_{\text{canal}}(n) \) was applied. The dynamic behavior of the canal waiting time looks similar to the one for the regular estimator, except for the fact that the average value dropped almost by a factor of 2 to approximately 10 days.

Similarly, Figure 5.4d illustrates the dynamic waiting time at the Suez Canal for an even more conservative estimator which estimates the canal waiting times three times higher as they actually are. In this case, we also observe a dynamic behavior similar to the one for the regular estimator, shown in Figure 5.4a. Correspondingly, the average value of the canal waiting time dropped almost by a factor of 3 to approximately 7 days.

It should be highlighted that the average waiting times for the different implementations of the estimators are real dynamic outcomes of the simulations and not artificially tuned by some specific parameters.

The waiting time equilibrates around a value which represents the average saved travel time when vessels take the shorter path through a canal. In case the waiting time is expected higher than it actually should be (by more conservative estimator), as done in the last two simulations, the average value of the simulated waiting time should adjust itself correspondingly to lower values. Exactly this could be observed, as shown in Figures 5.4c and d.

Generally, all mentioned results demonstrate the importance of the congestion estimator and its direct effect on the simulation outcome. Moreover, they validate the implementation of the estimator as the corresponding results turned out as expected.

Last but not least, as already mentioned in Section 5.1, all results shown in Chapter 5 only represent a small fraction of the potential application of this framework. Due to its generality and modularity, the framework really provides the basis to run many more simulations. For more details on potential applications and refinements, we refer to the outlook in Chapter 7.
5.3. Validation of the Congestion Estimator for Canals
Chapter 6

Conclusion

This chapter summarizes the most important observations and findings of this project in a final conclusion. At first, general conclusions are made regarding the approach and workflow of this thesis. Subsequently, the main results of the project are summarized and put into context. At the end, some final considerations complete the chapter as preparation for the outlook.

Based on the literature review of previous works in Section 1.3, we realized that there was only little work done that covers the crude oil shipping market in its entirety.

A large part of the reviewed papers only address a very specific problem of the maritime transportation system. These works are usually strongly based on specific data and therefore not generally applicable for other scenarios. In contrast to that, there are also works that cover more general aspects. However, they provide mainly analytical concepts and therefore do not actually solve a real problem.

Generally, most of the papers only discuss static aspects of maritime transportation, e.g. by considering different routing or path planning problems. In many cases, the main focus is then set on the mathematical approach for a specific optimization problem, for which the crude oil shipping market serves as an interesting application example. Thereby, the modeling aspect is only secondary.

Taking this into consideration, the goal of this project was to create a very general simulation framework based on real data from Signal in order to analyze, model, simulate and ideally optimize the crude oil shipping market.

The collaboration with Signal really helped us to better understand the challenges of crude oil shipping by providing access to their voyage data, online platform and API. Based on that, we made sure that our work addressed a real problem and started our work from there. The data was very important for creating the theoretical model for the framework and its implementation.

In parallel, a lot of effort was put into the simulation framework in order to make it as general and applicable as possible. Moreover, by running simulations over time, also dynamic aspects of the transportation system could be considered, e.g. congestion phenomena at canals.

Taking all these aspects into consideration, the project was initially challenging to start with, as there was not much reference available to our approach. Especially, the application of the data and the implementation of the graphs took a lot of time, including all the different program setups.

Once these steps were completed, the general workflow became more efficient in form of an iterative process of trial and error. We went through many iterations, as we stepwise improved the
model, implementation and application of the simulation framework. This iterative process was maintained until the final simulations had been run and their results been analyzed.

It should be mentioned at this point that the general goal of completely analyzing, modeling, simulating and optimizing the crude oil shipping market turned out to exceed the scope of a single Master’s thesis. Each individual aspect of the simulation framework, including the multiple sub-components, poses a challenge of its own and could be further refined.

Therefore, this thesis describes a first iteration in order to achieve the higher goal. It serves as a foundation for future works, as it covers the main aspects of theory, implementation and application:

• **Theory**

  We developed a general model for the crude oil shipping market in form of a simulation framework. This includes different aspects about efficiency, graphs as transportation network, demands, vessels, simulation procedures, preferences, etc. Moreover, it also allows the consideration of dynamic phenomena, e.g. congestion at canals. Approaches to describe and simulate AV-enabled mobility systems served as an inspiration for this model [3, 54, 55].

  The model offers great modularity and flexibility in its application and provides a lot of potential for further refinements. Last but not least, it is consistent with the voyage data of Signal and helps thus to solve a real problem.

• **Implementation**

  The simulation framework is fully implemented. During implementation, we focused on its modularity and its usability for later applications.

  An important part of the implementation is represented by the graphs. Multiple steps were required to actually implement and generate these graphs, e.g. read-in of voyage data, evaluation of ports, calculation of distances via API, calculation of arc weights, implementation of graph structure via BGL methods, consideration of canals, read-out to .dot-files, etc.

  The actual simulation framework with its simulation procedures represents another major part of the implementation. It contains many individual components that have been implemented, e.g. vessel and path assignment, vessel reallocation, congestion estimation, etc.

  All implemented framework components serve as a foundation for further refinements in future iterations.

• **Application**

  The application and validation of the simulation framework is represented by multiple simulations that were run and their results.

  Within the results, we observe interesting trends for the chosen metrics time, emission and cost, indicating the required resources during simulation. The results for these metrics show a strong dependence on the corresponding congestion at the Suez Canal, so that two different cases could be distinguished: low congestion (related to the preferences cost) and high congestion (related to preference time and emission).

  As a result, we find that some specific preferences lead to Pareto inefficient outcomes. Specific
implemented strategies may be locally optimal for each individual vessel but can lead in their sum to a generally inefficient result.

Furthermore, the implemented congestion estimator could be validated, as the results show conceptually reasonable trends, based on the theoretical model. Simulations exhibit significant differences in their outcome depending on the applied congestion estimator during path assignment.

For the sake of completeness, it should be mentioned that the performed simulations with their results only represent a very small part of all potential applications that could be performed with the current simulation framework. The main reason for that was the scope of the Master’s thesis with a usual timeline of 6 months (expanded for this project to totally over 9 months).

Conclusively, the simulation framework in its current iteration serves as a foundation for future work. Potential refinements and extensions for further iterations are described in the following outlook.
Chapter 7

Outlook

This final chapter presents the outlook for this project. At first, the potential updates regarding graphs are discussed. Subsequently, we describe potential refinements for the entire simulation framework with respect to theory and implementation. Then, the application of the framework is further outlined by describing interesting simulations that could additionally be realized. At the end, some final statements about the project in general are given.

Despite the interesting insights gained by this work on the shipping of crude oil, there remain open questions that need to be addressed. Therefore, it is important to discuss potential refinements, updates and extensions for this project. This holds especially in this case, where we have a very general simulation framework in its first iteration.

The graphs, introduced in Section 3.2, describe an important part of the current framework, as they represent the maritime transportation network. Regarding these graphs, there are potential refinements that may be done in the future.

First, the graphs could be updated based on corrected voyage data. As already mentioned in Section 4.2.2, the current voyage data of Signal shows minor inconsistencies in the description of ports and geosets. As soon as Signal provides a corrected version of the voyage data, the graphs can be updated accordingly.

Furthermore, additional Travel arcs between two ports could be added in order to consider alternative direct paths. Travel arcs represent the direct path between two ports avoiding canals. In the current implementation for Travel arcs based on the Signal API, we applied the corresponding path of shortest distance without canal. Therefore, additional parallel Travel arcs could be implemented in order to account for other direct paths between two ports without canal. This would enable even more travel alternatives.

Referring to the arc weights of the graphs, the implementation of a non-constant vessel speed, which depends on the local supply and demand of the system, poses an interesting task for a future iteration of the framework.

Regarding the model and the implementation of the entire simulation framework, there are other potential refinements which might be later realized.

In Section 3.3.1, it is mentioned that a demand only considers one Load and Discharge event, assuming full shiploads. The cargo size directly implies the required vessel class. Therefore, as an extension of the current framework, multiple Load and Discharge events per demand may be considered. This would allow to include continuous cargo sizes. Consequently, vessels of different
vessel classes would be qualified to service a specific demand, i.e., they would be compared with each other. For that, the arc weights are already normalized per ton of transported cargo.

Additionally, we could include transshipment in the simulation framework. Based on the structure of the graphs, the concept of transshipment would be applicable, as the subgraphs of different vessel classes are combined in the corresponding Land nodes, see Section 3.2.1.

Another important aspect that could be specifically worked on is the tuning of the simulation parameters. Each parameter within the framework could be individually varied, while analyzing its influence on the simulations. By that, the entire system of crude oil shipping could be better understood. As an example among many others, the ratio between demands and vessels for a specific vessel class could be further studied, by analyzing the relation between supply and demand.

Referring to the discussion in Section 3.1, it may also be possible to include profit within the framework, in addition to cost. For that, we would have to consider an additional financial structure, due to fact that profits are dynamically dependent on supply and demand in a specific region. Based on supply and demand, the charter rates for vessels could be dynamically determined in order to evaluate the corresponding profits.

Generally, each component of the simulation framework could be individually refined. As mentioned in the conclusion, this project with its current version of the framework serves as foundation for future iterations. Due to the modular structure of the implementation, one should be able to develop the different components of the framework almost individually. To mention some examples, these components include demand initialization, vessel initialization and assignment, path planning and assignment, congestion estimation and vessel reallocation, etc.

It is worthwhile to mention that there is a lot of potential in the application of the simulation framework. The performed simulations and their results, shown in Chapter 5, represent only a very small part of all the potential applications that could be done by the framework.

As an example, we could recreate the scenario of a vessel blocking the Suez Canal for a certain amount of time. By simulating this scenario, the corresponding consequences of this blockage could be analyzed and compared with the real case (happened in March 2021, see [16, 17]), e.g. by looking into the congestion at the Suez Canal. The results from such a simulation could serve as an additional validation of the framework, as a further reference for the parameter tuning or as a tool to better predict and anticipate similar events in the future. Similarly, many other potential scenarios could be simulated, e.g. by imposing specific geopolitical constraints. This directly refers to the recent question how potential quarantine restrictions between specific countries could impact the crude oil shipping market (based on article [66]).

A potential solution to overcome Pareto dominated simulation outcomes (predominantly based on preference time, while neglecting cost, see Section 5.2.2) could represent a redefinition of the simulation objective. The application of specific game theoretical concepts, e.g. competition or cooperation between specific vessel owners, could help to develop more efficient simulation strategies in the future, referring to [67, 68].

The current simulation framework and its future iterations will help to better understand the crude oil shipping market. As general long-term goal, the framework should ideally tell us how to optimize the current market or at least show us how to better operate within the market.

Last but not least, the collaboration between Signal and the Institute for Dynamic Systems and Control (IDSC) is key for pursuing this general goal and therefore should be maintained or ideally further extended.
Bibliography


