UiT The Arctic University of Norway

Faculty of Science and Technology
Department of Computer Science

Evaluating Continuous End-to-End Communication at Sea with Multi-Hop MANET Routing, Using AIS Data

Øyvind Arne Moen Nohr

INF-3981 Master's Thesis in Computer Science

Supervisors

Main supervisor:	Dagenborg, Håvard J.	UiT The Arctic University of Norway, Faculty of Science and Technology, Department of Computer Science
Co-supervisor:	Bjørndalen, John Markus	UiT The Arctic University of Norway, Faculty of Science and Technology, Department of Computer Science

To you, whom I have not yet met

“Your best and wisest refuge from all troubles is in your science”
–Ada Lovelace

Abstract

The marine sector has unique and challenging problems supporting high-bandwidth, low-latency internet connectivity, often unavailable or only available through satellite services. Multi-hop MANETs that utilise low-cost commodity hardware potentially offer a cost-effective solution compared to satellite services but come with their own limitations.

This thesis is motivated by the need for reliable and affordable communication at sea, especially in areas where satellite coverage is compromised by geographical features, high traffic, or downtime. By studying a fleet of more than *14,500* boats spread across Norway's EEZ using AIS positional data from the NCA, it aims to assess the viability of SANET as an alternative to satellite communication systems for maritime vessels operating within the EEZ.

This thesis proposes Aktan: a novel simulator that facilitates connectivity analysis of this fleet by modelling it as a graph. It does not extend to hardware specifications or physical-level simulation. Aktan uses a combination of tabular, key-value storage, and vantage point trees to support date-based graph queries. Results from this analysis provide insights into the practical considerations of SANET deployment in the EEZ, contributing to the broader discourse on improving maritime communication systems. We identify challenges in implementing and deploying an overlay MANET in this region while identifying key geographical areas of interest. It is shown that connectivity is low overall and has temporal dependencies. We identify key areas of interest for deploying infrastructure to support multi-hop MANETs in this region.

The study shows that MANETs alone will not be viable in this region without improvements in fleet management, infrastructure placement, technological breakthroughs, or combining technologies.

Acknowledgements

Addressing real-world challenges through computer science to enhance the quality of life in any shape or form is profoundly fulfilling to me. I thank my supervisors, Håvard D. Johansen and John Markus Bjørndalen, for allowing me to dive into new fields, supporting my exploration and guiding me through my mistakes. Thank you, Robert Pettersen, for being my mentor during the first years at IFI, always there to help me sort out my thoughts. Dag Johansen, thank you for allowing me to stay in the research group and being a big inspiration to me. John Markus, your faith can move mountains, thank you.

I want to thank the wonderful people in the Cyber Security Group who made my time at UiT unforgettable. In particular, I would like to thank Marius J. Ingebrigtsen. A friend like you is a rare treasure.

I am deeply grateful to the entire staff of the Department of Computer Science at UiT. Your open doors and the fantastic learning environment you provide have been instrumental in my academic journey. Jan Fuglesteg, you are invaluable.

Edvard and Helene, I am unsure where Eir(my dog) and I would be today without you.

Lastly, I want to express my deepest gratitude to my partner. Your patience and support have been my rock through what felt like an eternity apart. Thank you for spending countless hours reading through my master's, for pushing me out of bed when I needed it, and for always believing in me. I am looking forward to what is to come, knowing that I have you by my side.

This thesis is written by me, with great guidance from my supervisors and help from some tools:

Grammarly has been used throughout the writing process for spellchecking and paraphrasing.
Co-pilot has been used as an aid in plotting the graphs, as I keep failing to memorize the vast Matplotlib library.

ChatGPT/Large language models outside of the abovementioned have only been used to explore keywords. No prompt has been used to generate any of the text in this thesis.

Contents

List of Figures

List of Tables

Definitions

List of Abbreviations

ACM Association for Computing Machinery

AIS Automatic Identification System

AMSL Above Mean Sea Level

ANET Ad Hoc Networks

AODV Ad Hoc On-demand Distance Vector

BFS Breadth-first search

BTB Boat-to-Boat

COLREGS Convention on the International Regulations for Preventing Collisions at Sea

CPS Cyber Physical Systems

CSG Cyber Security Group

CSV Comma-Separated Values

DAO Distributed Arctic Observatory

DFS Depth-first search

DV Distance Vector

EEZ Exclusive Economic Zone

FMGT Forsvarets militærgeografiske tjeneste

GPS Global positioning system

IMO International Maritime Organization

IUU Illegal, Unreported and Unregulated

LOS Line Of Sight

LSR Link State Routing

MANET Mobile Ad Hoc Networks

MMSI Maritime Mobile Service Identity

MPI Message Passing Interface

NCA Norwegian Coastal Administration

NN Nearest Neighbour

NOU Norwegian Official Report

NS3 Network Simulator 3

OCN Ocean Communication Networks

OLSR Optimized Link State Routing

OMPI Open Message Passing Interface

OOW Officer of the Watch

PTP Point-to-Point

SANET Sea Ad Hoc Networks

SDN Software Defined Network

SHF Super High Frequency

UHF Ultra High Frequency

UiT UiT, The Arctic University of Norway

USD US Dollar

VANET Vehicular Ad Hoc Network

VHF Very High Frequency

VPN Virtual Private Network

VPTree Vantage Point Tree

WGS84 The World Geodetic System 84

1 Introduction

Fish is considered one of the world's most important food sources and is estimated to make up about 17% of the global protein production [1]. Commercial fishing is a multi-billion dollar business [2], and with this significant potential for profit, sustainability, and economic challenges related to Illegal, Unreported and Unregulated (IUU) fishing has dramatically increased. UN's Food and Agriculture Organization estimates that IUU fishing is responsible for a total loss of 11-22 tonnes fish with an estimated value of *10-23* billion US Dollar (USD) every year [3].

Technological solutions have been proposed to address the issues of IUU, and the Norwegian government plans to deploy video and sensor surveillance on board the vessels in the EEZ [4]. The Norwegian Official Report (NOU)19:21 [4] outlines five viable technologies for achieving communication of data in the context of monitoring and surveillance at sea: *drones, blockchain, cloud services, mobile networks, and satellite*. Blockchain, drones, and satellites are intended for tracking, monitoring, and surveillance. Cloud services are seen as the architectural choice for achieving scalability, availability, and security.

Access to reasonable communication solutions is a prerequisite for many systems proposed to achieve surveillance and control of fishing activities in the EEZ. Fishing boats often find themselves far off the coast, limiting access to reliable, low-cost communication solutions.

While recently deployed low-orbit satellites have made it possible to deliver

high-bandwidth networks in areas without fixed-cable infrastructure, they come at a high cost, with unresolved bandwidth and latency problems in densely populated areas or at particular times at day [5, 6]. While reliable and low-cost for the end-user, cellular or wireless local area networks on shore do not have coverage far into the open sea, and infrastructure is difficult to deploy out in the ocean. The concern is that surveillance and monitoring systems may be deployed without fully considering how they will disseminate the large amount of data they produce.

Mobile Ad Hoc Networks (MANET)s have been studied for decades as an alternative to satellite- or fixed infrastructure-communication. Multiple standardisations such as AIS (cooperative and non-cooperative) and technological advances have created new possibilities for realising communication between nodes in these networks.

The active Norwegian fishing fleet makes up about 5,700 vessels [7] that sail from Skagerrak in the south to the Barents Sea in the north as seen in Figure: 1.1.

For this study, AIS-positional data for over 14,500 marine crafts have been traced over three weeks. No emphasis has been put on the type of vessel, activity, or AIS broadcasting capabilities.

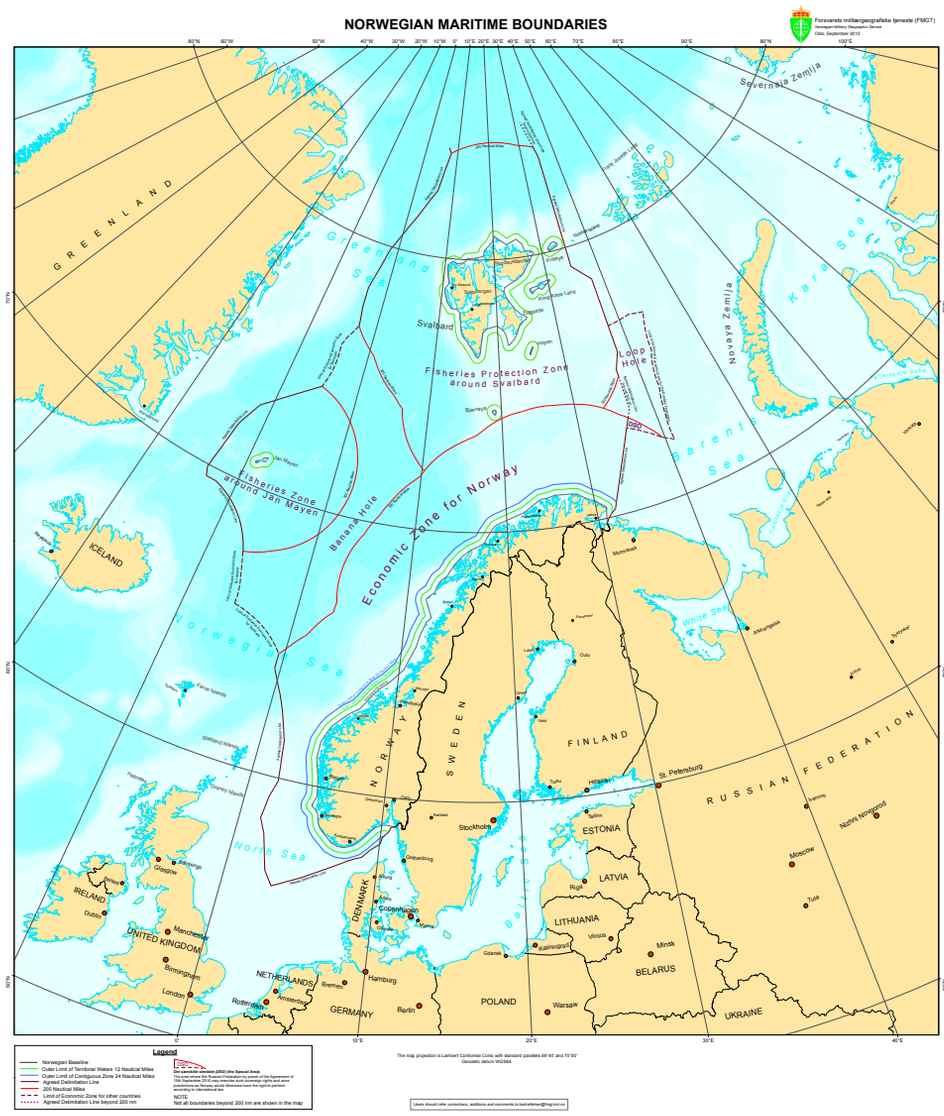

Figure 1.1: Norwegian economic zone and surrounding areas for fishing.
Source: FMGT

1.1 Problem Definition

This thesis explores multi-hop MANET in ship-to-ship and ship-to-shore communication by evaluating the viability of deploying an overlay network in the EEZ using low-cost commodity hardware to achieve vessel connectivity to the global network on shore. Our thesis is that:

> Multi-hop MANET routing enables continuous connectivity to on-shore infrastructure for boats in Norway's exclusive economic zone

1.2 Method

The Association for Computing Machinery (ACM)'s Task Force on The Core of Computer Science [8] defines three paradigms for the discipline of computing; *theory*, *abstraction (modelling)*, and *design*.

The paradigm of **Theory** consists of four stages rooted in mathematics and is the steps taken in the development of a *coherent*, valid theory: Characterise objects of study and hypothesise possible relationships among them. Determine whether the relationships are true, create meaning of the results and interpret them

The paradigm of **Abstraction (modelling)** consists of four stages rooted in the experimental scientific method and is the steps taken in investigating a phenomenon. A hypothesis must be formed. A model of the system investigated should be created, and predictions should be put down to reason about the hypothesis. Data should be gathered, and an experiment should be designed to test the model, predictions, and hypothesis. The results from the experiment should be interpreted, and conclusions drawn from them.

The last paradigm of **Design** consists of four stages rooted in engineering and is the steps taken in constructing a system, or device, to solve a given problem. A system should be outlined based on functional and non-functional requirements that the system must fulfil. Based on these requirements, a system specification should be stated to ensure the system can fulfil the set requirements. The design and implementation must be documented, and the system must be tested.

This thesis resides in the second paradigm: ***abstraction***, but as stated in the full report [8], there is never *just* one paradigm at play as their underlying processes are intricately intertwined. While this thesis resides mainly in the

second paradigm, elements from both *design* and *theory*, especially *design*, will be recognised. It is expected that these four steps of *abstraction* will be iterated and potentially re-iterated i.e. when a prediction based on the model does not agree with experimental results.

This thesis started out with observations in ocean communication networks [9, 10, 11]. Based on these observations, a hypothesis has been stated (Section: 1.1), a model of a real-world system has been made (Chapter: 3), and predictions have been made based on the model. Several experiments have been performed, and the results will be presented and analysed in Chapter 4. The real-world system is complex, which makes the model complex. A distributed system, Aktan, has been designed and implemented to run experiments with the model and facilitate analysis (Chapter: 3).

We set the sampling frequency to *1*-minute, giving us a tradeoff between computation time, storage requirements, and simulation accuracy. We acknowledge that some boats might achieve speeds above *50* kn and that they might travel more than *1.5* km between two samples with this granularity.

Thesis statement: Multi-hop MANET routing enables continuous connectivity to on-shore infrastructure for boats in Norway's exclusive economic zone.

To reason about this statement, we outline 4 research questions: 1. Given constraints on communication range in Norway's exclusive economic zone, what level of connectivity is achievable? 2. Is the level of connectivity in this network within the range that one would expect for continuous connectivity? 3. Can AIS positional data provide information to model and simulate the network topology? 4. Can AIS positional data identify regions of interest to improve overall network performance?

To answer these questions, we will: 1. Briefly look at previous work [12] where the topology was analysed horizontally without on-shore infrastructure. 2. Expand the previous work by including on-shore infrastructure and simulating the network topology with multi-hop MANET routing. 3. Simulate the network topology in discrete timesteps. At every time step, path searches from boat to shore are performed.

To answer question 4, we will: 1. Map the positional data of boats that have little or no connection to the global network and visually inspect the regions in which they reside. 2. Perform a neighbourhood analysis.

1.3 Scope and Limitation

This thesis uses information from a limited period of three weeks in June 2023; only a subset of the boats seen in a year is present. The positional data was given as a suggested period from the NCA. The type of vessel and sizes are not considered or filtered for. The NCA stated that June would likely have a high density of vessels, as it is the high season for maritime activity in Norway. This thesis models this data and simulates the network topology to investigate connectivity. Hardware deployment and physical layer simulations have been considered but found unsuitable for the study.

Entities in the marine sector have infinite possible combinations of assembly. Their hardware characteristics, such as shape, height, material, etc., are not limited, and each entity is often hand-tailored to a specific use. It is infeasible to model every detail of this system, at least for this thesis. Assumptions have been made, and here they are outlined:

- Radio communication range is equal for all entities under the same entity type. i.e. all *off-shore nodes* are considered equal, and all *on-shore infrastructure* is considered equal.

- The only attribute that affects a communication link is *distance*.

- No interpolation or extrapolation of data has been performed. Given that this is a discrete-time analysis, this can skew the results negatively by boats not being present in all timesteps of the analysis, more on this in Section:5.2.

- The assumptions for communication range are based on the work of Rao et al. [9, 10], and Surendran et al. [11].

1.4 Context

This work has been performed in the context of the Cyber Security Group (CSG) and Cyber Physical Systems (CPS) group. The Cyber Security Group (CSG)'s primary focus is fundamental systems problems that are practically applicable in interdisciplinary real-world scenarios. In the maritime domain CSG has over 30 years of experience, starting with distributed AI-based weather forecasting in StormCast [13] to more recently investigating and developing privacy-preserving monitor systems for fishing vessels [14, 15, 16]. The Cyber Physical Systems (CPS) group focuses on distributed, parallel systems, including their design, architecture, implementation, and behaviour. In the domain of Ad

Hoc Networks (ANET), the group is running the Distributed Arctic Observatory (DAO) - a project with interdisciplinary groups from both the Department of Computer Science and the Department of Arctic and Marine Biology at UiT, The Arctic University of Norway (UiT). The goal of the DAO project is to develop robust, efficient, and autonomous monitoring systems under challenging conditions in the Arctic.

This project investigates problems related to reliable means of communication, a problem all of these systems rely on or are challenged by; more precisely, it explores the potential for Mobile Ad Hoc Networks (MANET)s using low-cost, commodity, infrastructure as an alternative to or in collaboration with other technologies like satellite communication at sea.

1.5 Outline

The rest of this paper is organised as follows:

Chapter 2 outlines concepts and background information relevant to the rest of the paper.

Chapter 3 describes the data analysis processing pipeline.

Chapter 4 provides a view of the results from the analysis.

Chapter 5 a summary of the problem, the findings and its implications.

2 Background

2.1 Wireless communication

Communication between two nodes in a network without wired links is considered wireless, and are commonly classified as either *infrastructured* or *ad-hoc* networks. An infrastructure network typically consists of host nodes; end-system or user devices, and base stations; intermediate nodes connecting hosts to a part of the global network. In ad-hoc networks, nodes are not organised in pre-defined topologies, and hosts serve as intermediate nodes, relaying packets to and from other hosts. In both network classes, one typically separates on the amount of *hops*; the number of links a network packet will have to traverse to reach a base station or some other network gateway. For single-hop routing, a packet traverses precisely one link to reach a router or its destination. Meanwhile, for multi-hop routing, a packet traverses more than one link to reach a router or its destination.

The most common wireless infrastructure one interacts with today is single-hop networks with deployed infrastructure. Typical cellular networks and Wi-Fi zones fall into this category [17].

2.1.1 Frequency bands and properties

Frequency is the main factor when considering the properties a wireless network will achieve. Higher-frequency signals can carry more data than lower

ones, usually at the cost of reach and susceptibility to interference. The systems discussed in this thesis utilize Very High Frequency (VHF), Ultra High Frequency (UHF), and Super High Frequency (SHF), frequencies from 30MHz up to 30GHz [18].

It is worth noting that higher frequency bands are considered for the sixth generation cellular network (6G) and beyond, with the potential to increase the bandwidth substantially, opening new possibilities for larger data transfers in cellular networks [19]. The sixth-generation cellular network is entering the pre-standardization phase at the time of writing and is therefore omitted from this thesis.

All frequency ranges described in this thesis have the same coverage limitations of Line Of Sight (LOS). They are susceptible to terrain multipath propagation, while Super High Frequency (SHF)s are also vulnerable to weather [18]. An important metric to consider with LOS radio propagation is the fresnel zone. These are confocal prolate ellipsoidal regions between the sender and receiver; obstructing or deflecting objects between the sender and receiver can impact the signal strength due to interference. As a rule of thumb, the first fresnel zone should be 80% clear, and at least 60% clear [20]. The radius of a fresnel zone at a point p is given as:

$$r_n = \sqrt{(n \times \lambda \times d1 \times d2/(d1 + d2))} \tag{2.1}$$

Where n is the zone number, λ is the wavelength, $d1$ is the LOS distance from p to antenna 1, $d2$ is the LOS distance from p to antenna 2, total LOS distance between the two antennas is the sum of d1,d2: $D = d1 + d2$. It is most interesting to see the largest radius, which is directly in the middle of the two antennas: $d1 = d2 = \frac{D}{2}$. In this thesis, we are mostly concerned about free space ocean communication at the open sea, where the maximum radius is interesting. Equation (2.1) then simplifies to:

$$r_1 = \sqrt{\lambda \times \frac{D}{4}} \tag{2.2}$$

Finding the wavelength λ can be done by applying the wavelength formula:

$$\lambda = \frac{V}{f} \tag{2.3}$$

V, for the frequency ranges and medium considered, is assumed to be the constant speed of light in air, $C = 299792458\,\text{m/s}$, and f is the radio frequency.

AIS

AIS uses the four frequencies: [156.775, 156.825, 161.975, 162.025] MHz where the two lower ones are intended for satellite communication [21], which means that for ship-to-ship AIS communication, the first layer fresnel zone radius is then between:

$$\sqrt{\frac{c}{161.975MHz} \times \frac{D}{4}}, \sqrt{\frac{c}{162.025MHz} \times \frac{D}{4}} \tag{2.4}$$

depending on the distance between the transmitter and the receiver.

WIFI

The Ocean Communication Networks (OCN) [9] establish a multi-hop extension to an existing cellular network using 2.4 and 5 GHz using commodity WIFI hardware from Ubiquiti [9]. The fresnel zones for 2.4 and 5 GHz frequencies are given as:

$$\sqrt{\frac{c}{2.4\,\text{GHz}} \times \frac{D}{4}}, \sqrt{\frac{c}{5\,\text{GHz}} \times \frac{D}{4}} \tag{2.5}$$

Cellular

For cellular towers, this becomes tricky. In each generation of cellular network deployed, i.e. (2 G, 3 G, 4 G or 5 G), the frequency bands for each generation(G) do not necessarily have anything to do with GHz. To complicate the matter, each network provider buys licenses for different frequency bands for their implementation of a given generation network.

For this thesis, the assumption is that the base stations deployed by telecommunication companies reside high in the terrain, with optimal reach and use 2.4 or 5 GHz frequencies. Some coverage maps provided by the companies themselves suggest varying ranges at sea [22, 23, 24]. Rao et.al. [9] achieved links up to 51 km in the hop between shore and boat, given a base station at 56 m Above Mean Sea Level (AMSL) and antennas on board vessels at 8-9 m height [9, 25]. They achieved up to 21km range for boat-to-boat communication on 2.4 GHz.

No specific data on the link range for 5 GHz is presented, but one can assume it to be lower due to propagation loss[25]. The producer of the field equipment estimates that link ranges above 6,5km, with similar equipment in the 5 GHz band, is not achievable [26], for a bandwidth of 100MHz, but perhaps achievable for 10 MHz (lowest available at the time of writing) [18]. This is twice the

bandwidth that Rao et. al. used in their field trials [9].

2.1.2 Curvation Drop

Since the earth is not flat, we must consider its curvature and how that affects the perceived height between two point-to-point transceivers. The curvation drop between two points on a sphere can be written as:

$$\begin{aligned} drop &= \frac{R_s(1-\cos(\theta)}{\cos(\theta)} \\ &or \\ drop &= R_s - R_s \times \cos(D/R_s) \end{aligned} \tag{2.6}$$

R_s is Earth's radius, D is the distance between the two transceivers, and $\theta = D/R_s$. Which for small distances can be simplified to [27]:

$$drop = 7.848 \times D^2 \tag{2.7}$$

For the first *100km*, the deviation is small between the simplification and true drop, as seen in Figure 2.1. We will use the simpler one since we will not look at distances over 100 km.

Distance(km)	$(r - r \times \cos(\frac{D}{r}))$(m)	$7.848 \times D^2(m)$
1.0	0.08	0.08
10.0	7.85	7.85
20.0	31.39	31.39
30.0	70.63	70.63
40.0	125.57	125.57
50.0	196.20	196.20
60.0	282.53	282.53
70.0	384.55	384.55
80.0	502.27	502.27
90.0	635.68	635.69
100.0	784.79	784.80

Figure 2.1: Curvation drop in meter for different distances between points

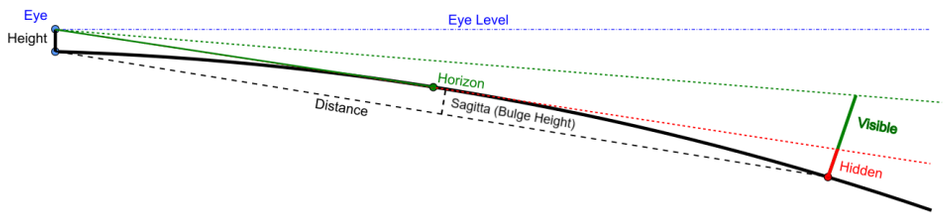

Figure 2.2: Curvation drop and LOS on Earth [27]

2.1.3 Properties of electromagnetic waves

Electromagnetic wave propagation is a very complex field, and while not the focus of this thesis, it is something one must consider. There are simulation tools out there, i.e. Ubiquiti has a network design visualisation tool [28], one can set up access points with specific hardware to estimate link quality and viability in a network topology, as seen in Figure 2.3.

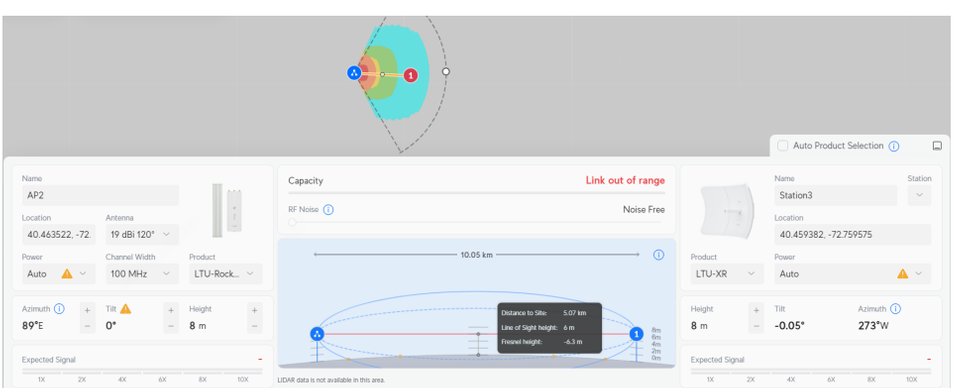

Figure 2.3: 5GHz 120degree Point-to-Point (PTP)-link 100MHz bandwidth

Calculations with the first fresnel zone and curvation drop estimates above yield results similar to those of the Ubiquiti simulation with full bandwidth. It's important to remember that simulations may have limited value in estimating actual system deployments where a broad range of external sources, such as sea conditions, weather, and varying noise sources, will impact the propagation and might not be included. The fact that the first fresnel and curvation calculations are similar suggests that this simulation is conservative, but note that refraction has *not* been considered.

A deeper discussion on the properties of electromagnetic waves is beyond the scope of this thesis. Still, we would suggest *Spectrum 101: An Introduction to Spectrum Management* [18], and *Fundamentals of Analogue and Digital Communication Systems* [29] for the interested reader.

2.2 AIS

AIS is designed to provide coastal authorities and other ships information about a ship. AIS is part of the navigational equipment that the International Maritime Organization (IMO) requires ships above 300(cargo ships:500) gross tonnage to carry. AIS should always be in operation unless international agreements provide for the protection of navigational information.

Figure 2.4 shows the architectural model of the system; it is capable of both ship-to-ship, ship-to-shore and ship-to-space communication, depending on the AIS transceiver [21]. There are two classes of AIS transreceivers, *class A*, and *class B*. Class A transceivers are intended for larger commercial ships that are required to report more information and at a higher frequency than leisure ships. Class A transceivers are also prioritised over class B transceivers [21]. In this thesis, Maritime Mobile Service Identity (MMSI), time, longitude and

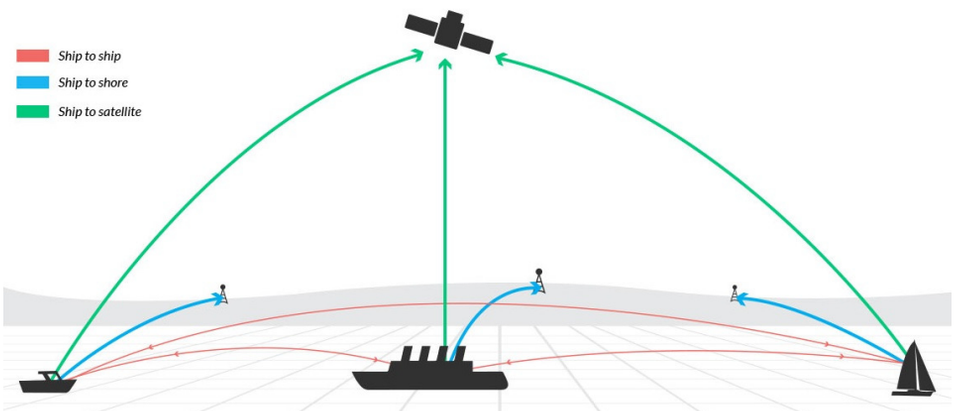

Figure 2.4: AIS architecture. Source: North Atlantic Treaty Organization [30].

latitude from AIS broadcasts will be used; these are broadcast via messages of type 1-5, 18(class B), and 27(outside base station coverage) [21]. See Appendix 6 for details on broadcast frequency, message content and long-range packet 27.

2.2.1 Problems

Broadcast frequency and disabling

The broadcast frequency of AIS information varies greatly throughout a ship's journey, depending on the type of vessel and speed. While it's generally not allowed to disable a transceiver on board a ship, there are several exceptions, and as a result, ships might seem to disappear in the records for varying

amounts of time. For example, a ship at port may turn off its AIS unit.

This problem with *cooperative* AIS is currently being addressed by using radar and satellite surveillance to achieve *non-cooperative* AIS systems where ships that disable their transceivers are still being tracked [31].

Spoofing

AIS can easily be spoofed by altering the information a transceiver broadcasts out. This can be any of the values in a message, not limited to MMSI, date, positional data. By doing this, ships can take the identities of other ships or create false positional tracks, for example, when hiding illegal fishing.

Availability

Ships in the maritime sector typically have high degrees of freedom and can move over vast distances. However, the electromagnetic waves AIS uses for communication cannot travel indefinitely far. This makes the network prone to partitioning, where ships cannot communicate with all the other ships in the system. To mitigate this, ships can propagate information on behalf of other ships through the network. AIS-compatible satellites are also being deployed, making the AIS resilient to network partitioning. [31, 21].

2.3 Overlay Network

Overlay networks are abstractions on top of physical networks, where links are logical. MANETs are typically implemented as overlay networks on top of a physical network. From a theoretical perspective, we can define an overlay network as a graph G containing vertices V that are connected by links E. Since the overall goal of this network is to extend the connectivity of a global network of vessels at sea, if a node is connected, it is in the global network. Still, one must remember that links in the overlay network are logical [17]. Links inside the overlay network may be hidden behind gateways, such as cellular towers or cable switches, to any node outside the overlay network. Cabled networks like ethernet typically have point-to-point communication between nodes, while anyone in the propagation zone can listen to a radio wave.

A much-used type of overlay network that this thesis draws inspiration from is the idea of Virtual Private Network (VPN)s [17]. Each base station could be fitted with middleware that can handle the special packets from the MANET network

and further propagate it to the correct destination in the global network or to the correct node in the MANET.

2.4 Graph

A powerful abstraction is modelling real-world systems as graphs. Graphing complex systems is frequently used in various disciplines, such as math, engineering, and biology to mention a few. For this thesis, the system consists of thousands of nodes that may or may not be connected, and it is natural to model a system like this as a graph.

2.4.1 Definition

A graph is a set of entities, typically called a *vertex*(V), and a connection between any two vertices is called an *edge*(E). A set of vertices and edges make up a *graph*(G).

The Definition 1. A graph *G* is a set of vertices *V* and edges *E*[32]

In this thesis, any entity that broadcasts AIS messages or can communicate with the other participants, i.e. on-shore base stations, is considered a *vertex* or *node*. An edge exists between any two nodes if they are within their respective communication range capabilities.

Graphs are typically segmented into two categories: *Directed* and *Undirected* graphs.

2.4.2 Directed graphs

A directed graph is a graph where edges are *uni-directional*, and information flow is *one way*. For any $V_1 \in (G)$ with an edge E_{12} to V_2 there is no guarantee the inverse relation exists:

$$v_1 \xrightarrow{e_{12}} V_2 \nRightarrow v_2 \xrightarrow{e_{21}} V_1 \tag{2.8}$$

2.4.3 Undirected graphs

An undirected graph, on the other hand, assumes that if there exists an edge E_{12} between two vertices V_1 and V_2, the flow of information is *bi-directional*

$$V_1 \xrightarrow{e_{12}} V_2 \Rightarrow V_2 \xrightarrow{e_2 1} V_1 \tag{2.9}$$

In a ship-to-ship/ship-to-shore network, depending on the type of communication, both directed and undirected graphs may be present simultaneously. This may cause problems in a network topology where the overarching goal is to route information from a to b. Undirected and directed graphs can be classified as either *cyclic* or *acyclic*.

An *acyclic* graph is a *directed* graph where traversal *cannot* re-visit a vertex V after being sent through one of V's outgoing edges, the paths do not form *closed loops*. A *cyclic* graph is a graph where traversal *can* re-visit a vertex V after walking one of V's outgoing edges, forming closed loops.

2.4.4 Graph search

Graph modelling does not provide much new insight in itself. A node in a graph or a boat in the sea alone provides little information. One of the strengths of graph modelling is that the study of relations between nodes in graphs comes naturally. One very powerful tool is graph searching, traversing edges between nodes to unveil connections, relations, and patterns that otherwise are difficult to uncover. There are many ways one might go about traversing a graph, and this thesis heavily relies on a technique called DFS, but another common technique is called BFS. To understand why one would choose one over the other, we need to remember two graph search properties. A search is considered *complete* if it can find a solution for any graph that has a solution. With no constraints on time, this might take infinite time. Further, a search is considered *optimal* if, for a search, it finds the *shortest path*.

A depth-first search is a *non-optimal non-complete* search technique, meaning that unless one takes precautions when setting up the graph or search, one may not find solutions, or if a solution is found, it is not guaranteed to be the shortest. Depending on the implementation, a DFS will apply the same *heuristic* when selecting a path among a set of paths. i.e. always starting a search by traversing the node's right-most outgoing path. A depth-first search is good for answering questions like: "Is there any path between node a and node b in a graph G".

An alternative is to use a BFS that is both *optimal* and *complete*. Instead of applying a heuristic way of selecting a path among a set of paths from a node, a breadth-first search will check any of the adjacent nodes before traversing a graph further and thus is good for answering questions like: "What is the shortest path between a node a and a node b".

Examples of both can be seen in Figure 2.5 and the different ways a graph traversal will be performed with the different techniques. The left graph is a unidirectional, cyclic graph, and a breadth-first search traverses the graph in a *hierarchical* way, visiting every adjacent node from a starting point 1. Since the graph is cyclic, a naive search could end up in an infinite loop in one of the subgraphs $[\{1, 2, 3, 6\}, \{1, 3, 4, 8\}, \{1, 2, 3, 4, 6, 8\}]$, but a record of visited nodes are typically recorded and checked for revisits to stop infinite traversal. The right graph is a directed, cyclic graph with unweighted edges. Here, a depth-first search traverses the rightmost outgoing edge for every node visited, and as with the BFS, if no caution is made, searches may end up stuck in infinite loops[32].

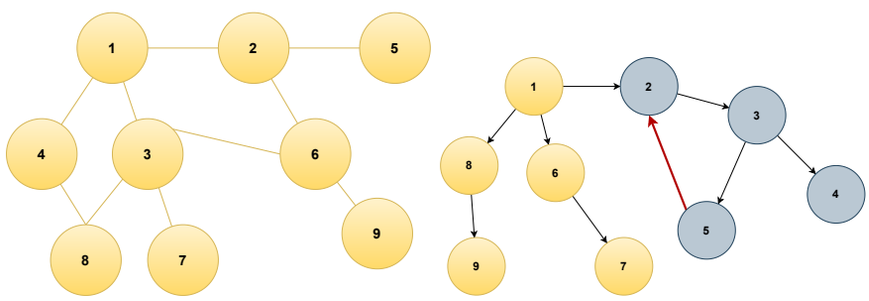

Figure 2.5: Left: BFS traversal in a cyclic, unidirectional graph. Right:DFS traversal in a directed cyclic graph.

2.5 Performing hops

For a packet to reach its destination, unless it's a one-hop, we must relay information through the network. In wired networks, this is usually performed by putting electricity on a fixed physical interface with a point-to-point connection to the next node. In a switched/routed fashion, the next node will take incoming unique IDs(IP) of adjacent devices and map them to outgoing adjacent devices according to some mapping rules. An intermediate device keeps a lookup table for each of its source-destination pairs. The lookup table is maintained either locally in each intermediate based on some routing protocol or governed by a separate control plane. The systems implementing separate data/control planes are often referred to as Software Defined Network (SDN)s [33]. This is analogously similar to wireless networks but much more

complicated, as we do not have stable linked connections between nodes but instead noisy, free-space channels [17, 18].

2.5.1 Data paths

In the work by Rao. et al. [9] Ocean Communication Networks (OCN) is implemented as a multi-hop network to extend cellular network coverage and tested in the Arabian Sea using Long-range WIFI. Their tests use commodity hardware from Ubiquiti, and Cisco[9, 26, 34], where each boat would have directional antennas increasing their potential range. The tests showed promising results, with first-hop shore-to-ship of *50* km, while subsequent hops managed *20* km. With WIFI as the underlying technology, this setup is minted on high bandwidth communication at the expense of range.

2.5.2 Control path

Following the idea behind Software Defined Network (SDN)s, the control path does not need to rely on the same technology as the data paths and could, i.e. be implemented using different frequencies. Since control path information is typically orders of magnitude smaller than data path information, lower frequencies could increase communication range at the cost of bandwidth. The path data takes in the network does not follow a universal rule of thumb. Numerous routing protocols have been proposed and implemented to solve problems in routing information through a network. For ad-hoc networks, Boukerche et. al.[35] have extensively classified routing protocols as seen in Figure 2.6 [35].

The two most discussed approaches to MANET routing are Link State Routing (LSR) and Distance Vector (DV), which both have seen many variants proposed to optimize for different scenarios. Most notably, LSR has been extended in Optimized Link State Routing (OLSR) version 1 and 2, and DV have been extended with Ad Hoc On-demand Distance Vector (AODV). These two approaches have very different underlying mechanisms; the LSR family tries to keep updated routing tables at all times, preventing path discovery when data is to be sent. AODV does not keep pre-discovered paths but instead performs a path discovery when data is to be sent. LSR optimizes for latency at the cost of communication overhead, keeping tables up to date. In contrast, AODV reduces maintenance overhead at the cost of latency by computing the path *on demand*[35, 36].

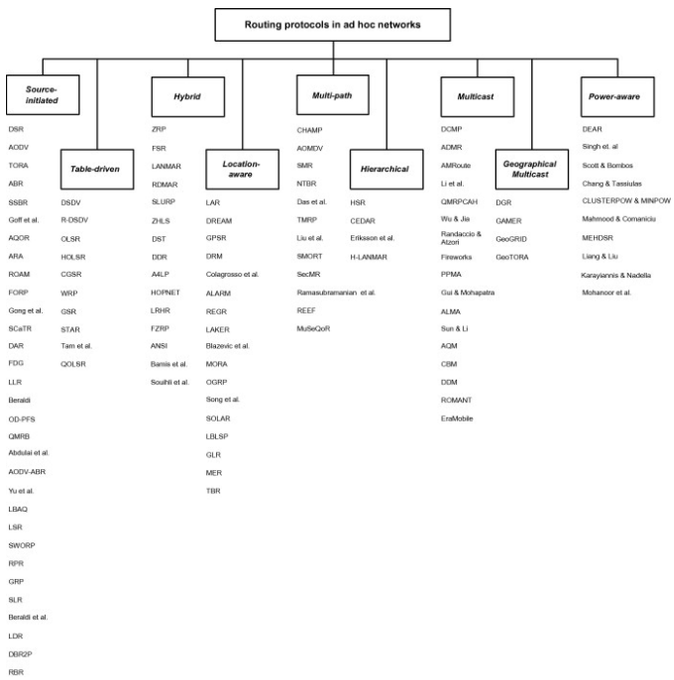

Figure 2.6: MANET routing protocol overview[35]

2.6 (Open) MPI

Today, it seems almost foreign that a message-passing interface has not followed a standard since the dawn of distributed computation, but in the early 1990s, this was very much the case. Extensive work from researchers and the industry eventually led to the standardization of Message Passing Interface (MPI) in the mid-1990s. In the next 30 years, several versions were released, and at the time of writing, version *3* of MPI is the latest, while the open source implementation of the standard OMPI has reached version 5. *MPI* hase become the *de facto* standard to coordinate communcation and collaborative calculations on clusters of machines [37, 38].

The MPI standard defines a general message-passing model with both collective and point-to-point communication capabilities. The later versions allow for remote memory access and non-blocking execution, include bindings for several languages, and even integrate with multi-thread libraries such as open-MP [38].

2.7 Vantage Point Tree

A VPTree is a partitioning data structure for segregating data in *metric* space. It separates the space by choosing a starting vantage point that encapsulates it. Then it splits the space into *subspaces* based on a distance d from it and separates points outside of this d in one *sub-tree* while points closer in another sub-tree, and is illustrated in Figure 2.7. The partitioning allows for average $O(logn)$ Nearest Neighbour (NN) searches with a construction complexity of $O(n)$ [39, 40].

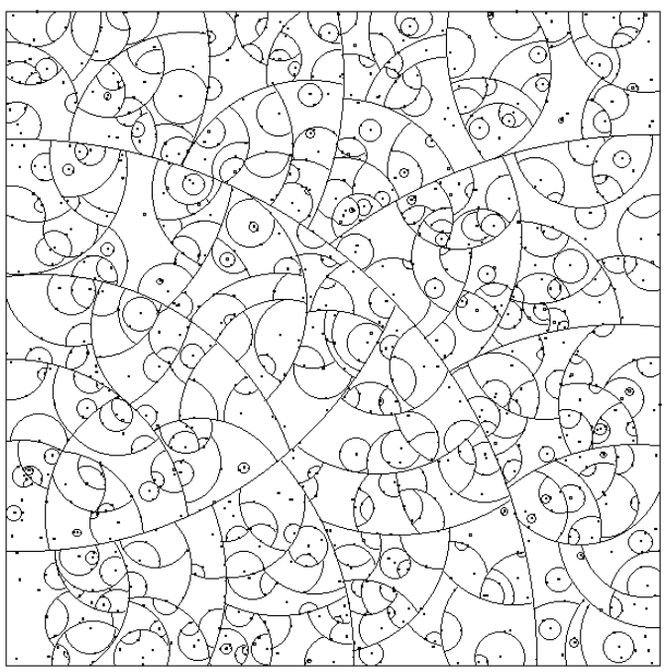

Figure 2.7: Vantage point tree decomposition, center points and circles showing subspaces [39].

2.8 World Geodetic System 84

Global positioning systems such as Global positioning system (GPS) and Galileo require a clearly defined coordinate system to precisely give a position on Earth's surface. The World Geodetic System 84 (WGS84) is a geocentric definition of such a coordinate system, Figure 2.8 shows the reference frame, and is under current development from the 1984 standard and is consistent with the meter level[41, 42].

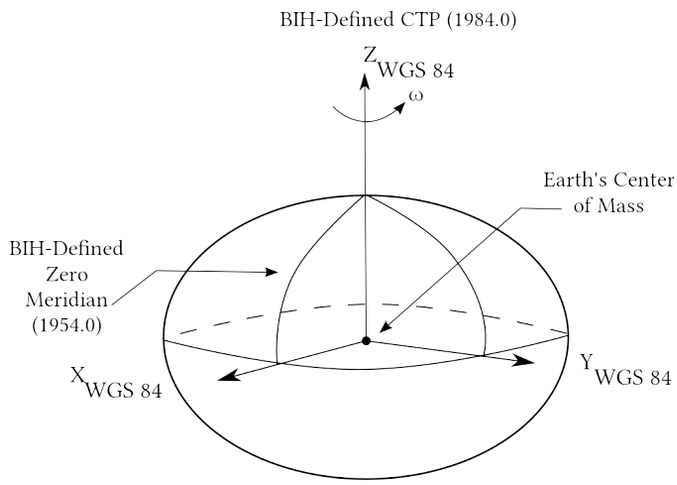

Figure 2.8: WGS84 reference frame [42]

2.9 Ocean Communication Network

The OCN [9, 10, 11] is a multi-hop MANET overlay network that tries to extend cellular/IP reach at sea by using low-cost commodity hardware. OCN's architecture is seen in Figure 2.9, and there are 3 entities in this system: Supernodes(nodes with 2 set of backhaul equipment), adaptive nodes(nodes with 1 set of backhaul equipment) and access nodes(no backhaul equipment).

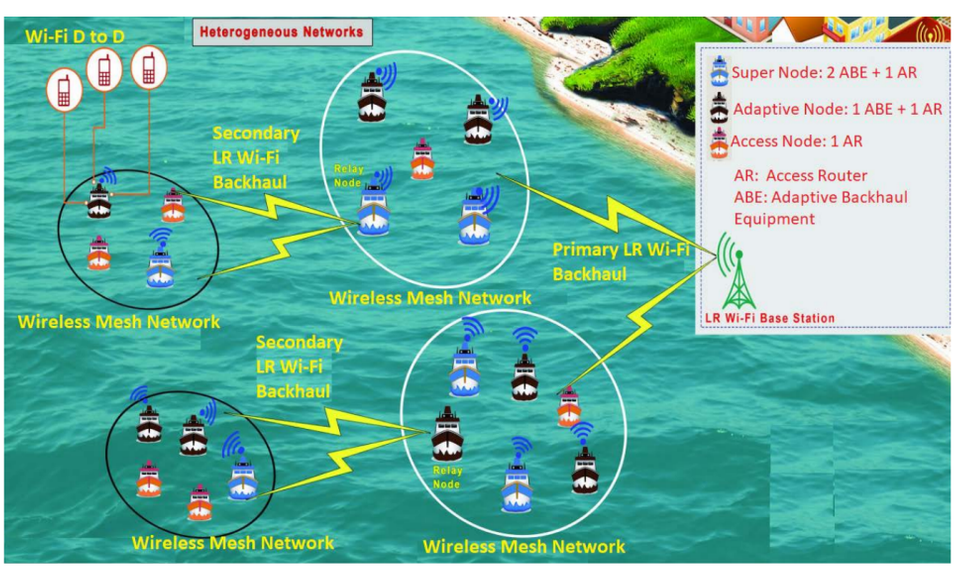

Figure 2.9: The OCN architecture [11].

This work achieved boat-to-boat communication upto 20 km and boat-to-shore ranges up to 50 km.

2.10 Previous work

This thesis extends and improves previous work by the authors [12] where a pipeline to preprocess and structure data was designed and implemented. This pipeline was designed to run locally on the computer where analysis occurred, using multi-processing [43] on shared objects in memory and on disk. It spreads the workload by segmenting timesteps evenly across microprocessor cores. Since the problem was memory-bound, the number of active cores would be determined dynamically based on available memory.

Every object was stored in a tabular data storage(pandas), and adjacency matrices were generated for each boat to traverse in neighbourhood analysis. The system was designed for a workload of 4500 boats in 9600 timesteps, and was able to process this data in about 20 minutes [12]. This proved insufficient when the problem was scaled up to the workload for this thesis. With 14500 boats over 28800 timesteps in 9 independent experiments this would take $20min \times \frac{14500}{4500} \times \frac{28800}{9600} \times 9 = 1740minutes$ or 29 hours.

The work in [12], analysed vessel positions without regarding on-shore infrastructure and looked at neighbourhood clustering and message propagation over time. Figure 2.10 shows that it takes more than *3* days for a packet to traverse the whole network without infrastructure. This is measured by finding all neighbourhoods for every timestep and merging any neighbourhoods that share nodes in previous timesteps. This assumes that any information a node wants to broadcast to its neighbours is completed in the time they see each other. The x-axis shows the logarithmic time since the start of the simulation, and the y-axis shows the number of neighbourhoods seen. Note: This study's sampling frequency was 18 minutes.

Figure 2.11 is the plot over adjacent neighbours for each node in the network. The y-axis is the number of adjacent nodes, the x-axis is the ID of a node(boat). It is measured by performing a query for all boats within a set range, in this instance, 20 km, for each node. Based on this study, it is difficult to reason about centrality, but it highlights a problem. A node with over 200 adjacent neighbours will have problems with TDMA limitations for multiple access over the same channel [12, 44], if these nodes are central, they will become bottlenecks for data propagation.

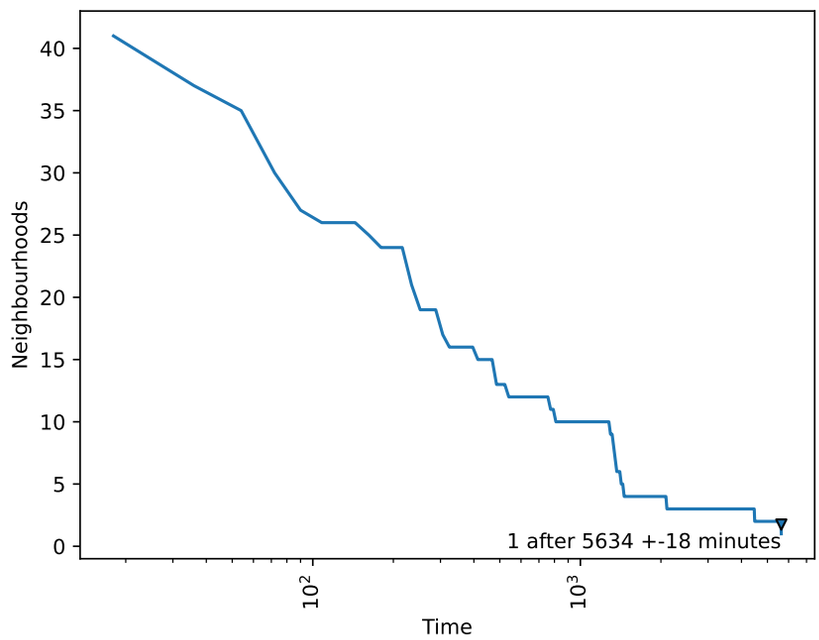

Figure 2.10: Neighbourhood reduction over time with *20* km node-to-node range [12].

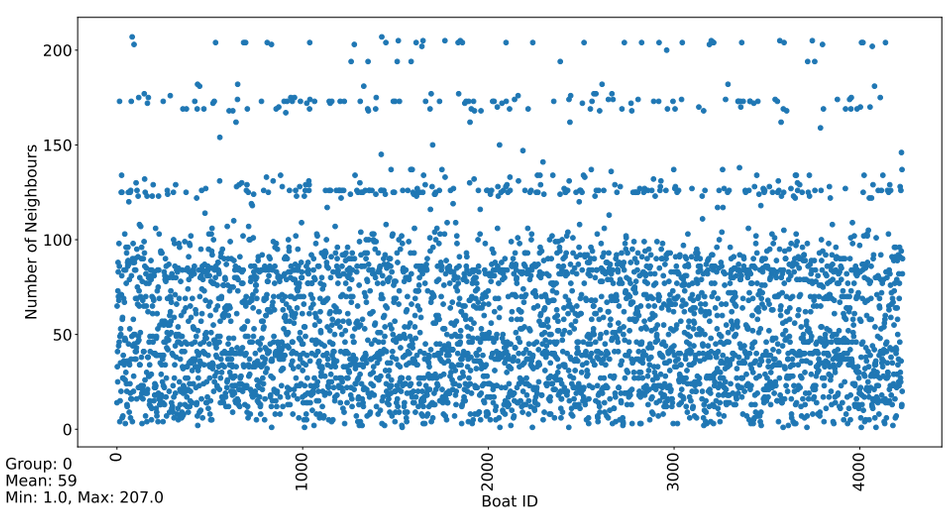

Figure 2.11: Number of adjacent nodes, per node [12].

2.11 NS3

This thesis originally created a simulated network using Network Simulator 3 (NS3). Network Simulator 3 (NS3) is an open-source, powerful discrete event network simulator that allows researchers to model and simulate very complex problems[1]. It is implemented in C++ and has bindings for Python, allowing network simulations to be scripted in Python. The simulator gives access to all abstraction layers of the network stack from physical to application, allowing fine-grained control of the simulation environment. NS3 has been used in thousands of publications to date[2].

In the initial phases of this project, work was put into using this tool to achieve an accurate model and simulation of the real-world system. A small simulation with varying numbers of mobile nodes was implemented in a 1500×1500m area and OLSR as the routing protocol.

Some problems occurred: 1. The initial proposition of this project was to look at centralised calculation and distribution of optimal paths in the network. Implementing this protocol in NS3 turned out to be very difficult. 2. When we increased the area size and the number of nodes, the simulation took a very long time to complete. The exact cause of this is unknown to the authors, and unfortunately, we could not find enough information about this within the ns3 community. If this is a limitation in the simulator, then *ns3* cannot model the system we are looking at. 3. The Python bindings have a very long(15min+) setup overhead when run with or without visualisation. This is most likely due to the runtime building of C++ to Python modules using cppyy[3].

Due to these problems, development time skyrocketed, and at some point, the project had to find another solution to maintain its schedule.

1. https://www.nsnam.org/
2. https://www.nsnam.org/research
3. https://www.nsnam.org/docs/manual/html/python.html

3 Aktan

To find answers to our thesis: *Multi-hop MANET routing enables continuous connectivity to on-shore infrastructure for boats in Norway's exclusive economic zone*, we have created a novel simulation system, Aktan. This system creates a simplified model of a real-world topology and simulates network topology in discrete time intervals.

3.1 Model

The real-world antennas, boats, and infrastructure we are looking at can be created, placed, and moved in infinite combinations, with ever-changing characteristics and properties such as hardware specifications, mobility capabilities, height, material, etc. It is infeasible to try to achieve a replication of the real world. In this thesis, two entities are modelled: boats (nodes) and on-shore infrastructure, nodes with different assumptions on communication capabilities. Entities of each type have the same underlying assumptions in capabilities: *boats* are assumed to have the same capability of communication range, even though that cannot be true, i.e. for vessels of varying height as seen in Section: 2.1.1. *Base stations* also have hardware and geographical placement differences, with varying noise sources around them or obstacles that can greatly impact their communication range but are assumed to have equal capabilities and optimal omnidirectional reach.

3.1.1 Base stations

Without information from telecom companies about the placement of base stations, having a realistic view of the network is difficult. Instead, we have used the coastline shape files from Natural Earth Data [45] and assume every point along the coastline is a base station. This is a very optimistic assumption.

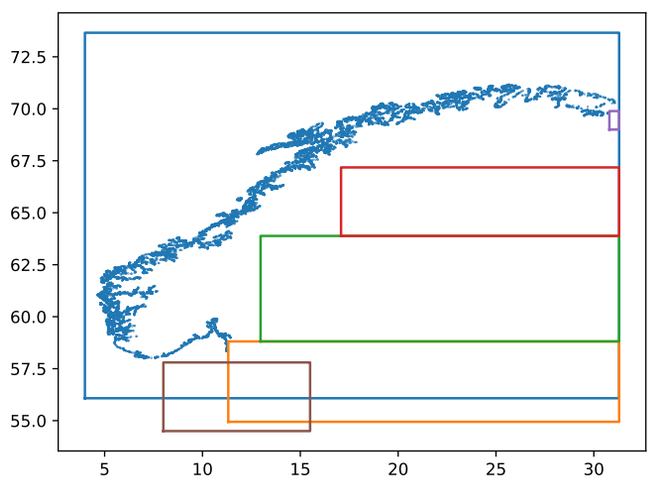

Figure 3.1: Overview of area covered, base station placement and bounding boxes for filtering vessels outside the EEZ

3.1.2 Sampling frequency

The system has a lower bound sample frequency of *one (1)* second.

Frequency limitations

A more fundamental problem in high-frequency sampling is the limited update frequency of AIS broadcasts. Boats are not required to update their position more frequently than every 2 to 10 seconds [21]. This is sufficient for this analysis, but finer granularity would be preferred for some cases, i.e., low-level physical layer simulations or filling in missing data points when AIS broadcasts are lost.

One could increase the granularity by inter- and/or extrapolating between known positions, but this is not unproblematic. Boats have a high degree of freedom, and unlike nodes in, i.e. Vehicular Ad Hoc Network (VANET)s, they

are not bound by designated paths, except for narrow fjords, gulfs, and straits. This makes interpolation and extrapolation of vessel positions particularly difficult, and numerous methods have been applied to predict boat movement from Kalman filters to deep neural networks [46, 47, 48].

3.1.3 Geographical limitations

The original dataset contained vessels outside of the EEZ; we limit the scope to the EEZ in this thesis. As seen in Figure 3.1, the *blue* rectangle marks the bounding box for Norway, the outer periphery a boat can sail. The *purple*, *red*, *green*, *orange*, and *brown* are bounding boxes used to filter boats in the Baltic Sea/Gulf of Bothnia, Denmark, and Russian waters. As shown in Figure 4.1, the dataset also contains quite a few boats "inland", i.e., boats travelling the Telemark Canal; these have been manually filtered out.

3.1.4 Tiling

The boats are segmented into what we call 'layers' to optimise the search for paths. As seen in Figure 3.2 each layer is a band around the coastline. Nodes in *layer 1* are assumed to reach infrastructure directly(one hop), while boats in the other layers need at least *two* hops to reach infrastructure. The width of a layer, except the first, is equal to the assumed node-to-node communication range. In the first layer, the communication range of base stations is assumed to be *50* km based on field trials from Rao et al. [9]. The node-to-node communication range is an independent variable that takes values ranging between *5* and *110* km for independent experiments.

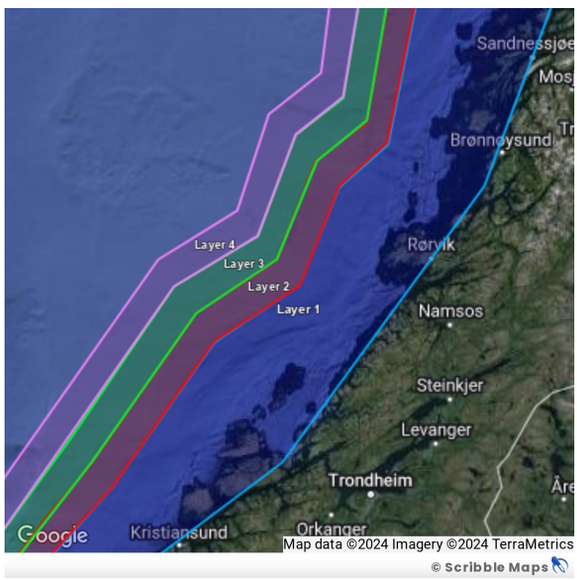

Figure 3.2: Illustration of communication layers. There were up to 30 layers when boat-to-boat ranges were set to 5 km.

3.2 Design

Aktan is a distributed parallel system that relies on a shared filesystem to store information on which the different modules can operate.

3.2.1 Process pipeline

To preprocess and structure the data, a pipeline has been rewritten from previous work [12] and extended to run distributed in parallel with the OMPI as seen in Figures 3.3, 3.4, and 3.5. Each module can be run independently or simultaneously but requires access to shared storage to pass on snapshots and processed data.

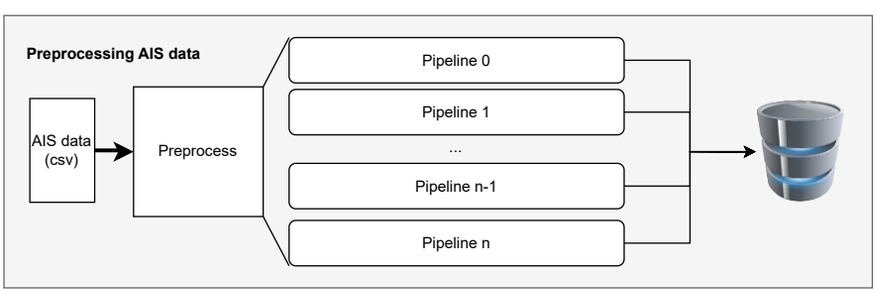

Figure 3.3: Preprocessing of AIS-data.

The preprocessing module seen in Figure 3.3 reads AIS-data from one or more input files and supports a varying amount of filetypes, but defaults to Comma-Separated Values (CSV) as that is what NCA exported. Each file is processed independently in its own pipeline. The pipeline structures the input into tables, performs datatype conversion, filters unwanted data points, i.e. boats outside the EEZ, and saves the tables to parquet files. Optionally, the output can be combined into one file.

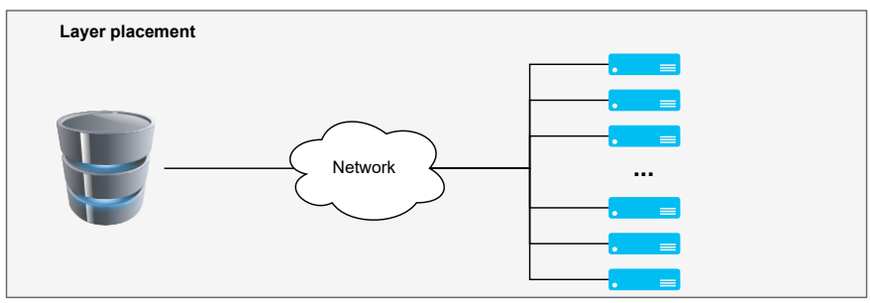

Figure 3.4: Calculation of layer placement for boats.

After preprocessing, we can start to segment boats into their respective layers as seen in Figure 3.4. Each layer is a zone within a given distance from shore, depending on the assumed first-hop and BTB distance, i.e. with 20 km BTB range and 50 km first-hop range layer 2 will stretch from 50- to 70- km from shore. This is an intensive computation, depending on the sampling frequency one uses; therefore, it is set up with the option to distribute it to other computers using OMPI. Each worker can independently determine what part of the data they should work on, and the system requires *no* orchestration.

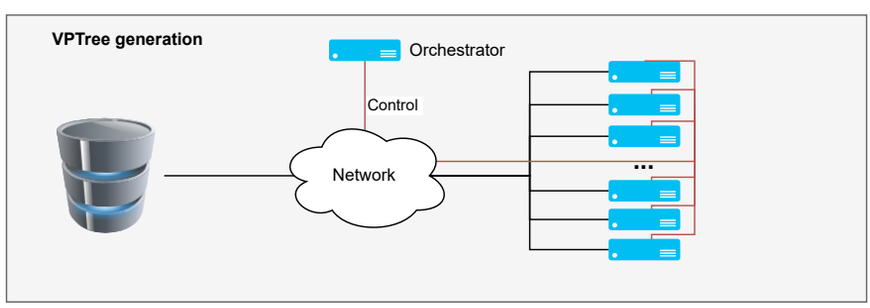

Figure 3.5: VPTree generation from AIS data.

With boats classified into layers for each timestep, we generate VPTrees for each layer as seen in Figure 3.5. Since this project ran several configurations simultaneously(different sampling frequencies, BTB ranges etc.) and these are independent computations, an orchestrator distributes the workload to the workers in chunks.

Each worker begins by communicating its processing capabilities to the orchestrator. The orchestrator then distributes an amount of operations it believes the worker should be able to compute. When a worker finishes the operations, the results are returned to the orchestrator. The orchestrator orders every operation in key-value storage, and once the VPTrees for a configuration has been completely generated; it stores the result in binary format. For a 1-minute granularity, this accumulates to $60\frac{trees}{h} \times 24\frac{h}{day} = 1440$ trees per day.

3.2.2 Path search and data segmentation

Path searches can be very expensive if no data preprocessing is done. Nearest-neighbour searches in tabular data scale linearly $O(n)$ and, in its most naive, becomes a $O(n^2)$ operation [40]. The layering seen in Figure 3.2 is an alternative to rectangular bounding box tiling around nodes and drastically narrows the search space for each query. Each layer is, in turn, structured as a vantage point tree [40], allowing for $O(logn)$ nearest neighbour and in-range searches in these layers.

The path problem, reaching shore s from a node n in a global graph G is broken down into 4 cases:

1: There is a path from a node $n \longrightarrow s$ with at least one hop in each layer closer to shore than layer i, creating a graph g whose edge count (E_c) is bound between

$$layer\# <= E_c < \sum_{1}^{N} n \begin{cases} 1, & \text{if } n \in \{layer_1, layer_2, \ldots, layer_{i-1}\} \\ 0, & \text{otherwise} \end{cases} \tag{3.1}$$

2: There is a path from a node $n \longrightarrow s$ with at least one hop in each layer, including the one a node is in, creating a graph g whose edge count is bound between

$$layer\# <= E_c < \sum_{1}^{N} n \begin{cases} 1, & \text{if } n \in \{layer_1, layer_2, \ldots, layer_{i}\} \\ 0, & \text{otherwise} \end{cases} \tag{3.2}$$

3: There is a path from a node $n \longrightarrow s$ with at least one hop in each layer; creating a graph g whose edge count is bound between

$$layer\# <= E_c < \sum_{1}^{N} n \begin{cases} 1, & \text{if } n \in \{layer_1, layer_2, \ldots, layer_{max}\} \\ 0, & \text{otherwise} \end{cases} \tag{3.3}$$

4: Otherwise there is no path from a node $n \longrightarrow s$.

The system implements a depth-first(Algorithm: 1) and a breadth-first(Algorithm: 2) search from nodes to shore. As we have seen in chapter 2, our depth-first search is not optimal and can only answer questions like "Is there any path from a node to shore?". This is sufficient for a lot of the analyses we want to support. I.e. to start evaluating connectivity, it's enough to find *any* path to shore. It is also interesting to see how long these paths must be to, in other words, "what is the shortest path to shore?". This question cannot be answered with our depth-first search, we need a search algorithm that is both complete and optimal. A breadth-first search is both optimal and complete and can answer this question. We apply a heuristic that we are likely to find direct paths by traversing layer by layer inwards, breadth-first search is only used where the shortest path is important.

Distance calculation

The vantage point trees calculate the distances between two points on the earth's surface by solving the inverse geodesic problem[49], relying on WGS84 as the reference frame [41].

The layers with VPTrees are generated for each timestep and stored in a key-value storage, as shown in Figure 3.6. This allows for fast querying of the wanted trees with an average complexity of $O(1)$.

Timestep	Layer: [Tree]
time	{layer_id:tree, layer_id:tree,. . . }
1	{1: VPTree, 2:VPTree, . . . }
2	{1: VPTree, 2:VPTree, . . . }
. . .	. . .

Figure 3.6: Key-value storage for VPTrees, indexed on *time* since simulation started

3.2.3 Storing and serving data

Keeping everything in memory becomes difficult, with the amount of data rapidly increasing as we unfold and structure it. The storage complexity of VPTrees is $O(n)$. For the three weeks covered, with nine independent experiments running simultaneously, the total memory consumption accumulates to ĩ7 GB per day simulated.

Even when it is feasible to hold everything in memory, synchronising access to shared objects becomes a bottleneck. The data is distributed on several

servers to achieve better performance. The servers are set up with an API, and chunks of information are passed on request to the working clients, as seen in Figure 3.7.

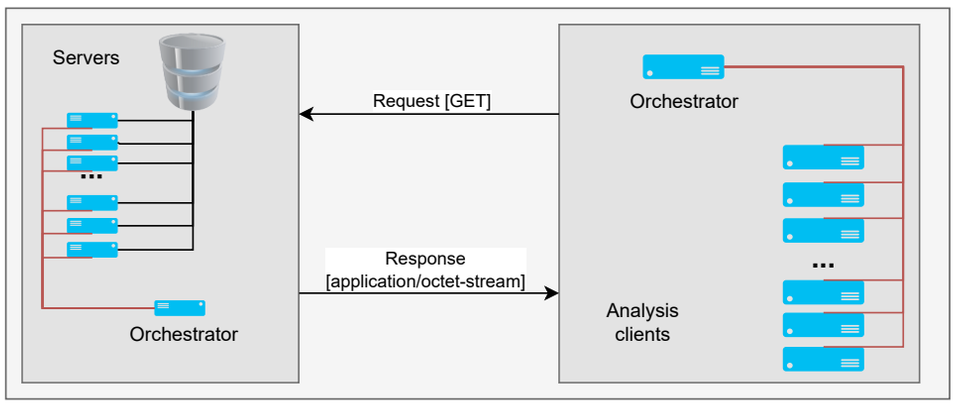

Figure 3.7: Communication of data with an exposed API. Both server and client side run in an orchestrated controller-worker fashion using OMPI.

3.3 Implementation

The system relies on OMPI [37] for scattering and gathering of tasks. Each node runs computations in parallel, and tasks are split into time intervals, layer assumptions(node-to-node range) or both.

3.3.1 Preprocessing

The AIS-data is received as one or more files. The raw AIS-data contains excess information, such as name and ship dimensions; these are filtered out, and only information defined in Table 3.8 is kept. The preprocessing pipeline uses the *pandas* [1] data analyst library for structuring the data in tabular data frames.

MMSI: str	date-and-time: iso8601	longitude: float	latitude: float

Figure 3.8: Data preserved after AIS preprocessing.

The preprocessing step of the pipeline in Figure 3.3 stores this information in an assumed shared filesystem. The preprocessing stores tabular data in the Apache *parquet* [50] file format for two reasons: *1.* Parquet allow information

1. https://pandas.pydata.org/

Location	Format	Space (MB)	Relative change
on-disk	csv	840	1
on-disk	parquet	170	0.2
in-memory	dataframe	1836	2.19

Table 3.1: Storage consumption for 18364747 records(one day) CSV, parquet, and in-memory dataframe, CSV as reference.

on data types to be maintained. 2. Parquet is a compressed file format, greatly reducing disk space. As seen in Table 3.1, parquet only requires 20% of the disk space compared to CVS, while in-memory objects require over twice the memory for ≈18 **M** records.

3.3.2 Layering

To reduce the search space for each node when searching for a *path to shore*, the nodes are segmented into different layers based on their distance to shore. To facilitate this, the positional data for the coastline of Norway [45] is structured into a VPTree. This allows us to easily query for "*how far from shore is this node*" for every node in the graph by traversing the vantage point tree for a node's nearest |neighbour. This yields the point along the coastline closest to the node, and we can solve the inverse geodesic problem [41] to find the distance in *meters*. This is done for *all* unique records in the preprocessed data and appended to the data in a new column 'layer'. For this, we use Rickard Sjögren's Python module: vptree [2] combined with Charles Karney's geographiclib module [3].

Granularity

The system does not operate on infinite time granularity, and in segmenting the data, a 'timestep' for the simulation is introduced. This is the lower granularity with which the system samples the data. The first record is kept if there is more than one record within the sampling frequency.

3.3.3 Converting to trees

Tabular data is great for storing columnar data, but we would like a transparent way to retrieve and query for NN at arbitrary times in our simulation. We can

2. https://github.com/RickardSjogren/vptree
3. https://geographiclib.sourceforge.io/html/python/index.html

transform the tabular AIS data into vantage point trees with one VPTree per layer, per timestep. With the tabular data transformed, we need an intuitive way to retrieve VPTrees that interest us. The trees are stored in key-value storage indexed by date and boat-to-boat range and served through an API from several servers, see: 3.3.5.

3.3.4 Search

The search follows the pattern of a DFS or BFS to traverse the layers, in the worst case traversing all nodes in the network, which is unlikely given the distribution, as seen in Figure 2.10 and Figure 4.4.

Before looking at the algorithms 1 and 2 which shows the implemented searches, we must define the operations and functions we will use, but not explicitly outline.

The Definition 2. We define ψ as a destructive operation returning the first element *e* from a set *s*.

The Definition 3. We define *connected* as a function returning a boolean value if the input object has a path to the shore.

The Definition 4. We define *neighbours* as a function returning all neighbouring boats in a vantage point tree within a range r from a position p.

Algorithm 1 Aktan Depth-First Search

Input Position p, VP Trees t, Range r, Layer l, Set v, Set s
Output $\sum e \in p$ if $p \leftarrow path(p, shore) \neq \emptyset$ else 0

```
procedure DFS(p,t,r,l,v,s)
    if p ∈ v then
        return 0
    end if
    if p ∈ s then
        return s.get(p)
    end if
    if l == 1 then
        return 1
    end if
    c_s ← {l − 1, l, l + 1}
    for all c ∈ c_s do
        if c ∉ t & c == l then
            Abort
        end if
        n_s ← neighbours(t, p, r)
        for all n ∈ n_s do
            if h ← DFS(n, t, r, c, v, s) ≠ 0 then
                return h+1
            end if
        end for
    end for
    return 0
end procedure
```

Algorithm 2 Breadth-First Search

Input Position p, VP trees t, Range r, Layer l, Set v, Set s, Queue q
Output $\sum e \in p$ if $p \leftarrow path(p, shore) \neq \emptyset$ else 0

procedure BFS(p,t,r,l,v,s,q)
  $v \leftarrow v \cup \{p\}$
  $q \leftarrow q \cup \{(p, l, 0)\}$
  while $q \neq \emptyset$ **do**
    $p_n, l_n, h \leftarrow \psi(q)$
    if $l_n == 1$ **then**
      return h
    end if
    if $p_n \in s$ **then**
      if $connected(p_n)$ **then**
        return $s.get(p_n) + h$
      else
        continue
      end if
    end if
    $l_s \leftarrow [l_n - 1, l_n, l_n + 1]$
    $n_s \leftarrow \{\}$
    for all $l_i \in l_s$ **do**
      if $l_i \notin t$ & $l_i == l$ **then**
        Abort
      end if
      $n_d \leftarrow neighbours(vp, p, r)$
      for all $n \in n_d$ **do**
        if $n \neq p_n$ & $n \notin v$ **then**
          $n_s \leftarrow n_s \cup n$
        end if
      end for
    end for
    $q \leftarrow q \cup n_s$
    $v \leftarrow v \cup n_s$
  end while
  $s \leftarrow s \cup v$
  return o
end procedure

3.3.5 API

The VPTrees do take up a decent amount of data, and for a sampling frequency of *1*-minute one day of trees take up about 1.9 GB. The VPTrees are served on-demand from a set of web servers, and a client can query the web servers on two endpoints:

Path: /tree/{date}/{time}/{range}/{layer}

Method: GET
Description: Retrieve a VPTree for a specific layer at a specific date and time with an assumption of BTB communication range.
Parameters:

- **date** (required): ISO 8601 formatted date
- **time** (required): ISO 8601 formatted time in UTC
- **range** (required): Assumption on BTB communication range
- **layer** (required): Specific layer of interest

Path: /trees/{date}/{time}/{range}

Method: GET
Description: Retrieve all VPTrees for a specific date and time with an assumption of BTB communication range
Parameters:

- **date** (required): ISO 8601 formatted date
- **time** (required): ISO 8601 formatted time in UTC
- **range** (required): Assumption on BTB communication range

The path parameters *date* and *time* follow ISO8601 [51] on the form: *YYYY-MM-DD* and *HH:MM:SS*, setting a lower bound for the granularity for the system at *seconds*. This is an artificial limitation, but as mentioned in Chapter 2, interpolation and extrapolation for achieving sub-second granularity do not come without problems in a SANET. The high degree of freedom for ships combined with sudden changes in trajectory makes predictions difficult.

4

Results

In this chapter, we will present, discuss and evaluate the results from applying Aktan to analyse three weeks of AIS-data from Norway's EEZ. Since the network is highly dynamic, boats may join or leave the network at **any** time; averaging over fixed times such as per day can, therefore, skew the results negatively.

4.1 Analysis

In this analysis, the focus is *being able to reach shore* for continuous connectivity to the global network. The range for Boat-to-Boat (BTB) hops is up to *110* km, and ship-to-shore communication is set to *50* km. Ideally, we want to have as low a distance between the boats as possible to increase bandwidth, as discussed in section 2.1.2.

4.1.1 Network layering

We start by looking at where the boats are located. Figure 4.1 shows the layering of boats throughout the whole time series of three weeks. Unfortunately, high-resolution vector graphics of this take up about 4 GB and cannot be included. Every boat position is plotted, and the colour is selected based on what layer the boat is in. This shows that boats visit large parts of the EEZ throughout

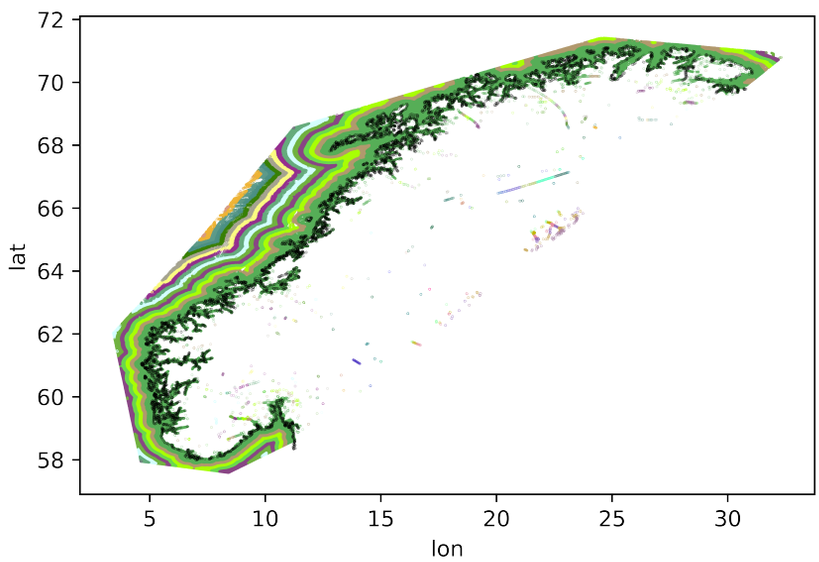

Figure 4.1: Overview of boat position between 2023-06-01, 2023-06-20. Colours represent different layers; in-land and Swedish boats are shown, and these are filtered out in the rest of the analysis.

the three weeks.

The topology is sparse, as seen in Figure 2.10. By including infrastructure and changing the research question to ship-to-shore connectivity, looking at boats' vertical distribution out from shore becomes more interesting.

To investigate this, the layer information is analysed for each day, plotting what layers have and do not have boats. As seen in Figure 4.2, with a 5 km BTB hop assumption, there are boats in all layers until 95 km from shore. The network will have many empty and isolated layers with this assumption for boat-to-boat communication range. With an increased BTB range of 20 km, boats are present in all layers except for the 13th and 17th of June as seen in Figure 4.3. The blue lines mark sequential layers with boats in them, blue dots mark layers with isolated boats, and red dots mark layers without boats.

Figure 4.4 shows a snapshot of the first minute of analysis. Y-axis is latitude, x-axis is longitude, blue marks the Norwegian coastline, and orange dots mark boat positions. Boats are spread over the entire coastline of Norway. Denser clusters are observed near shore and in the southern and western regions.

Boats inside the first layer (less than 50 km from shore) are expected to be

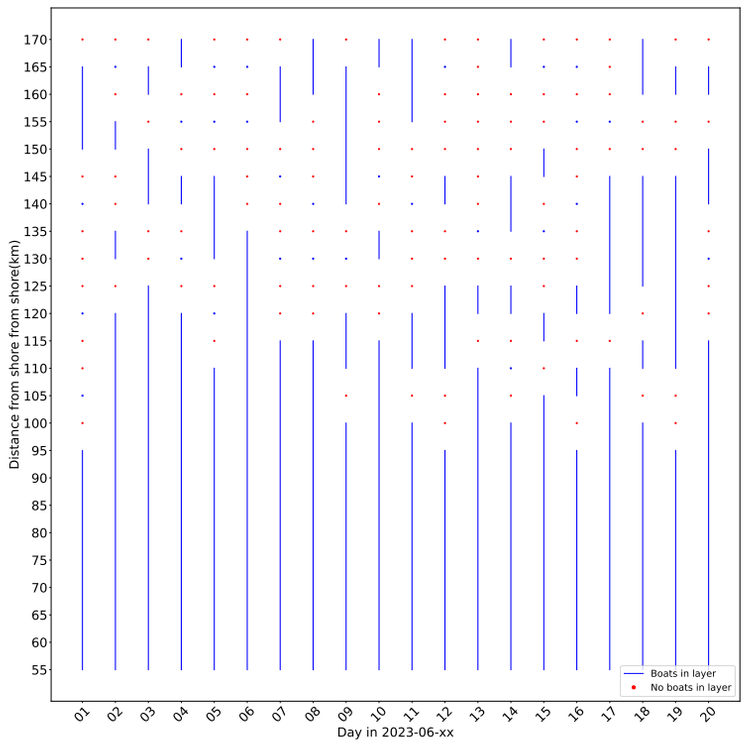

Figure 4.2: Vertical distribution of boats, 5 km BTB range

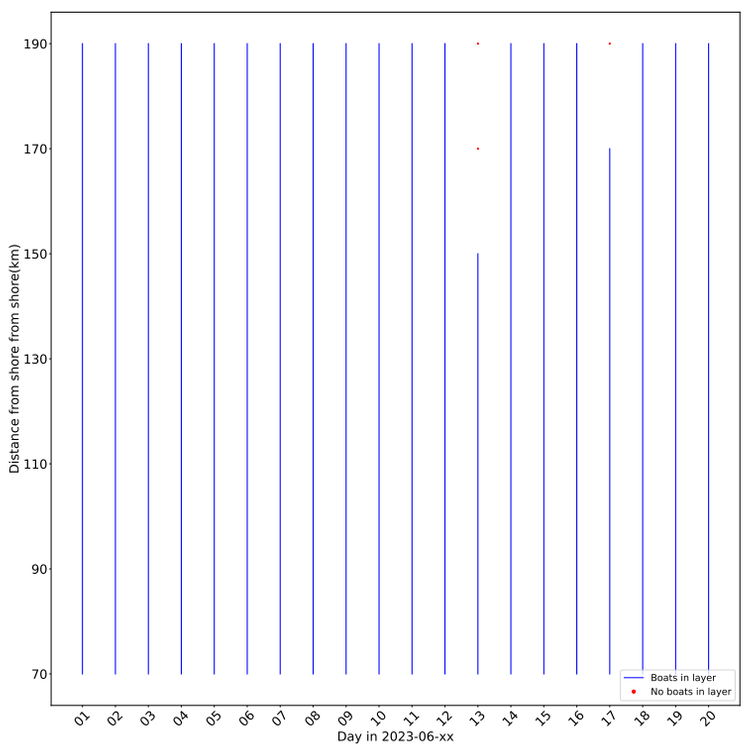

Figure 4.3: Vertical distribution of boats, 20 km BTB range

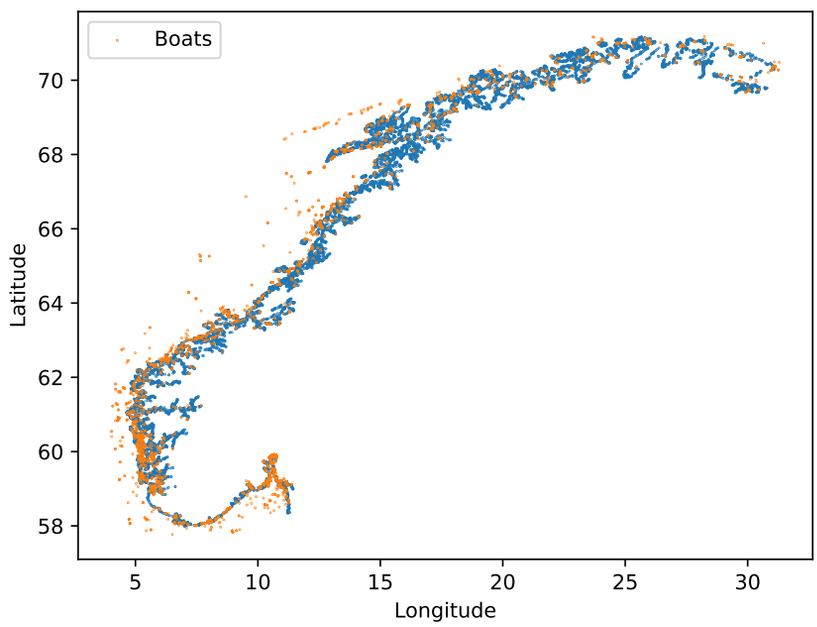

Figure 4.4: Boats distribution, first minute of simulation.

able to reach on-shore infrastructure directly. As seen in Table 4.5 a total of *14510* unique boats are observed in the three weeks the dataset spans. Of these boats, 14349 travel within the first layer within three weeks. This makes up 99% of the network; these boats will have a direct connection with the on-shore infrastructure sometime within these three weeks. Further, 1715 boats sail in any of the outer layers(layer $>$ 1, more than 50 km from shore) and will require multi-hop routing to reach on-shore infrastructure. 12785 (88%) of the boats are *never* outside layer 1, these boats will never require multi-hop routing. 161 (1.1%) of the boats are never inside layer 1 and require multi-hop routing to reach on-shore infrastructure.

Out of this *two* important observations are made.

- 1. Most boats do not need multi-hop routing of data.
- 2. 1715 boats require multi-hop routing. These are the boats we will analyse further.

In layer	decimal (count)
All	1 (14510)
1	0.99 (14349)
> 1	0.12 (1715)
Never in layer	
1	0.011 (161)
> 1	0.88 (12795)

Figure 4.5: Boat distribution for first layer(one hop to shore) and layer > 1(multi-hop), assuming a *50* km shore-to-boat range.

4.1.2 Connectivity

Boats within *layer 1* are of little interest, as they are assumed to reach the global network directly via on-shore infrastructure. As we saw in Table 4.5, there are *1715* boats that travel outside the first layer, these are the boats we investigate further.

Connected boats

We have investigated the network's topological characteristics. To answer whether MANETs can be leveraged for end-to-end communication, we must examine whether the topology supports link establishments to shore. We investigate the overall connectivity for each unique boat. Figure 4.6's y-axis is the percentage of time a boat can establish a path to shore, the x-axis is the boat ID. Blue dots mark connectivity for a boat, and the red dotted line marks the global average for the network. The average connectivity is calculated as the sum of timesteps a boat can establish a path to shore over the total number of timesteps the boat sees in the network. The network's connectivity is low; the global average over the three weeks is only 29%.

The data shows that there are big connectivity differences between the layers. In Figure 4.7, we look at each layer separately and calculate the daily average connectivity for all boats in the layer. On the y-axis, we have the average connectivity per day. The blue line is connectivity for layer 4, the red line is connectivity for layer 3, the green line is connectivity for layer 2, and the maximum is included in the legend as 7. Layers 4 and up achieve effectively no connectivity.

Layer 4, as seen in the figure, does have some very small percentage connectivity, but after rounding to 2 decimal points, this comes out as 0. Each day has $60 \times 24 = 1440$ timesteps with a sample frequency of 1 minute. Rounding with

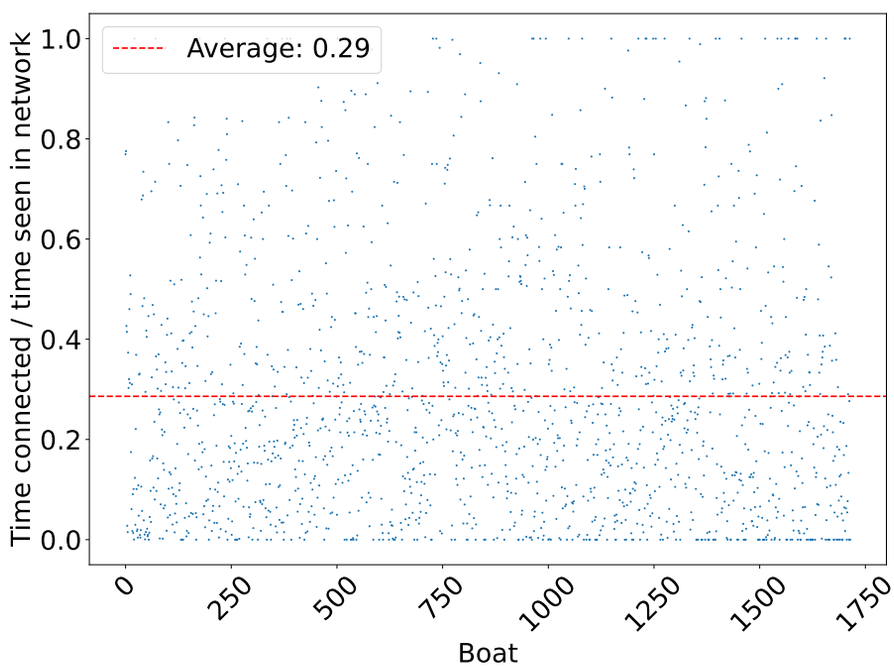

Figure 4.6: Average connectivity and global average for each of the 1715 boats outside layer 1.

two decimal points to 0 requires less than $1440 \times 0.005 \approx 7$ paths per day.

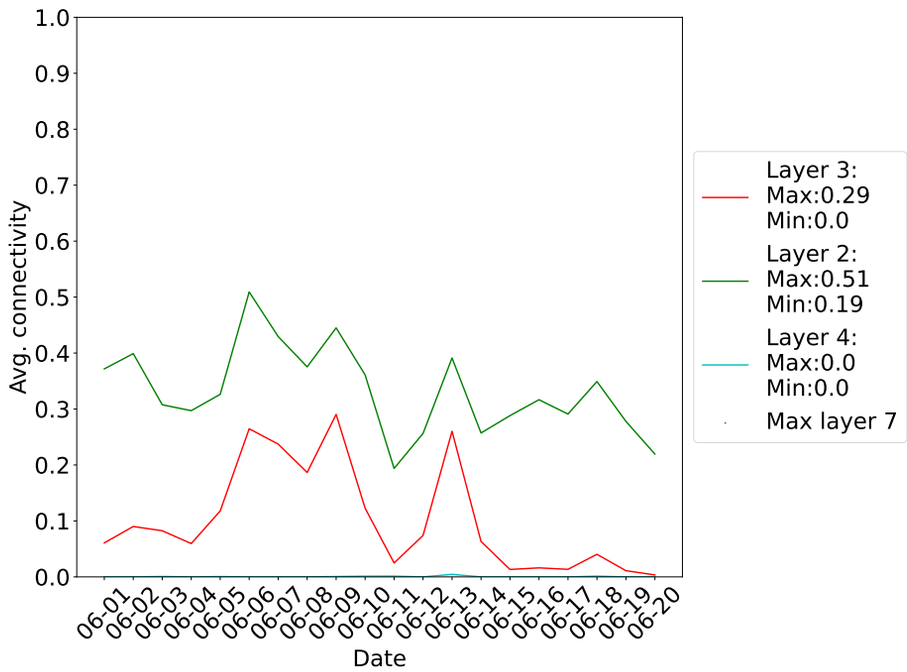

Figure 4.7: Average connectivity per layer, per day. Note: layer 4 is reported as 0.0, meaning it achieves less than 7 minutes of daily connections.

We have seen that the average connectivity is very low in the network, but how many minutes does a boat have a connection to the network? We can retrieve this information by calculating the sum of all minutes a boat can achieve a path to shore. Figure 4.8 has a total time connected on the y-axis, the x-axis depicts the day, the blue dots mark the connection time for a boat on a day, and the red line is the global average for each day. There is a high variance between the boats, with some boats having a connection of more than 440 minutes out of the 1440 in a day. This graph does not show how much this makes up a boat's total time in the network. Is 440 a lot? It is only $\frac{440}{1440} = 0.30$ per cent of a day, but if this is all the time the boat is 'online,' then this is great.

To account for boats' highly dynamic network time, we have examined how much a boat's connected time makes up for its time in the network. While Figure 4.6 shows the boat's average and global average, it does so over the whole time period. Figure 4.9 better evaluates this for each boat by looking at the daily connectivity and shows that none of the days achieve more than 40% connectivity in the network overall. In Figure 4.9 y-axis, we have the boat's

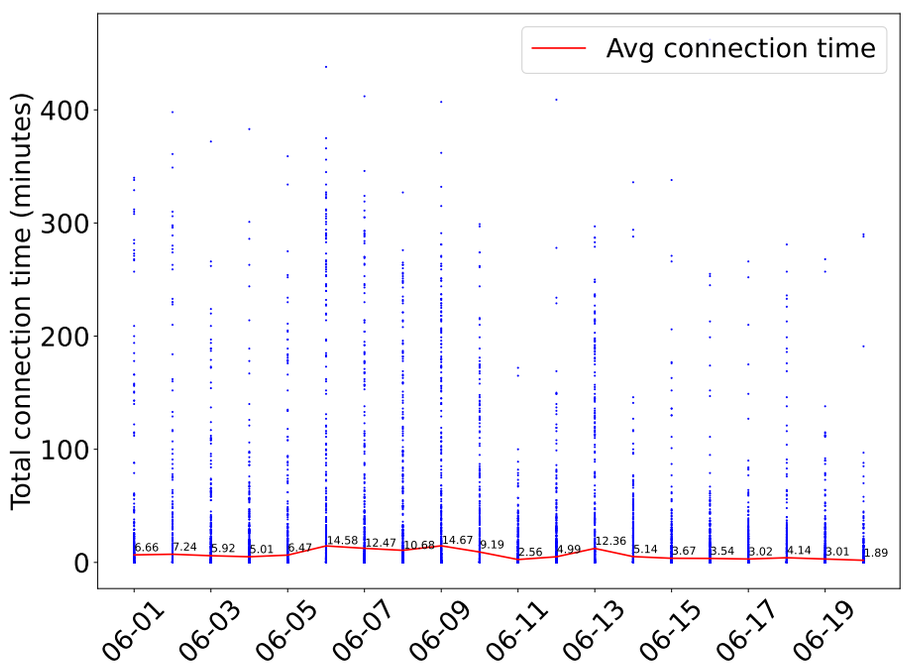

Figure 4.8: Average connection time over the three weeks. With a sampling frequency of 1 minute, averaged each day.

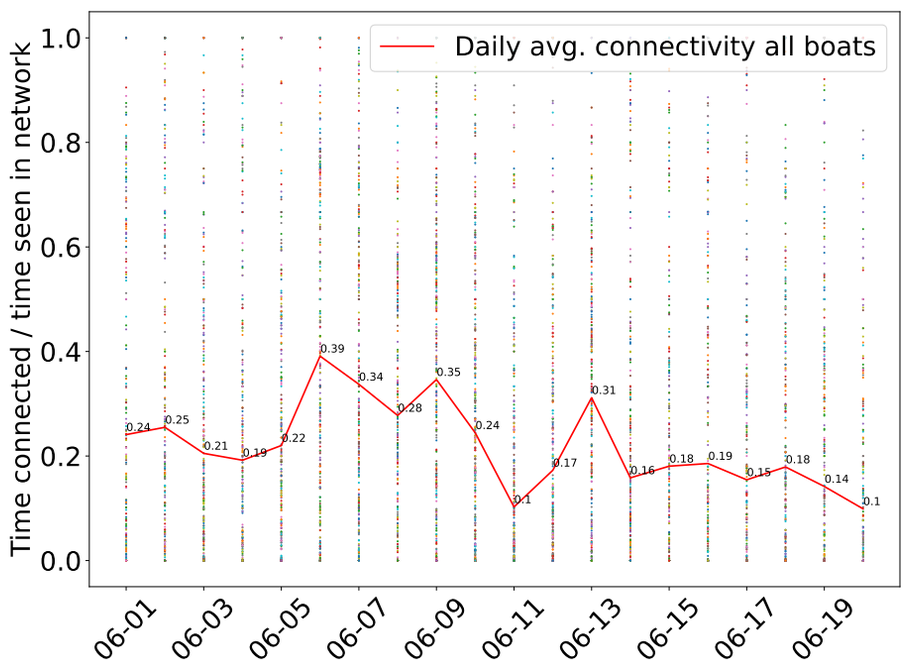

Figure 4.9: Daily connectivity per boat and daily global average.

time connected over time seen in the network; on the x-axis days, coloured dots are this relation for each boat, and the red line is the daily average for all boats.

We further analyse how many boats can achieve connectivity above 99% of the time visible in the topology. We do this by filtering out the connected boats that have connected time divided by time seen in the network above 99 ($\frac{c_t}{t_s} > 99$). Of the 1715 boats analysed, only 55 manage a connection time above 99%. As seen in Figure 4.10, some have quite substantial time in the network, but they are all found in the *second* layer. Note: this does not scale linearly, there will not be 110 boats for 98% and so one, as seen in Figure 4.9.

Knowing that we have very limited time to send data, how much time does a boat have on average per achieved link? We calculate this similarly as for total connection time. For each boat, we find the length of sequential connectivity divided by the sum of a boat's uptimes. In Figure 4.11 we have this ratio on the y-axis, boat ID on the x-axis, blue dots mark the ratio for a boat, the red dotted line marks the average, but as we have quite some outlines, we also plot a green dotted line for the median value.

This data must traverse at least 1 intermediate boat, but how many can we

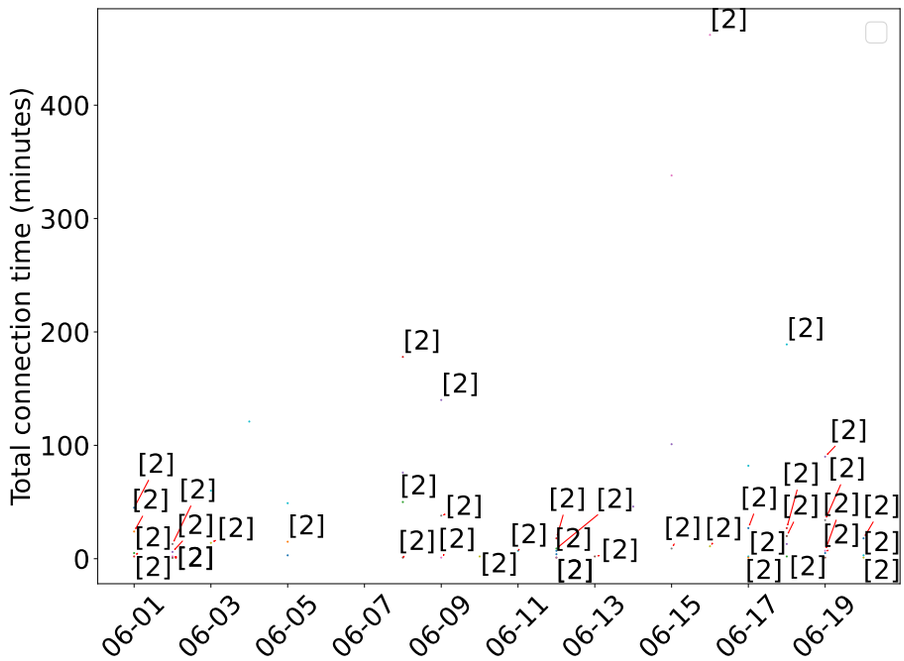

Figure 4.10: boats with more than 99% connection time, annotated with layer.

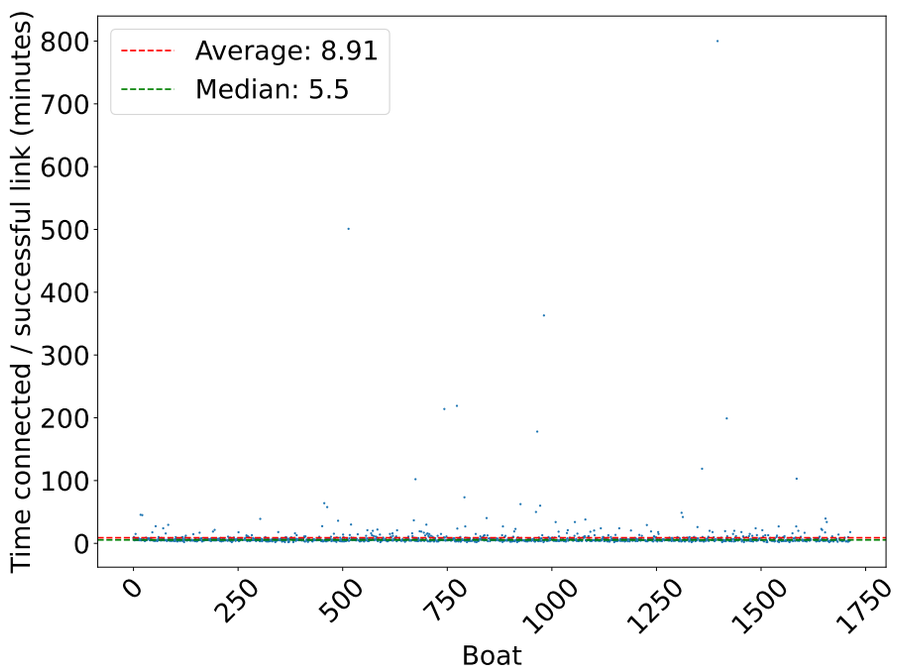

Figure 4.11: Time connected per achievable link.

expect to traverse? We find this by counting the number of hops for each connection to shore. Figure 4.12 shows this, with the average number of hops on the y-axis, days on the x-axis. Since only layers 2 and 3 have connectivity, these are the only ones included; even though layer 4 had a few paths in the middle of the time series, this has been left out as it provides little value. The orange line shows the average number of hops taken for boats in layer 2. The black dots are the minimum hops required for layer 2, and the red dots mark the maximum required for layer 2. The blue line shows the average number of hops for boats in layer 3; the yellow dot is the minimum required, and the green dots are the maximum.

An important thing to remember is that this DFS is not *optimal*, so even though it gives some indication of the length of the paths in the network, these paths are not guaranteed to be the shortest paths, and can only be greater or equal to the shortest path.

For most of the three weeks, both layers had fewer than 10 hops to shore. An unexpected pattern is shown, and boats in both layers seem to need longer paths to reach shore in the middle of the time period. The reason for this increase has not been investigated, but it is something to look into. Are there vents causing this, like changes in fishing areas, cargo or tourism or perhaps search and rescue operations impacting this?

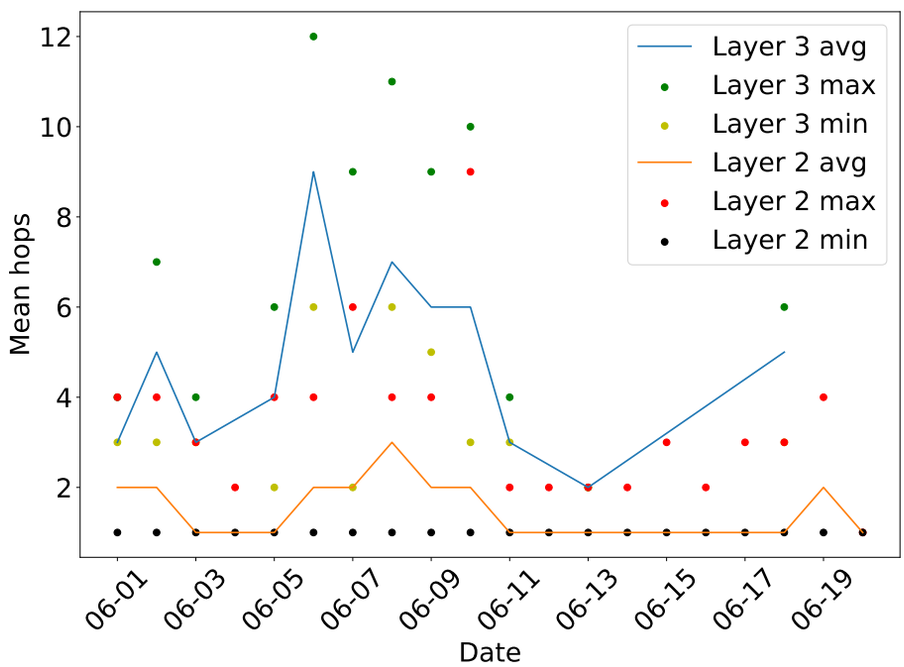

Figure 4.12: Average, max and min hops needed for connected boats, per day.

Never connected

Until now, we have looked at the 1715 boats that travel further than 70 km from shore and focused on those who can achieve connection to shore. In this set, 138 boats are *never* able to reach shore through any path. It's interesting to see where these are, are they isolated in far away regions, are they clustered together or spread out in small pockets without connectivity? To investigate this, Figure 4.13 gives us a visual view on the positions throughout the three weeks that these boats are in. To the authors' surprise, many of them are found in areas of the region where we expected high connectivity due to the number of boats in these areas, except one outliner far up in the Barents Sea.

Figure 4.13 shows never-connected boats outside the coast of Norway, Figure 4.14 shows the neighbourhoods that these boats create, and Figure 4.15 shows the positions of isolated boats(grey dots), over connected boats(green dots). These plots show that from outside Møre og Romsdalen in the south, through the sea outside Helgeland to the outside of Lofoten, scattered small neighbourhoods of boats are found that cannot reach the global network via multi-hop MANET routing. Interestingly, larger and denser neighbourhoods of boats never connected are found in areas where connected boats are also found, suggesting that the network has temporal dependencies.

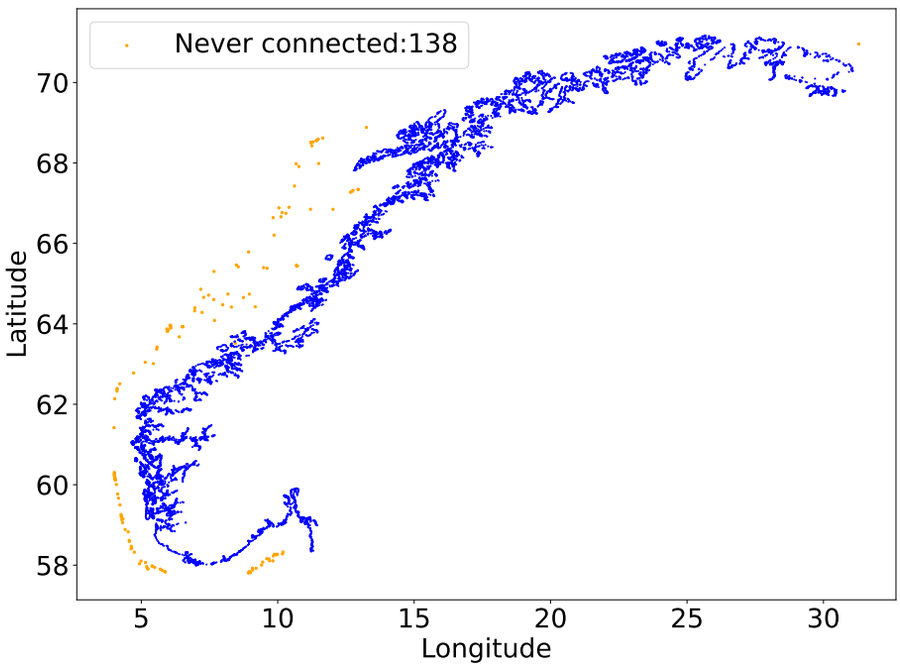

Figure 4.13: boats that never have a connection with the global network.

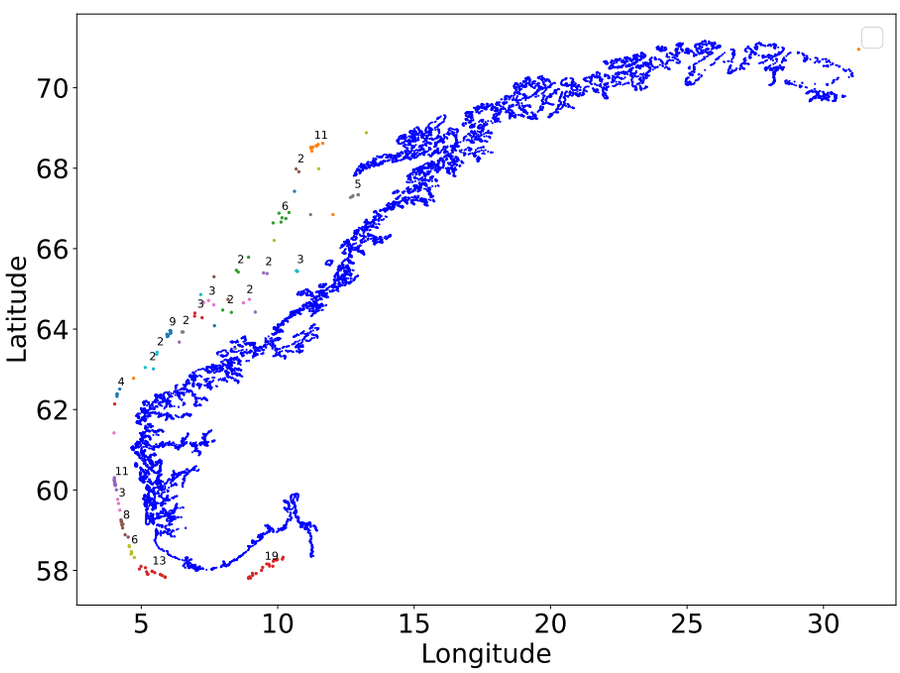

Figure 4.14: Neighbourhoods for the boats never connected to the global network.

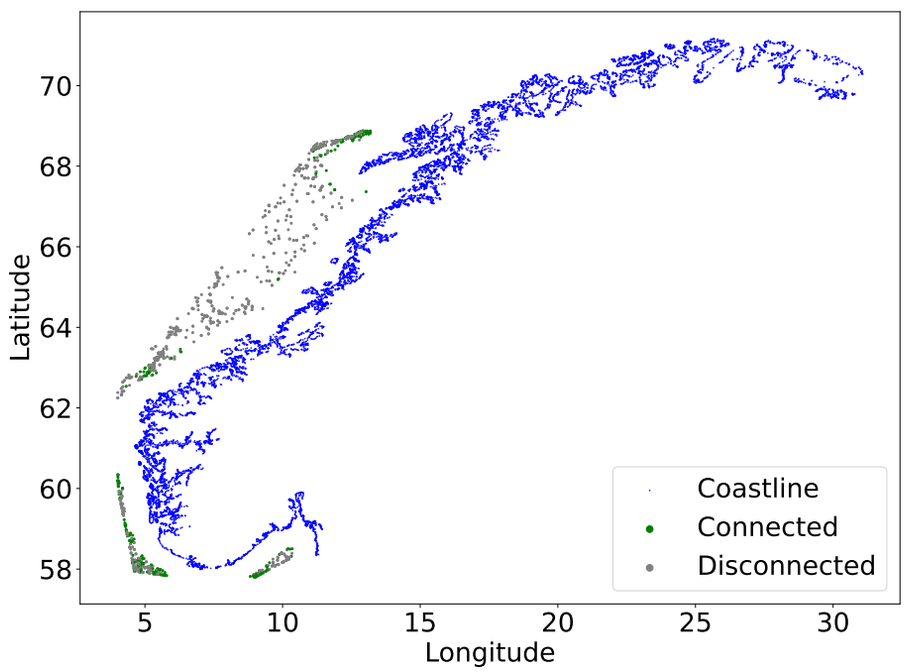

Figure 4.15: Position for disconnected over connected boats.

4.2 Increasing boat-to-boat range

Until now, we have only looked at 5 and 20 km boat-to-boat ranges with a fixed 50 km shore-to-boat range. We provide insight into alternative configurations by increasing the boat-to-boat communication range.

In Figure 4.16 we see the time boats are connected over the time they are seen in the network for each day with a 30 km boat-to-boat range. We see a 20% increase in connectivity, compared to 20 km boat-to-boat range. When increasing this range to 70 km, we see an increase by over 60% average connectivity, with average levels never falling below 80%. Even with 70 km range assumptions, some boats do not achieve connectivity.

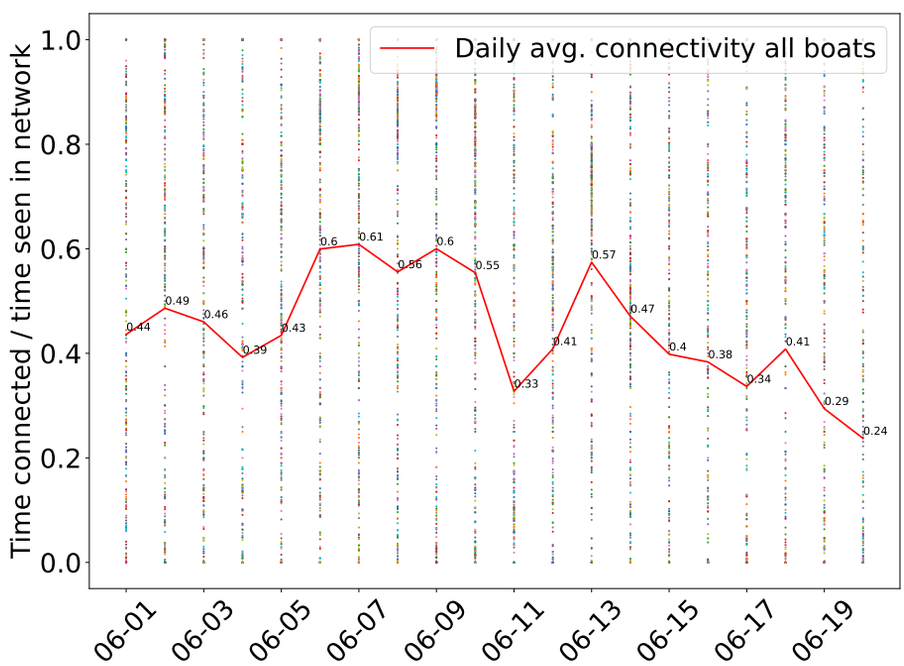

Figure 4.16: Daily connectivity per boat and daily global average with 30 km boat-to-boat range.

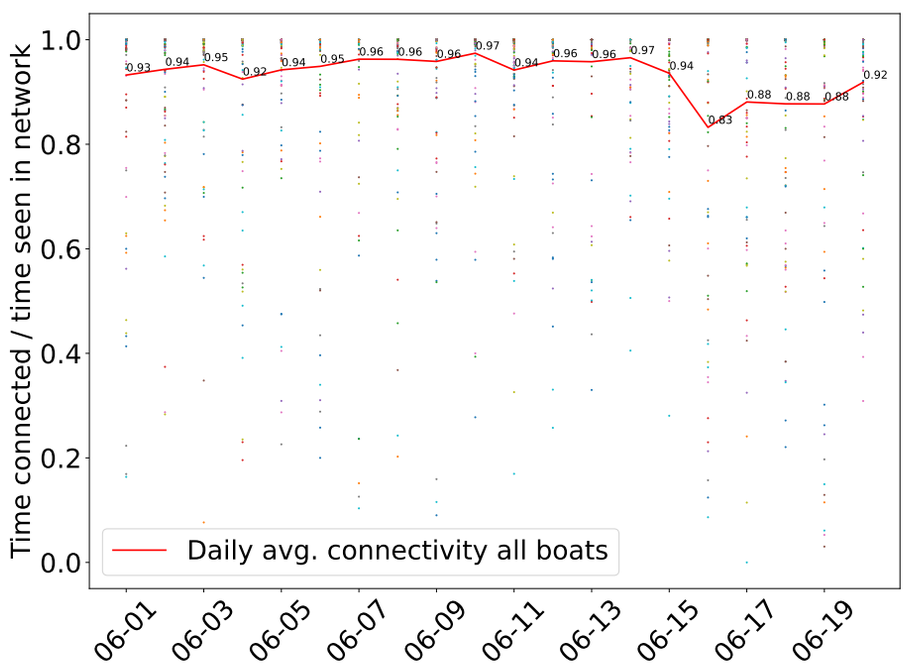

Figure 4.17: Daily connectivity per boat and daily global average with 70 km boat-to-boat range.

5

Conclusion

In this chapter, we want to combine the results from the Aktan analysis with some related work. We interpret the results in the context of our research questions, derived from our overarching thesis statement, and then summarise with concluding remarks.

5.1 Research questions

We will go through each research question individually and interpret the relevant results to answer it.

Thesis statement: Multi-hop MANET routing enables continuous connectivity to on-shore infrastructure for boats in Norway's exclusive economic zone.

Question 1: "Given constraints on communication range in Norway's EEZ, what level of connectivity is achievable?"

Based on the assumptions from related work [9, 10, 11], the average level of connectivity for any day is above 40%, and overall throughout the three weeks, this is reduced to 29%. This being said, as there are shortcomings such as no inter- and extrapolation of data in Aktan in its current form, these numbers might be skewed negatively. The large area with boats that never achieve

connection does, however, indicate that multi-hop MANET routing using boats *alone* might not be sufficient. Drones have been studied in [52] and have the potential to dramatically increase a boat's communication range, although for shorter periods of time. This and equipping boats with satellite capabilities or deploying supply boats acting as routers have the potential to mitigate some of the problems observed in this analysis.

Question 2: "Is the level of connectivity in this network within the range that one would expect for continuous connectivity?"

There is no one answer to this question, and it highly depends on what we define as continuous connectivity and whether we require *all* boats to have it simultaneously. For some boats, the level of connectivity reaches close to 100% of their time online in the network, while other boats do not achieve connection at all, with averaging numbers below 40%. Further analysis should be conducted before one dismisses multi-hop MANET routing as an alternative to achieve continuous connectivity in this region.

Question 3: "Can AIS positional data provide information to model and simulate the network topology?"

Throughout building Aktan and analysis of the results, AIS proved adequate for *some* modelling and simulation. The positional data, time and unique IDs are everything we need for high level analysis, and while AIS can broadcast vessel dimensions, it is not included for all types of AIS units or all message types.

However, it comes with two severe problems that *must* be considered when using it for this type of modelling: 1. The AIS broadcasts can happen at any time without synchronisation between the boats and may be turned off by the crew. 2. AIS is not required by all types of vessels and will, therefore, not show the complete potential network topology.

This may result in boats being missing in parts of the dataset. Analysis on top of this data *must* be aware that without inter-, extrapolation, filtering, or care when analysing, this may have a big impact on the results, i.e. boats in close proximity may broadcast within different timesteps and appear "missing" when one looks at connectivity analysis.

Question 4: "Can AIS positional data identify regions of interest to improve overall network performance?"

Based on the AIS data, we identified regions where boats seemed isolated. Boats in some of these areas did achieve connectivity, indicating that the network has temporal dependencies. Cluster analysis in the same regions showed that the boats form about 40 neighbourhoods. These are regions where infrastructure could remove the temporal dependency and potentially connect these regions. Future work should investigate this.

5.2 Future work

By using Aktan and interpreting the results from the analysis, we observed some things that could be improved. Since Aktan is a discrete-time analysis tool that does not offer inter- and extrapolation of data points, it has one important weakness: boats that are in close proximity but have interleaving broadcast timesteps, the connectivity analysis will fail to report these boats as present in the proximity. For example, we have 3 boats, b_1,b_2, and b_3, that report their position every 3 minutes, b_1 starting at t_1, b_2 at t_2, and b_3 at t_3. When we run path searches in t_1, boats b_2 and b_3 are not included; similarly, in t_2, b_1 and b_3 are not included. This means that the results may be skewed negatively.

Although inter- and extrapolation potentially solve this problem, it does not come without its own problems. Boats have a high degree of freedom, and the deviation may become large for less frequent sampling. This problem is highly situational, and therefore, we suggest that Aktan is extended to support inter/extrapolation based on user-defined sampling frequencies and let the end-user decide whether it is sufficient or if further action is required.

When inspecting the boats that were not connected, we saw that the network had temporal dependencies and the isolated nodes formed clusters. This brought up the idea of extending Aktan to model a new type of mobile node with global connectivity, i.e. on-demand placed infrastructure such as supply boats with satellite connectivity or drones. This would allow us to investigate infrastructure deployment and its impact on a MANET's capability to enable continuous connectivity.

5.3 Concluding remarks

IUU fishing is a global problem that is highly problematic for ecosystems and costs society billions of US Dollars every year. New regulations, monitoring and surveillance systems are being designed and implemented by governments around the world. Many of these systems require connectivity to the global network, but the maritime sector is a challenging sector to achieving this.

This thesis has investigated the viability of MANET routing enabling end-to-end communication in Norway's EEZ using AIS data over a time period of three weeks. Based on previous work, we propose a novel simulation tool, Aktan, for discrete-time analysis on AIS data.

Aktan is written in Python and uses MPI to distribute tasks to compute nodes. The system is modular by design and uses snapshots to store intermediate computations, allowing modules to run independently and previously modelled systems to be loaded precomputed. This allows multiple analyses to be run simultaneously or re-iteration of the same experiment without the need to recreate the models.

Our analysis shows that with constraints in communication range, multi-hop MANET routing alone does not perform particularly well in this region but might be viable with fleet management or in combination with other technologies such as satellites and drones.

Bibliography

[1] Christopher Costello et al. “The future of food from the sea.” In: *Nature* 588.7836 (2020), pp. 95–100.

[2] Hans-Martin Straume, Ursula Landazuri-Tveteraas, and Atle Oglend. “Insights from transaction data: Norwegian aquaculture exports.” In: *Aquaculture Economics & Management* 24 (Oct. 2019), pp. 1–18. DOI: 10.1080/13657305.2019.1683914.

[3] FAO. *IUU day|Illegal, Unreported and Unregulated (IUU) fishing*. URL: https://www.fao.org/iuu-fishing/fight-iuu-fishing/en/.

[4] Industry Ministry of Trade and Fisheries. *Nou 2019: 21*. Accessed: 2024-06-02. Nov. 2024. URL: https://www.regjeringen.no/no/dokumenter/nou-2019-21/id2680187/?ch=1.

[5] Mohamed M. Kassem et al. “A browser-side view of starlink connectivity.” In: *Proceedings of the 22nd ACM Internet Measurement Conference*. IMC ’22. New York, NY, USA: Association for Computing Machinery, 2022, 151–158. ISBN: 9781450392594. DOI: 10.1145/3517745.3561457. URL: https://doi.org/10.1145/3517745.3561457.

[6] Ogutu B. Osoro and Edward J. Oughton. “A Techno-Economic Framework for Satellite Networks Applied to Low Earth Orbit Constellations: Assessing Starlink, OneWeb and Kuiper.” In: *IEEE Access* 9 (2021), pp. 141611–141625. DOI: 10.1109/ACCESS.2021.3119634.

[7] Fiskeridirektoratet. *The active fishing fleet*. Accessed: 2024-06-02. Apr. 2024. URL: https://www.fiskeridir.no/Yrkesfiske/Tall-og-analyse/Fiskere-fartoey-og-tillatelser/Fartoey-i-merkeregisteret.

[8] Peter J. Denning et al. *Report of the ACM Task Force on The Core of Computer Science*. Tech. rep. New York, NY, USA, 1988.

[9] Sethuraman N Rao, Maneesha Vinodini Ramesh, and Venkat Rangan. “Mobile infrastructure for coastal region offshore communications and networks.” In: *2016 IEEE Global Humanitarian Technology Conference (GHTC)*. 2016, pp. 99–104. DOI: 10.1109/GHTC.2016.7857266.

[10] Sethuraman N. Rao, P. Venkat Rangan, and Maneesha Vinodini Ramesh. *US20170230841A1 - mobile infrastructure for coastal region offshore communications and Networks*. URL: https://patents.google.com/patent/US20170230841A1/en.

[11] Simi Surendran et al. “Link Characterization and Edge-Centric Predictive Modeling in an Ocean Network.” In: *IEEE Access* 11 (2023), pp. 5031–5046. DOI: 10.1109/ACCESS.2023.3235387.

[12] Øyvind Arne Moen Nohr. “Characterizing Boat-to-Boat Communication in The Norwegian Economical Sector Using The Ocean Communication Network.” In: *UNPUBLISHED* (2023).

[13] G. Hartvigsen and D. Johansen. “Stormcast — A Distributed Artificial Intelligence Application for Severe Storm Forecasting.” eng. In: *IFAC Proceedings Volumes* 21.12 (1988), pp. 99–102. ISSN: 1474-6670.

[14] Joakim Aalstad Alslie et al. “Áika: A Distributed Edge System for AI Inference.” In: *Big Data and Cognitive Computing* 6.2 (2022). ISSN: 2504-2289. DOI: 10.3390/bdcc6020068. URL: https://www.mdpi.com/2504-2289/6/2/68.

[15] Tor-Arne S. Nordmo et al. “Dutkat: A Multimedia System for Catching Illegal Catchers in a Privacy-Preserving Manner.” In: *Proceedings of the 2021 ACM Workshop on Intelligent Cross-Data Analysis and Retrieval*. ICDAR ’21. New York, NY, USA: Association for Computing Machinery, 2021, 57–61. ISBN: 9781450385299. DOI: 10.1145/3463944.3469102. URL: https://doi.org/10.1145/3463944.3469102.

[16] Aril Bernhard Ovesen et al. “File System Support for Privacy-Preserving Analysis and Forensics in Low-Bandwidth Edge Environments.” In: *Information* 12.10 (2021). ISSN: 2078-2489. DOI: 10.3390/info12100430. URL: https://www.mdpi.com/2078-2489/12/10/430.

[17] James F Kurose. *Computer networking : a top-down approach*. eng. Boston, 2017.

[18] J.A. Stine and David Portigal. “Spectrum 101: An Introduction to Spectrum Management.” In: (Mar. 2004), p. 220.

[19] Mohammed Banafaa et al. “6G Mobile Communication Technology: Requirements, Targets, Applications, Challenges, Advantages, and Opportunities.” In: *Alexandria Engineering Journal* 64 (2023), pp. 245–274. ISSN: 1110-0168. DOI: https://doi.org/10.1016/j.aej.2022.08.017. URL: https://www.sciencedirect.com/science/article/pii/S111001682200549X.

[20] Jhihoon Joo, Hong-Jong Jeong, and Dong Seog Han. “Verification of Fresnel Zone Clearance for Line-of-sight Determination in 5.9 GHz Vehicle-to-Vehicle Communications.” In: *Wireless Personal Communications* 101 (2018), pp. 239–249.

[21] *International Maritime Organization*. Accessed: 2024-06-02. Dec. 2015. URL: https://www.imo.org/en/OurWork/Safety/Pages/AIS.aspx.

[22] *Telenor Norge*. Accessed: 2024-06-02. URL: https://www.telenor.no/dekning/#dekningskart.

[23] *Ice Communication Norge*. Accessed: 2024-06-02. URL: https://www.ice.no/dekning/kart/.

[24] *Telia Norge AS*. Accessed: 2024-06-02. URL: https://www.telia.no/nett/dekning/.

[25] Sethuraman N Rao et al. "Realizing cost-effective marine internet for fishermen." In: *2016 14th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*. 2016, pp. 1–5. DOI: 10.1109/WIOPT.2016.7492904.

[26] *Ubiquiti*. Accessed: 2024-06-02. Nov. 2023. URL: https://www.ui.com/.

[27] Walter Bislin. URL: http://walter.bislins.ch/bloge/index.asp?page=Advanced+Earth+Curvature+Calculator.

[28] *Ubiquiti ispdesign*. Accessed: 2024-06-02. Apr. 2024. URL: https://ispdesign.ui.com/#.

[29] Sunil Bhooshan. *Fundamentals of Analogue and Digital Communication Systems*. eng. 1st ed. 2022. Vol. 785. Lecture Notes in Electrical Engineering. Singapore: Springer, 2021. ISBN: 9811642761.

[30] *NATO*. Accessed: 2024-06-02. URL: https://shipping.nato.int/nsc/operations/news/2021/ais-automatic-identification-system-overview.

[31] Daichi Hirahara. "Simulation Results of Satellite AIS when Utilizing Khatri-Rao (KR) Product Array Processing." In: *2020 International Symposium on Antennas and Propagation (ISAP)*. 2021, pp. 165–166. DOI: 10.23919/ISAP47053.2021.9391136.

[32] Robert Sedgewick. *Algorithms in C : Part 5 : Graph algorithms*. eng. Boston.

[33] Wenfeng Xia et al. "A Survey on Software-Defined Networking." In: *IEEE Communications Surveys & Tutorials* 17.1 (2015), pp. 27–51. DOI: 10.1109/COMST.2014.2330903.

[34] *Cisco*. Accessed: 2024-06-02. Nov. 2023. URL: https://www.cisco.com/.

[35] Azzedine Boukerche et al. "Routing protocols in ad hoc networks: A survey." In: *Computer Networks* 55.13 (2011), pp. 3032–3080. ISSN: 1389-1286. DOI: https://doi.org/10.1016/j.comnet.2011.05.010. URL: https://www.sciencedirect.com/science/article/pii/S1389128611001654.

[36] Manmohan Sharma et al. "Comprehensive Study of Routing Protocols in Adhoc Network: MANET." In: *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. 2019, pp. 0792–0798. DOI: 10.1109/IEMCON.2019.8936135.

[37] Edgar Gabriel et al. "Open MPI: Goals, concept, and design of a next generation MPI implementation." In: *Recent Advances in Parallel Virtual Machine and Message Passing Interface: 11th European PVM/MPI Users' Group Meeting Budapest, Hungary, September 19-22, 2004. Proceedings 11*. Springer. 2004, pp. 97–104.

[38] Frank Nielsen. "Introduction to MPI: The Message Passing Interface." In: *Introduction to HPC with MPI for Data Science*. Cham: Springer International Publishing, 2016, pp. 21–62. ISBN: 978-3-319-21903-5. DOI:

10.1007/978-3-319-21903-5_2. URL: https://doi.org/10.1007/978-3-319-21903-5_2.

[39] Peter N. Yianilos. "Data structures and algorithms for nearest neighbor search in general metric spaces." In: *Proceedings of the Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '93. Austin, Texas, USA: Society for Industrial and Applied Mathematics, 1993, 311–321. ISBN: 0898713137.

[40] Frank Nielsen, Paolo Piro, and Michel Barlaud. "Bregman vantage point trees for efficient nearest Neighbor Queries." In: *2009 IEEE International Conference on Multimedia and Expo*. 2009, pp. 878–881. DOI: 10.1109/ICME.2009.5202635.

[41] B LOUIS Decker. "World geodetic system 1984." In: *Defense Mapping Agency Aerospace Center St Louis Afs Mo* (1986).

[42] Wikipedia. *World geodetic system*. Accessed: 2024-06-02. May 2024. URL: https://en.wikipedia.org/wiki/World_Geodetic_System.

[43] David Padua. *Encyclopedia of Parallel Computing*. Springer Publishing Company, Incorporated, 2011. ISBN: 0387097651.

[44] Ping Zhang et al. *Model division multiple access for Semantic Communications - Frontiers of Information Technology & Electronic Engineering*. June 2023. URL: https://link.springer.com/article/10.1631/FITEE.2300196.

[45] Natural Earth. *Natural earth " blog archive " coastline - free vector and raster map data at 1:10m, 1:50m, and 1:110m scales*. Nov. 2009. URL: https://www.naturalearthdata.com/downloads/10m-physical-vectors/10m-coastline/.

[46] Shaojun Gan et al. "Ship trajectory prediction for intelligent traffic management using clustering and ANN." In: *2016 UKACC 11th International Conference on Control (CONTROL)*. 2016, pp. 1–6. DOI: 10.1109/CONTROL.2016.7737569.

[47] Petra Virjonen et al. "Ship Movement Prediction Using k-NN Method." In: *2018 Baltic Geodetic Congress (BGC Geomatics)*. 2018, pp. 304–309. DOI: 10.1109/BGC-Geomatics.2018.00064.

[48] Huanhuan Li, Hang Jiao, and Zaili Yang. "Ship trajectory prediction based on machine learning and deep learning: A systematic review and methods analysis." In: *Engineering Applications of Artificial Intelligence* 126 (2023), p. 107062. ISSN: 0952-1976. DOI: https://doi.org/10.1016/j.engappai.2023.107062. URL: https://www.sciencedirect.com/science/article/pii/S0952197623012460.

[49] Charles FF Karney. "Algorithms for geodesics." In: *Journal of Geodesy* 87 (2013), pp. 43–55.

[50] Deepak Vohra. "Apache Parquet." In: *Practical Hadoop Ecosystem: A Definitive Guide to Hadoop-Related Frameworks and Tools*. Berkeley, CA: Apress, 2016, pp. 325–335. ISBN: 978-1-4842-2199-0. DOI: 10.1007/978-

1-4842-2199-0_8. URL: https://doi.org/10.1007/978-1-4842-2199-0_8.

[51] Misha Wolf and Charles Wicksteed. *iso8601*. Accessed: 2024-06-02. Aug. 1997. URL: https://www.w3.org/TR/NOTE-datetime.

[52] Rui Tang et al. "NOMA-based UAV communications for maritime coverage enhancement." In: *China Communications* 18.4 (2021), pp. 230–243. DOI: 10.23919/JCC.2021.04.017.

/6

Appendix A

AIS-DATA	
Information item	Information generation
Static	
MMSI	Set on installation, note that this might need amending if the ship changes ownership
Call sign and name	Set on installation Note that this might need amending if the ship changes ownership
IMO Number	Unique ID of vessel. Set on installation, stay until vessel is scrapped
Length and beam	Set on installation or if changed
Type of ship	Select from pre-installed list
Location of electronic position fixing system (EPFS) antenna	Set on installation or may be changed for bi-directional vessels or those fitted with multiple antennas
Dynamic	
Ship's position with accuracy indication and integrity status	Automatically updated from the position sensor connected to AIS The accuracy indication is approximately 10 m.
Position Time stamp in UTC	Automatically updated from ship's main position sensor connected to AIS
Course over ground (COG)	Automatically updated from ship's main position sensor connected to AIS, if that sensor calculates COG
Speed over ground (SOG)	Automatically updated from the position sensor connected to AIS. **This information might not be available**
Heading	Automatically updated from the ship's heading sensor connected to AIS
Navigational status	Information has to be manually entered by the Officer of the Watch (OOW) and changed as necessary. In practice, since all these relate to the Convention on the International Regulations for Preventing Collisions at Sea (COLREGS), any change that is needed could be undertaken at the same time that the lights or shapes were changed
Rate of turn (ROT)	Automatically updated from the ship's ROT sensor or derived from the gyro. **This information might not be available**
Voyage-related	
Ship's draught	To be manually entered at the start of the voyage using the maximum draft for the voyage and amended as required
Hazardous cargo (type)	To be manually entered at the start of the voyage confirming whether hazardous cargo is being carried, namely: **DG (Dangerous goods)**, **HS (Harmful substances)**, **MP (Marine pollutants)**. Indications of quantities are not required
Destination and ETA	To be manually entered at the start of the voyage and kept up to date as necessary
Route plan (waypoints)	To be manually entered at the start of the voyage, at the discretion of the master, and updated when required
Short safety-related messages	Free format short text messages would be manually entered, addressed either a specific addressee or broadcast to all ships and shore stations

Table 6.1: AIS-DATA specification pr. 2015. See [21] for further details.

AIS frequency per class	
Type of ship	**Reporting interval**
Class A	
Ship at anchor or moored, speed <= 3knots	3min
Ship at anchor or moored, speed > 3knots	10s
Ship 0-14 knots	10s
Ship 0-14 knots and changing course	3 1/3s
Ship 14-23 knots	6s
Ship 14-23 knots and changing course	2s
Ship >23 knots	2s
Ship >23 knots and changing course	2s
Class B	
'SO' shipborne equipment not moving faster than 2 knots	3min
'SO' shipborne equipment moving 2-14 knots	30s
'SO' shipborne equipment moving 14-23 knots	15s
'SO' shipborne equipment moving > 23 knots	5s
'CS' shipborne equipment not moving faster than 2 knots	3min
'CS' shipborne equipment moving faster than 2 knots	30s

Table 6.2: AIS frequency pr. 2015. See [21] for further details

Class 27(LR) AIS-message	
Message ID	Always 27
Repeat indicator	Always 3
User ID	MMSI number
Position accuracy	
RAIM flag	
Navigational status	
Longitude	
Latitude	
SOG	
COG	
Position latency	
Spare	Set to 0

Table 6.3: AIS message type 27