INTERNATIONAL
HELLENIC
UNIVERSITY

Towards a marine opportunistic network across the Mediterranean Sea

Vasiliki Charisi

SID: 3301140010

Supervisor: Dr. Merkouris Karaliopoulos

SCHOOL OF SCIENCE & TECHNOLOGY

A thesis submitted for the degree of

Master of Science (MSc) in Information and Communication Technology Systems

DECEMBER 2015

THESSALONIKI – GREECE

Acknowledgments

The presented thesis was undertaken during the winter semester of 2015. The thesis was submitted for the degree of Master of Science (MSc) in Information and Communication Technology Systems, School of Science and Technology, at the International Hellenic University. Before proceeding to the presentation of the research need for the completion of my dissertation, I would like to acknowledge the people that helped me and contributed in order the dissertation to be completed.

Initially, I would like to thank my supervisor, Dr. Merkouris Karaliopoulos, for the interest and trust he showed to me in assigning the thesis. His advice was particularly valuable for my research and the successful completion of the dissertation. Our frequent and detailed communication helped me to better understand the suggested, in my thesis, problem and the way I should had proceed in order to solve it. The assistance provided, especially during the development of the application's code in Matlab, was very helpful and made me better understand the simulator's nature and its specifications and guided me to the correct developer perception.

Also, I would like to thank my family, who they have always been supporting in everything I have done since now. I would specially like to thank my friends who have always been the receiving end of my complaints and irritation, and never complain.

Abstract

The broader Mediterranean Region is an area where maritime transportation is one of the most important ways of communication. Daily movement of a large number of vessels is the Mediterranean Sea creates the most densely populated marine routed in the world. The goal of this dissertation is to study the feasibility of an opportunistic (aka Delay Tolerant) network in the Mediterranean area that will be leveraging the (semi - deterministic) trajectories of ferries, ships and boats across the sea. AIS data transmitted from vessels for maritime purposes will be used for the network's creation.

The dissertation's motivation is to leverage the mobility of marine nodes across the sea to set up an opportunistic network. The network could serve traffic with no strict delay requirements and it would be helpful in the ever – expanding mobile networks, because there are several regions in the sea, without cellular coverage and satellite communication is much costlier.

The first chapter of the dissertation is an introduction to the basic idea proposed in it. Background information about marine traffic and maritime signal transmission is given. It also includes a brief introduction to the opportunistic networks, the dissertation's objectives and the methodology followed.

The second chapter is dedicated to the network characteristics and metrics that were used in the simulation of the opportunistic network. Inter – contact times, contact graph and contact duration are studied.

Chapter three includes the study of the AIS existing in the data set used. The definitions of the used data, as well as the pre – processing followed before the development of the application. There is also the algorithmic approach of the Matlab application, as well as the analysis of some of the critical points of it.

Chapter four contains the network characterization results, as obtained from the application. The results are taken from different scenarios for the time threshold and the distance. Graphs visualize the results are also presented. Finally, chapter five contains the conclusions drown and the further development toward a full network application.

Table of Contents

1.Introduction ... 1
1.1 Motivation – Background ... 1
1.2 Opportunistic Networks: a Brief Introduction ... 1
1.3 Objectives ... 5
1.4 Methodology ... 6
2. Opportunistic Networks ... 7
2.1 Characterizing an Opportunistic Network ... 7
2.2 Inter – Contact Times ... 7
2.2.1 Pair – Wise Inter – Contact Time ... 8
2.2.2 Aggregate Inter – Contact Time ... 8
2.3 Contact Graph ... 8
2.4 Contact Duration ... 9
2.5 Other Characteristics ... 9
3. Datasets and Methodology ... 10
3.1 AIS Data ... 10
3.1.1 MMSI – Maritime Mobile Service Identity ... 11
3.1.2 Latitude ... 11
3.1.3 Longitude ... 12
3.1.4 SOG – Speed Over Ground ... 12
3.1.5 COG – Course Over Ground ... 13
3.1.6 Timestamp ... 14
3.2 Pre – Processing of AIS Data Trace ... 15
3.2.1 Pre – Processing of Data ... 15
3.2.2 Main Processing of Data ... 16
3.2.2.1 Why Matlab? ... 16
3.2.3 Algorithmic Steps from AIS Data to Contact Traces ... 16
3.3 Code Analysis ... 17
3.4 Baseline Scenario Definition ... 20
4. Characterization Results ... 21
4.1 Scenarios Results ... 23
4.1.1 Baseline Scenario – *dist=1* and *thresh_t=1* ... 23
4.1.1.1 Day1 ... 23
4.1.1.2 Day2, Day3, Day4 – Baseline Scenario ... 29
4.1.2 Scenario1 – *dist=0.5* and *thresh_t=1* ... 40

4.1.3 Scenario2 – *dist=2* and *thresh_t=1* 46
4.1.4 Scenario3 – *dist=1* and *thresh_t=2* 50
4.1.5 Scenario4 – *dist=0.5* and *thresh_t=2* 55
4.1.6 Scenario5 – *dist=2* and *thresh_t=2* 58
4.2 Distributions Comparison 62
5. Epilogue 64
5.1 Conclusions 64
5.2 Further Work 65
Appendix 66
References 72

Table of Figures

Figure 1.1: Message propagation in an opportunistic network from node A to node D over successive node encounters occurring at different time epochs. 2
Figure 1.2: Binary Step of the Spray - and - Wait Protocol 3
Figure 2.1: A Typical Facebook Contact Graph 9
Figure 3.1: Latitude and Longitude on the Globe 11
Figure 3.2: Course Over Ground 13
Figure 4.1: Contact duration day 1 23
Figure 4.2: Histogram of inter - contact times day 1 24
Figure 2.3: Cumulative Distribution Function of ICTvec 25
Figure 4.4: Complimentary Cumulative Distribution of ICTvec with different fits day 1 26
Figure 4.5: Number of ships per number of different contacts day1 27
Figure 4.6: Contact graph created for the baseline scenario day1 28
Figure 4.7: Contact duration day2 29
Figure 4.8: Histogram ICTvec day2 30
Figure 4.9: CDF ICTvec day2 30
Figure 4.10: CCDF ICTvec day 2 31
Figure 4.11: Number of ships per contact day 2 31
Figure 4.12: Contact graph day 2 32
Figure 4.13: Contact duration day 3 33
Figure 4.14: Histogram ICTvec day 3 33
Figure 4.15: CDF ICTvec day 3 34
Figure 4.16: CCDF ICTvec day 3 34
Figure 4.17: Number of ships per contact day 3 35
Figure 4.18: Contact graph day 3 35
Figure 4.19: Contact duration day 4 36
Figure 4.20: Histogram ICTvec day 4 37
Figure 4.21: CDF ICTvec day4 37
Figure 4.22: CCDF ICTvec day 4 38
Figure 4.23: Number of ships per contact day 4 38
Figure 4.24: Contact graph day 4 39
Figure 4.25: Contact duration – Scenario 1 –day 2 40
Figure 4.26: Histogram ICTvec – Scenario 1 – day 2 41
Figure 4.27: CDF ICTvec – Scenario 1 –day 2 42
Figure 4.28: CCDF ICTvec – scenario 1- day 2 43
Figure 4.29: Number of ships per contact – Scenario1 – day 2 43
Figure 4.30: Contact graph – scenario 1- day 2 44
Figure 4.31: Contact duration – scenario 2 –day 2 46
Figure 4.32: Histogram ICTvec –scenario2 –day 2 47
Figure 4.33: CDF ICTvec – scenario 2 –day2 47
Figure 4.34: CCDF ICTvec –scenario2 –day2 48
Figure 4.35: Number of ships per contact – scenario 2 – day 2 48
Figure 4.36: Contact graph – scenario 2 – day 2 49
Figure 4.37: Contact duration – scenario 3- day24 51
Figure 4.38: Histogram ICTvec – scenario 3 –day 2 51
Figure 4.39: CDF ICTvec – scenario 3 – day 2 52
Figure 4.40: CCDF ICTvec – scenario 3 – day 2 52
Figure 4.41: Number of ships per contact – scenario 3 – day 2 53

Figure 4.42: Contact graph – scenario 3- day 2 ... 53
Figure 4.43: Contact duration – scenario 4 – day2 ... 55
Figure 4.44: Histogram ICTvec – scenario 4 – day 2 ... 55
Figure 4.45: CDF ICTvec – scenario 4 – day 2 ... 55
Figure 4.46: CCDF ICTvec – scenario 4 –day 2 ... 56
Figure 4.47: Number of ships per contact – scenario 4 –day 2 ... 56
Figure 4.48: Contact graph – scenario 4 –day 2 ... 56
Figure 4.49: Contact duration – scenario 5 – day 2 ... 58
Figure 4.50: Histogram ICTvec – scenario 5 – day 2 ... 59
Figure 4.51: CDF ICTvec – scenario 5 – day 2 ... 59
Figure 4.52: CCDF ICTvec – scenario 5 – day 2 ... 59
Figure 4.53: Number of ships per contact – scenario 5 –day2 ... 60
Figure 4.54: Contact graph – scenario 5 – day2 ... 60
Figure A.1: Contact duration - scenario 1 day 1 ... 66
Figure A.2: Histogram ICTvec - scenario 1 -day1 ... 66
Figure A.3: CDF ICTvec - scenario 1 -day1 ... 67
Figure A.4: CCDF ICTvec -scenario 1 -day 1 ... 67
Figure A.5: Number of ships per contacts - scenario 1 -day1 ... 68
Figure A.6: Contact graph - scenario 1 - day 1 ... 68
Figure A.7: Contact duration - scenario 5 - day 3 ... 69
Figure A.8: Histogram ICTvec - scenario 5 - day 3 ... 69
Figure A.9: CDF ICTvec - scenario 5 -day 3 ... 70
Figure A.10:CCDF ICTvec - scenario 5 - day 3 ... 70
Figure A.11:Number of ships per contact - scenario 5 -day 3 ... 71
Figure A.12:Contact graph - scenario 5 -day 3 ... 71

Tables

Table 1: Parameters of different distributions for ICTvec day1 27
Table 2: Distributions Parameters 32
Table 3: Distribution Parameters 35
Table 4: Distribution Parameters 39
Table 5: Concentrated Table with Distribution Parameters 45
Table 6: Concentrated Table of Distribution Parameters 50
Table 7: Concentrated Table of Distribution Parameters 54
Table 8: Concentrated Table of Distribution Parameters 57
Table 9: Concentrated Table of Distribution Parameters 61
Table 10: Distributions mathematical comparison 612

Abbreviations

AIS - Automatic Identification System

ICTvec - Inter Contact Time Vector

ECDF - Empirical Cumulative Distribution Function

CCDF - Complementary Cumulative Distribution Function

1. Introduction

1.1 Motivation – Background

The goal of this dissertation is to study the feasibility of an opportunistic (aka Delay Tolerant) network in the broader Mediterranean Region that will be leveraging the (semi – deterministic) trajectories of ferries, ships and boats across the sea. The basic underlying idea is that the daily routes of many ships in the Mediterranean Sea can be used to create an opportunistic network.

More specifically, hundreds of ships travel within the Mediterranean Sea at any point. Their type varies from smaller boats that serve intra–island communication or touristic tours, to fishing ships, passenger ferries and cargos, up to bigger cruising ships. Besides the size factor, the ships vary according to their trips and trajectories. Some of them travel faster, while others are slower. Moreover, they may have one or multiple destinations. Last, but not least, ships vary with respect to the number of humans they carry onboard. Hence, ships may carry from a single passenger (small vessels) to few thousand passengers (cruising ships) depending on their type.

This dissertation explores how to leverage the mobility of marine nodes across the sea to set up an opportunistic network. This network could serve traffic with no strict delay requirements, such as updates to social media, local news, offline browsing of prefetched content etc. Such an opportunistic network, besides being in line with the ever – expanding use of mobile networks, would represent an interesting alternative for leisure-time communications in the sea environment when compared to the intermittent and costlier cellular coverage and t --much costlier and usually available only at big ships-- satellite coverage.

1.2 Opportunistic Networks: a brief introduction

Opportunistic networks, interchangeably called delay – tolerant networks, represent a distinct networking paradigm that delivers information from one point to another through store-carry-and forward mechanisms. They take advantage of the mobility nodes to realize communication paths between nodes over time (space-time paths), even if no such end-to-end path is available at any single point in time. As they move, nodes encounter other nodes storing messages for different destinations. They may become the next hops for this message or even get a message copy, in the hope that they might meet the destination node earlier than the encountered node will.

Figure 1.1 depicts how a message generated at node A (message source node) reaches its destination node D through successive encounters of third-party nodes (intermediate nodes and relays) that make copies of the original message. The decision to copy and store a message upon each encounter is pertinent to the first and simplest forwarding protocol devised for such networks, called epidemic protocol. Other protocols manage differently the number of message copies in the network.

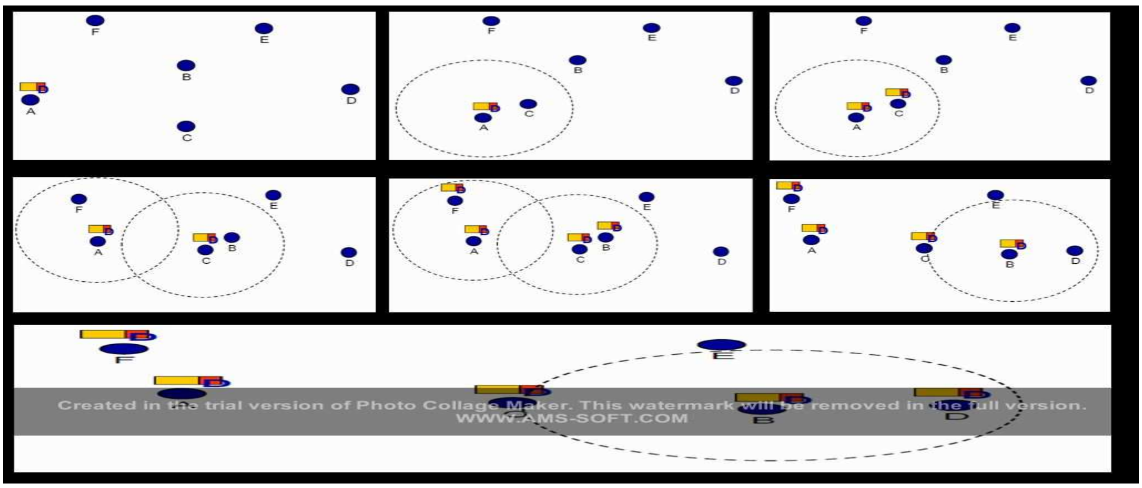


Figure1.1: Message propagation in an opportunistic network from node A to node D over successive node encounters occurring at different time epochs.

Another way of distributing the messages throughout the network is based on the idea of the spray – and – wait protocol, instead of the epidemic protocol described above. It is often favored against the epidemic protocol because performs significantly fewer transmissions than the epidemic protocol, has low number of conflicts, even under high traffic loads and achieves a delivery delay better than other single or multi – copy schemes and it is close to the optimal from the epidemic protocol. The spray – and – wait protocol for the opportunistic network in the Mediterranean Sea is preferable because the nodes move independently and the L is equal to two. It disassociates the number of copies generated per message and the number of transmission with the network size. The disassociation helps the vessels to transmit only once and then the rest of the nodes carry the message and continue to transmit it further in the network.

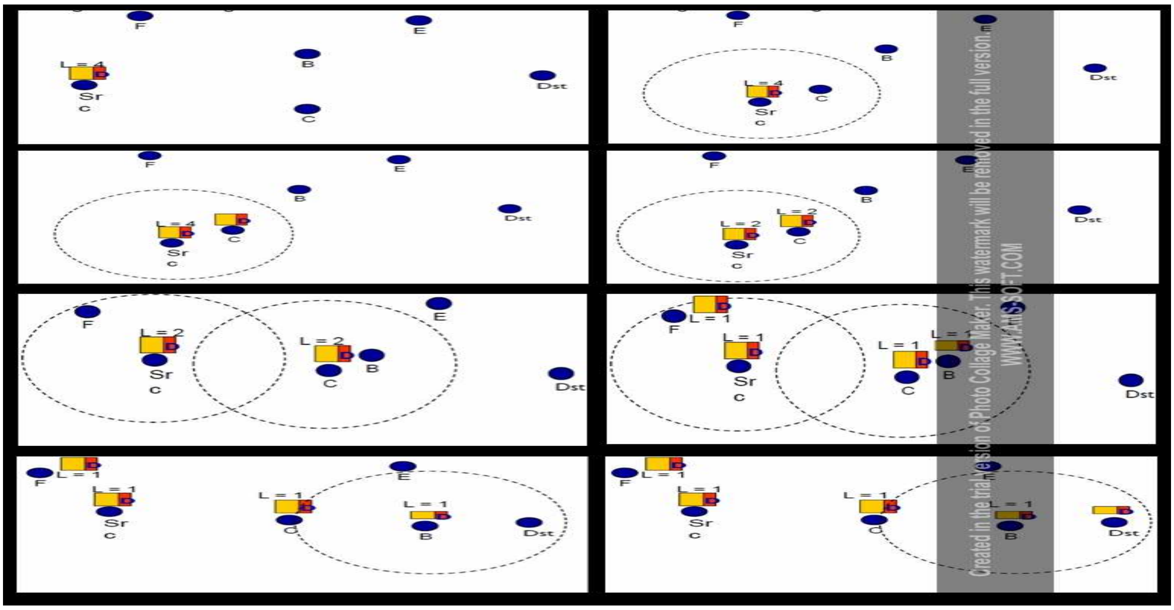


Figure 1.2: Binary Step of the Spray - and - Wait Protocol

Obviously, the node mobility is the most important “resource” of these networks. The characteristics/patterns of this mobility determine the ultimate performance of the network (how good can it theoretically be in delivering messages) as well as what message forwarding policies are appropriate for it (how to practically achieve its theoretical potential).

At a first level, the node mobility patterns are separated into deterministic and non-deterministic (stochastic). Examples of the first case are fixed-track public transportation networks in cities (subway, suburban railway and tram networks) that have periodic trajectories across the city. Another example of deterministic mobility is represented by low- and medium earth satellite constellations, whose orbits in space are pre – defined. The marine transport, where this study is focusing on, is another instance of deterministic mobility, even though vessels have more degrees of freedom in their tracks. Under deterministic node mobility patterns, it is possible to offline optimize the forwarding decisions in order to achieve some optimization objective, e.g., minimize the message delivery time or path hop – count[1].

With human nodes, mobility patterns are stochastic. Yet stochastic does not mean random; instead, there is structure in this mobility, dictated by the daily routine (commute to workplace and back, visit the gym, meet with friends). However, there is no single well-defined approach to optimize for such patterns and the range of forwarding solutions is much broader. As a first classification, these are split into randomized forwarding and utility-based forwarding solutions, based on how a node decides whether to forward a message (copy) to a node it encounters.

With randomized forwarding there is no discrimination against the nodes. Each node the sender encounters is an equally good and bad option for forwarding the packet. Protocols of this category have two parameters that can be configured: the number of per message copies M and the hops H each one is allowed to make. Therefore, the network nodes keep on generating message new message copies as long as their number does not exceed M and forwarding further a given message as far as it has not traversed more than H different nodes. The two variables define a tradeoff between protocol efficiency (how fast the message can be delivered) and network overhead (how many message copies are generated for a single message to be delivered to its destination).

Utility – based forwarding, on the other hand, acknowledges and tries to cater for the heterogeneity of node mobility. Protocols under this category maintain some notion of utility for each network node, which may depend on the specific message, the message source node and/or its destination. Then, the protocols try to recruit nodes with high utility scores as message relay nodes to forward the message. Typically, practical measures of node utility account for the frequency of node encounters with other nodes or the message destination, in particular; the duration and or recency of these contacts; or some notion of similarity or other social relation of the nodes to the message destination node[2]. Utility-based forwarding protocols require more state and computation effort in return for better performance and use of network resources.

---

[1] Jain, S., Fall, K. & Patra, R. 2004. Routing in a delay tolerant network. *SIGCOMM Comput. Commun.* Rev. 34, 4.pp.145-158

[2] Jain, S., Fall, K. & Patra, R. 2004. Routing in a delay tolerant network. *SIGCOMM Comput. Commun.* Rev. 34, 4.pp.145-158

1.3 Objectives

The ultimate objective of this study is to explore whether an opportunistic network can result from the mobility patterns (actually trajectories) of the vessels circulating in the Aegean Sea and the broader Mediterranean Sea.

To this end, several questions would need to be addressed, a non-exhaustive list of which includes:

a) Is the density of the marine routes adequate to set up such a network? Namely, do the mobility patterns of the marine nodes mix sufficiently well to let data hop through the network faster that what users would get if they waited to use their cellular connection upon reaching their trip destinations intermediate harbors with cellular coverage?

b) What is the amount of data that could be served through such a network?

c) Are there any cost – effective measures that could enhance the network performance, making it a competitive alternative, at least technically, to other solutions?

The dissertation seeks to make a first step towards addressing these questions by analyzing the opportunistic network structure underlying the ship routes across the sea. Namely, it treats the crowds of ships as potential nodes of an opportunistic network and computes characteristic attributes of these networks.

1.4 Methodology

Our work in the context of this dissertation, proceeds in discrete steps. The first step is the knowledge and process of data concerning the moving vessels. Real datasets about marine traffic in the Mediterranean Sea have been obtained after an agreement with the site www.marinetraffic.com. The provided data is AIS data. Ships are equipped with Automatic Identification Systems (AIS) that regularly (semi – periodically) generate data concerning the position of the ship, their speed, name, id etc. These data let tracking the position of a given ship throughout its route through the sea. The challenge then is to convert these data about individual ships' trajectories to lists of pairwise "encounters" of the ships (in opportunistic network terminology).

The second step is the analysis of the ships' movements and trajectories. Namely we need to infer instances that two ships come within range of each other so that they can communicate with each other. The communication radius will consider the available radio technologies determines the density of encounters between ships. Also, the time is an important factor to determine whether two or more vessels are within the communication radius and may transmit and contact each other. Vessels' movement analysis provides the metrics and characteristics needed for network characterization.

2. Opportunistic Networks

2.1 Characterizing an Opportunistic Network

Basic measures of interest for the characterization of the opportunistic network include inter–contact time (ICT) distributions, contact graphs, contact duration times and a few other statistics reflecting the role of individual ships in the network. All the above characteristics are thoroughly discussed in the following sections.

2.2 Inter – Contact Times

Two interacting mobile nodes is a crucial aspect of wireless network communication, such as opportunistic networks. The direct encounters between two mobile nodes are largely exploited in opportunistic network for message delivery, across multi – hop paths. When the mobile nodes communicate, a contact between them is made. The starting point of discussion in a contact is inter – contact times. Inter – contact times are defined[3] as the time intervals between two consecutive interactions (contacts) between the same pair of nodes.

2.2.1 Pair–wise Inter–Contact Times

The pair–wise inter–contact times (ICTs) is the time series resulting from inter – contact times between one specific pair of nodes. For a network with N nodes, there are $\frac{N(N-1)}{2}$pair-wise ICT distributions.

2.2.2 Aggregate Inter–Contact Times

Aggregate inter – contact times result from the superposition of the pair –wise inter – contact times over all possible pair of nodes that exist in a network. Their distribution is

[3] From Pereto Inter – Contact time to residuals, Chiara Boldrini et al.

an important characteristic of the opportunistic networks and is an indication of the time needed for a message to be delivered to its destination. For a network with N nodes, there is a single aggregate ICT distribution.

2.2.3 Inter–any Contact Times

The inter-any contact times is the time series resulting from inter – contact times between a specific node and any other network node. For a network with N nodes, there are N different inter-any contact time distributions.

2.2.4 Aggregate Inter–any Contact Times

Aggregate inter –any contact times result from the superposition of individual node inter-any contact times over all nodes that exist in a network. For a network with N nodes, there is a single aggregate Inter-any contact time distribution.

2.3 Contact Graph

The formal definition of contact graphs comes from graph theory. Given a set of objects in some space, the associated contact graph contains a vertex for each object and an edge implied by each pair of objects that touch in some prescribed fashion. While graphs have been extensively studied for objects such as curves, line segments, and even strings[4], their use in wireless networking is equally extensive and helps the visualization of the patterns the mobile nodes follow in the network during a time period. A typical example is the social networks graphs where the vertexes represent the mobile nodes and the edges represent the actions the nodes take in the network and the whole graph represents the way the contacts are made in the network.

[4] Rectangular Layouts and Contact Graphs, Adam L. Buchsbaum et al.

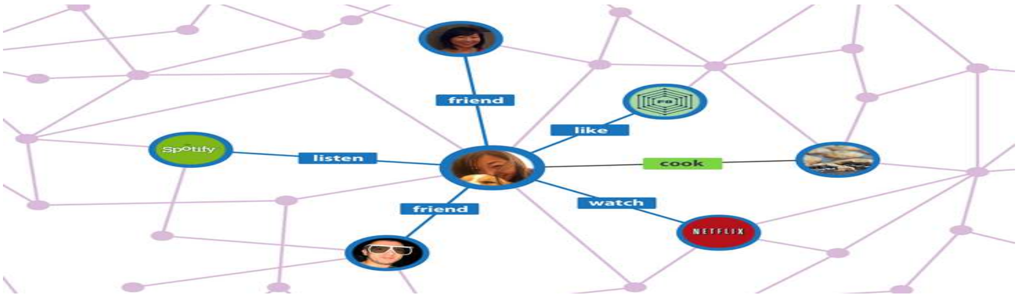


Figure 2.1: A Typical Facebook Contact Graph

2.4 Contact Duration

Contact duration is an opportunistic network characteristic, which shows how much time a contact between two nodes lasts. The time is measured from the moment a contact between two nodes starts till the moment they are out of range of each other. The opportunistic network studied in this dissertation considers as contact duration the time period in which two vessels can exchange AIS data, within a certain range. Usually, contact duration’s size is a small number, because in a network with mobile nodes that continuously moving, the time the nodes spend contacting other nodes is small. However, there are also long contacts; such contacts imply that the two nodes either move together in the network or they do not move at all and are stationary.

2.5 Other Characteristics

Other opportunistic network characteristics studied in this dissertation are the number of contacts that exist in the network. Also, the number of nodes which have a certain number of contacts is shown because such a characteristic reveals the contact graph complexity. This characteristic is essential for the connectivity of the network. As many nodes have as many contacts with different nodes the contact graph becomes denser. The edges of the graph denote the number of nodes’ contacts. In the marine opportunistic network, the vessels are considered to be the nodes, while the edges are the transmission of AIS data between two vessels. For all the above characteristics, their empirical probability mass distribution, cumulative probability distribution and complementary cumulative probability distribution are measured.

3. Datasets and Methodology

In order to determine an opportunistic network derived from the vessels' movements across the sea, a dataset containing AIS data is obtained. Ships are equipped with Automatic Identification Systems (AIS) that regularly (semi – periodically) generate data concerning the position of the ship, their speed, name, id etc. The data lets tracking the position of a given ship throughout its route through the sea. The challenge is to convert these data about individual ships' trajectories to lists of "pairwise" encounters of the ships. Deriving from the data, there is a need to infer instances that two ships come within range of each other, so they can communicate with each other. In order to succeed in inferring two ships within range, the study of the AIS data in the dataset is required, as well as a preprocessing analysis.

3.1 AIS Data

The Automatic Identification System (AIS) is an automatic tracking system used on ships and vessel services to locate and identify vessels by exchanging data with nearby vessels, AIS base stations and satellites. AIS data provide information about the position of the ship, speed, name, id and route. Navigator sensors, such as a gyrocompass or a rate of turn indicator are used in gathering the data.

The primary intention for gathering AIS data is to allow the ships to monitor the marine traffic in their area and also to be seen by the same traffic. A secondary, unintentional use of AIS data can make a vessel publically viewable over the Internet. Aggregate data is accessible by every internet – enabled device and vessel movements are available online and in real – time applications. The www.marinetraffic.com website provides AIS data. Most of the data is available free of charge, but satellite data and special services such as searching in the archives are supplied at a cost. The dataset used in order to develop the marine opportunistic network contains the following types of data: MMSI – Maritime Mobile Service Identity, Latitude, Longitude, SOG – Speed over Ground, COG – Course over Ground and Timestamp, which will be further analyzed.

The main applications for collecting AIS data are: collision avoidance, fishing fleet monitoring and control, vessel traffic services, maritime security, aids to navigation, search and rescue, accident investigation, fleet and cargo tracking. Hopefully, a marine opportunistic network can be developed from the AIS data provided.

3.1.1 Maritime Mobile Service Identity

A Maritime Mobile Service Identity is a series of digits which are sent in digital form over a radio frequency channel in order to uniquely identify ships, ship earth stations, coast stations etc. These identifiers are formed in such a way that can be used by general telecommunications. In the dataset, there are MMSIs only concerning the identities of ships.

3.1.2 Latitude

Latitude is a geographic coordinate that specifies the north – south position of a point on the Earth's surface. Latitude is an angle which ranges from 0^o at the Equator to 90^o (North or South) at the poles. In the dataset the point of latitude is the latitude of a ship

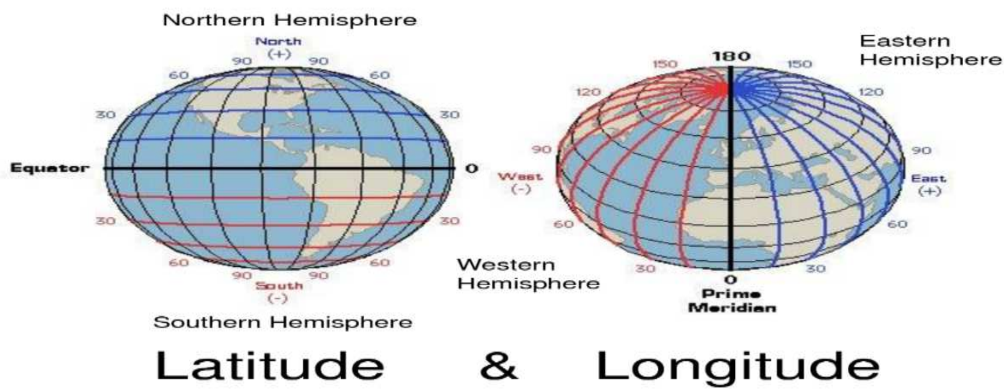


Figure 3.1: Latitude and Longitude on the Globe

3.1.3 Longitude

Longitude is a geographic coordinate that specifies the east – west position of a point on the Earth's surface. It is an angular measurement, usually expressed in degrees. Points with the same longitude lie in lines running from the North Pole to the South Pole. By convention, one of these lines, the Prime Meridian, which passes through the Royal Observatory, Greenwich, England, was intended to establish the position of zero degrees longitude. The longitude of other places was to be measured as the angle east or west from the Prime Meridian, ranging from 0^o at the Prime Meridian to $+180^o$ eastward and -180^o westward. In the dataset the point of longitude is the longitude of a ship.

3.1.4 Speed Over Ground (SOG)

Speed over ground (SOG) is a different measurement from the vessel's speed, in which it sails through the water. Speed over ground is the combination of the vessel's speed and the tidal current speed that may exist. When the tide is travelling directly with or against the boat, the speed over ground is the vector sum of the two, directing to an increase or decrease of the speed. To compute the speed over ground is more difficult when the tidal current is travelling with an angle to the ship's direction. When there is an angle between the tidal current and the vessel speed, the Pythagorean Theorem is used to compute the final speed, as well as the direction of the ship. Because of this effect, sailing and rowing boats should essentially travel in parallel with the tide, in order to gain speed. On the other hand, for larger power vessels, which travel with much higher speed when the sea is flat, speed over ground is not an essential factor. It may even be faster for those vessels to go against the tide, if the wind and the tide is in the same direction, a situation that produces flatter sea conditions, and may allow the vessel to travel at a higher speed.

3.1.5 Course Over Ground (COG)

The line connecting the vessel's consecutive positions in the sea is referred as track. The track the vessel was intended to follow is called route. For aircrafts, but as well as for ships too, the route is represented by the great circle line that connects the previous waypoint with the next waypoint. Navigators should make the tracks coincide as much as possible with the routes. Course over ground, is the actual path of a vessel with respect to the seabed. Course may be relevant to the true north (true course) or the magnetic north (magnetic course). Direction (course over ground) is typically measured from north, either true or magnetic, in degrees from 0^o to350^o, following compass convention (0^0 being north, 90^o being east etc.). Contrary to vessel movements and aviation, where wind and currents may cause the two to differ significantly, for land based vehicles heading and course are typically identical.

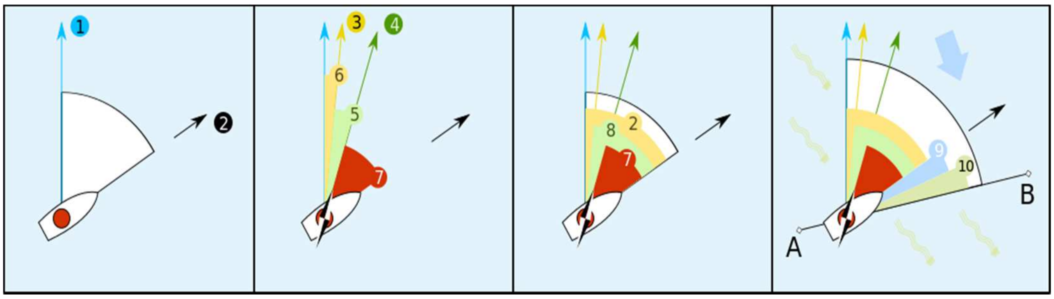


Figure 3.2: Course over Ground

1. Heading (2) is the angle the direction in which the vessel's nose is pointing and a reference direction (e.g. the north).
2. Any reading from magnetic compass refers to compass north (4), which is supposed to contain a two – part compass error:
   a. The Earth's magnetic field's north direction, or magnetic north (3), almost always differs from the true north by magnets variation (6), the local amount of which may be found in nautical charts.
   b. The vessel's own magnetic field may influence the compass by so – called magnetic deviation (5). Deviation only depends on the vessel's own magnetic field and the heading, and therefore can be checked out and given as a deviation table.
3. The compass heading (7) has to be corrected first at deviation (the "nearer" error), which yields the magnetic heading (6). Correcting this for variation yields true heading (2).
4. In case of crosswind (9), and/or tidal or other current (10), the heading will not meet the desired target, as the vessel will continuously drift sideways; it

becomes necessary to point the heading away from the course to counteract these effects and make the track coincide with the great circle.

3.1.6 Timestamp

A timestamp is a sequence of characters or encoded information identifying when a certain event occurred, usually giving date and time of day, sometimes accurate to small fraction of a second. The term derives from the rubber stamps, used in offices to stamp the current date and sometimes time, in ink on paper documents, to record when the document was received.

Common examples of this type of timestamp on a letter or the “in” and “out” times on a time card. However, in modern times usage of the term has expanded to refer to digital date and time information attached to digital data. For example, computer files contain time stamps that tell when the file was last modified, and digital cameras add timestamps to the pictures they take recording the date and time the picture was taken.

A digital timestamp is the time at which an event is recorded by a computer, not the time of the event itself. In many cases the difference may be inconsequential; the time at which an event is recorded by a timestamp (e.g. entered into a log file) should be close to the time of the event. This data is usually presented in a consistent format, allowing for easy comparison of two different records and tracking progress over time. The practice for recording timestamps in a consistent manner, along with the actual data is called timestamping. The sequential numbering of events is sometimes called timestamping. Timestamps are typically use for logging events or in a sequence of events (SOE), in which case each event in the log or the SOE is marked with a timestamp. In filesystems, timestamp may mean the stored date/time of creation or modification of a file.

The dataset from the collected AIS data contains a timestamp. The timestamp shows the date and time an AIS receiver collected and stored in a log file the data sent from different vessels. The timestamp in the dataset has the form: DD-MM-YY hh:mm, according to the ISO 8601 standard for the representation of dates and times.

3.2 Processing of AIS – Data Traces

3.2.1 Pre-processing of Data

Developing a marine opportunistic network across the Mediterranean Sea requires data that is gathered from vessels moving in the area. Vessels transmit many different AIS data. Log files are created to store the data and then they can retrieved in order to be used in different applications.

According to the pre – processing processes for the AIS data set, obtained from the www.marinetraffic.com website, it is obvious that data selection and cleansing have already occurred. All the suitable data values for a network application are gathered during the data selection process. The data set contains MMSI, as an identifier of the vessels, the vessels' coordinates, latitude and longitude, speed and course over ground, as well as a timestamp, which is the day and time the AIS data was transmitted.

The provided data set has already been through data cleansing because the data set is filled with accurate data values. Missing or corrupted values have been erased or corrected and the data is ready to proceed to the next processes of pre - processing: normalization, transformation and feature extraction. The given data set does not need a normalization process because the data will not be more easily presented in their canonical form.

On the other hand, data transformation is required to provide more eloquent attributes. The first transformation is the mapping of the existing, in the data set, MMSIs to smaller and consecutive integers. The transformation is one – to – one mapping transformation, and each unique MMSI is mapped to a unique number. The purpose of the transformation is to create smaller identifiers for the vessels, which will be more easily be used in the application. The second transformation is the creation of two new values derived from timestamp. Each new value contains the date and time respectively. The main reason for separate date and time is the ease of mathematical computations. The final data transformation is the transformation of time to an integer value. Each time corresponds to an integer and the reason is again use in the application and speed for mathematical computations. The feature extraction process is not used, because the data set has a few variables and there is no complexity and need to reduce the dimensions. As a result, the data set acquired from the data transformation process if the training set, which is used in the application.

3.2.2 Main Processing of Data

3.2.2.1 Why MATLAB?

The next logical step towards the programming solution is the use of high – level program language in order to create the project, in which the network simulator will be written.

The choice of the language used in a program depends on different criteria with each different application. An application which has thousands of entries of data in tabular form, as the data obtained from the website is very cost – effective in computation resources. A language that has the computational power and the structure to perform large amounts of computations as well as supports tabular forms as its main element is preferable for such kinds of application. Matlab was chosen because both requirements of the application were met. Matlab's main types are the cell, the matrix and the array. It is also a mathematical oriented language that has many predefined functions that help accelerate computations. Another reason for Matlab's obvious choice is that the data is contained in tabular form in a .csv file. The form of the data even after preprocessing remains tabular, and Matlab is the dominant language for mathematical processing of such data.

3.2.3 Algorithmic Steps from AIS Data to Contact Traces

The MATLAB application developed for the transformation of the AIS data to contact traces follows the steps below:

- Import the data in a script
- Imported data has different types, other is numerical data and other is string data
- Create a table containing all the numerical data

- Transform the string data into numerical data and the concatenate to the numerical data table
- Time and date data must be taken into consideration and also be transformed into numerical data
- Concatenate also numerical time and date data in the numerical data table
- Delete data containing zero speed, because they are docked in a port and do not move, and they do not need a new network.
- Data must be sorted according to time and date
- Create a loop in which all sorted data is parsed and then while a condition is true the output is a structure that contains the vessel ids and the times and the coordinate of the first vessel, in which two vessels have a contact. The structure is similar a contact table of an opportunistic network.
- The second loop iterates through the elements of the structure and find the Inter – any – contact times between the vessel, as well as the duration of each contact.
- Third loop used to compute the contact duration
- Forth loop that measures how many contacts has each ship
- The final step of the simulator contains several plots, showing different network characteristics and metrics.

Since the most computation trial of the code happen in the four loops that create the contacts find the inter – contact times and compute the contact duration and count how many contacts has each ship, an analysis of these steps of the application should be made. The parameters defined for these steps and how they generally work will be also studied.

3.3 Code Analysis

The pseudo code snippet below sort the data containing in the table MM by the values that exist in the ninth column. Afterwards the time threshold to consider a contact is set at two minutes and the threshold of whose value and below, it is considered the same contact is set at twenty minutes. There are also initialization of matrices and variables that are going to be used in the following code.

```
MM = sort(M,9)
thresh_t = 2
SameContactThresh = 20
numships = length(unique(M(:,1)))
maxshipid = max(unique(M(:,1)))
contacts = zeros(maxshipid,maxshipid)
lasttime = -1*ones(maxshipid,maxshipid)
numContacts = 0


for i = 1 to length of the data table MM
       t_now = MM(i,9)
       lat1 = lat(i)
       lon1 = lon(i)
       j = i + 1
       rowset_t = find the indices that  MM(j:end,9) - t_now <= thresh_t is true and store them to rowset_t
       rowset_t = rowset_t + i
       numrows_fwd = length(rowset_t);
       latlon1 = do a table of [lat1 lon1] multiple times the size of the table [numrows_fwd,1]
       latlon2 = do a table to contain the coordinates of rowset_t
       dist = compute the distance using the two tables(latlon1, latlon2)
       contactss  = find  the indices that dist <= 2 is true
       for  c = 1 to length of the table contactss
              orig_rowid = is the value that takes rowset_t when contactss index is c
              if  the value of contact is 0 for the indices i AND orig_rowid
                     create new contact
                     increment numContacts
                     create structure s of numContacts that contains ship1, ship2, tstart, tend, lat2, lon2
              else
                      if   t_now - lasttime<=  SameContactThresh   is true can assume this is the same contact
                             for indC = 1 to numContacts
                                    if  ((ship1 is equal MM(i,1)) AND   (ship2 is equal MM(orig_rowid,1)))  OR
                                       ((ship2 is equal MM(i,1) AND ship1 is equal MM(orig_rowid,1)))  AND
                                          tend is equal lasttime(MM(i,1),MM(orig_rowid,1))
                                           tend take the value of   MM(orig_rowid,9);
                                    end
                             end

                      else
                             increment numContacts
                             create structure of numContacts that contains ship1, ship2, tstart, tend, lat2, lon2

                      end
              end
       end
end
```

The above snippet of pseudo code is the loop that creates the structure containing the vessels that communicate with other vessels and have contacts. The structure contains the ship id for the first and the second ship, the starting and the ending time of the contact, as well as the latitude and longitude of the second vessel. Analyzing the snippet, the variable t_ now takes the value of time and changes every time. The initial for loop runs from 1 till the length of the data table, the coordinates from all vessels and then there is a condition imposed for the time threshold. The time threshold is set at two minutes and it is a step that shows how much further in the data set the execution proceeds in each run. If the condition is true all the indices complying with the condition are found and a new table containing the corresponding values is created. Afterward there is a call to an external function that computes the distance according two sets of pairs of coordinates. There is a check for the distance and, if the distance is smaller than two kilometers, and the check is true, the corresponding row of the contact table is forwarded in order to create the structure. The inner for loop runs from one till the length of contact table and if both indices are zero, creates a new row in the structure and increments the entries for the time. When the condition is not satisfied, there is a second condition that is satisfied when there is a contact already, and the check is to find whether the new contact is a continuing one or not. If the check is true the new contact is a continuing one and the corresponding row of the structure is updated with the new ending time. This happens when the check is smaller than the same contact threshold. When the condition is false, the same contact threshold is smaller than the check and the contact is considered as a new one.

```
create temvec table with columns ship1 and ship2
take the unique values of tempvec
sort tempvec
ICTvec = 0
ind = 0
for i=1 to length of tempvec
      lastt(i) = 0
      for j=1 to length of the structure
            if (  ( s(j).ship1 is equal tempvec(i) ) AND  ( s(j).ship2  > s(j).ship1 ) ) OR (  ( s(j).ship2 is equal
tempvec(i) ) AND  (  s(j).ship1>s(j).ship2 )  )
                  if  lastt(i) is equal to 0
                        lastt(i) takes the value of s(j).tstart;
                  else
                        increment ind
                        ICTvec(ind)  takes value equal to: s(j).tstart - lastt(i)
                  end
            end
      end
end
```

The above code creates a table that has a columns the ids of the first and the second ship. The table is created to contain only the unique values and afterwards is sorted. A new table ICTvec which will contain inter – contact times is created. For loop runs from one to the length of tempvec and the variable lastt is initiated. The inner for loop runs for the length of the structure and there are two conditions to be satisfied in order an entry to be written in the ICTvec table. The first condition is: *(s(j).ship1 is equal tempvec(i)) AND (s(j).ship2 > s(j).ship1)) OR ((s(j).ship2 is equal tempvec(i)) AND (s(j).ship1>s(j).ship2)* and the second say that last time should be equal to zero. When both the conditions are satisfied then the ICTvec table is updated with new entries.

```
A = [s.ship1 s.ship2];
for i=1:length(s)
    contact_duration(i) = s(i).tend - s(i).tstart;
end
```

The third snippet shows how the contact duration is counted. The snippet below creates a table with the unique values of the ship ids and the runs for the length of the table. In the loop the variable occurancy is initiated. The inner loop runs the length of the structure and if the condition is true the occurancy is incremented and a new entry to the structure ship_contacts is created. Ship contacts contains the ship id and the number of occurancies.

```
create table ships with columns ship1 and ship2
take unique values of ships
for i=1 to the length of ships
    occurancy(i) = 0;
    for j=1 to the length of the structure
        if (s(j).ship1 is equal ships(i) OR s(j).ship2 is equal ships(i)) AND (s(j).ship1 not equal (j).ship2))
            increment occurancy
            create a structure ship_contact that contains the ship id and the number of occurancies of the ship
        end
    end
end
```

3.4 Baseline Scenario Definition

Three variable parameters are configurable in the simulator. The first parameter is the *SameContactThreshold*. This parameter shows the time in which two vessels can have contacts and the multiple contacts to be considered one continuous contact. In the network, there are ships that have contacts many times according to the data set. The threshold imposed by the same contact threshold parameter provides the knowledge, the vessels which have contacts in times smaller than the threshold are considered to have only one continuous contact with a larger contact duration. The threshold is set at twenty minutes, meaning the same two vessels are found to have a contact after twenty minutes, this contact is considered to be a new contact and it is also logged in the simulator as a new one. The *SameContactThreshold* does not change, during the different scenarios, because it is considered to be a realistic threshold to be applied in an opportunistic network.

The second parameter to be examined is the *thresh_t* parameter. It is a control parameter that changes from the baseline value of one minute to two minutes in the different scenarios applied. The time threshold shows the abstract between the second ship's time and the first ship's time. The check conducted with the time threshold parameter provides the ships possible to have a contact. It is set, in one minute in the baseline scenario, because of the format of time in the data set. The time AIS data is received, each minute, but not all vessels transmit simultaneously and there are vessels that do not transmit each minute. As a result there is a time space of approximately two minutes in which two vessels can contact each other and an ambiguity is created when a vessel contacts another just before the first AIS signal is sent and just after the second AIS signal. The time passed during this two actions is two minutes, extracted from the format of AIS data. The choice of the time threshold to be set at one minute, is the most logical, as it is set as the mean. The time threshold parameter changes in different scenario to two minutes, in order to check how the number of contacts created changes.

The third parameter that changes from the baseline scenario to the different scenarios applied to simulate the network is the distance between two vessels which contact each other. The distance *dist* is set at one kilometer in the baseline scenario. The choice of one kilometer is based upon the distances given in most of the network protocols, in which two nodes are able to contact each other without the deterioration factors of the protocols play a major role in the transmission and the signal loss can be ignored. Based on the radio transmitters used for the AIS data, the distance can be changed for different simulation scenarios. Radio transmitters have bigger range than mobile transmitter and exploited the radio transmitter for the creation of an opportunistic network across the moving vessels is essential to the application. The distance changes from one kilometer to five hundred meters or two kilometers in order different scenarios to be tested.

The data set containing the AIS data used in the application is sorted by date. There is data collected from four different days. Three out of four day contain data collected over a period of approximately eight hours, usually starting from a time in the morning and finishing at some time in the evening. One the other hand, the third day contains data obtained only on a span of forty five minutes in the morning. The size of data belonging to the third day is significantly smaller than the rest of the days. Another crucial point concerning the data set is the number of vessels that have contacts are smaller than the total number of vessel. This can be explained, when considering a vessel entering a port at the starting time of data set. At this time the vessel only manages to transmit a few times and then it is securely docked in the port and it does not transmit again. On the other hand, there are transmissions which are ignored. These transmissions come from vessels that move with small speeds, in the ports, where there is no need for a new network.

4. Characterization Results

4.1 Scenarios Results

4.1.1 Baseline Scenario – *dist =1* and *thresh_t = 1*

4.1.1.1 Day 1

The baseline scenario for the first day of the entries in the data set contains 86907 entries during almost eight hours, starting from 9:49 in the morning and finishing at 15:05 in the evening. The length of the data does not cover the whole day, but an essential part of it, during the rush hours of traffic in the Mediterranean Sea. The date is the fourth of August, 2012, a summer day, which provides much traffic and it is ideal in order to study if an opportunistic network can be created from the vessels moving in the sea. Extracted from the results of the simulation, the following diagrams show the characteristics and the metrics of an opportunistic network.

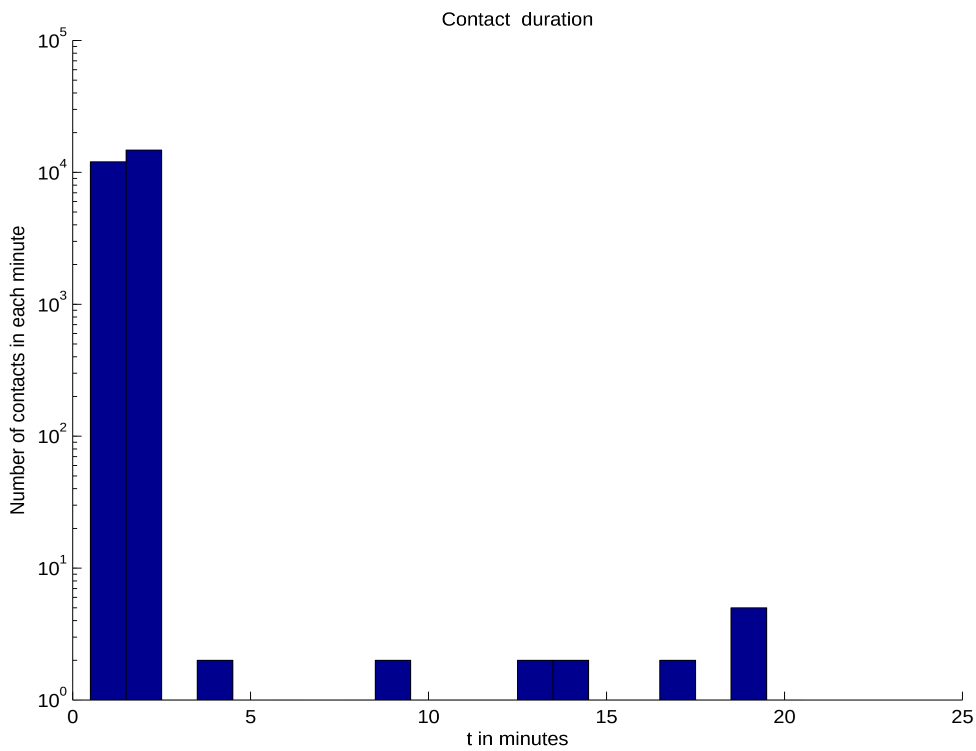


Figure4.1: Contact duration day 1

The histogram shown above indicates that most of the contacts last one or two minutes, when the time threshold is one minute and the distance is one kilometer. There are also contacts, which their duration is bigger, but the number of these contact drops dramatically. While the contacts of one or two minutes have a number of more than ten thousand for each minute, all the other contacts, with bigger times, are about twenty each.

The second network characteristic considered in the simulation is inter – any contact time. The variable corresponding to inter – any contact time is the *ICTvec* variable. For the ICTvec, the histogram shows the concentrations of inter – contact times of different vessels. The other metrics used are: the empirical cumulative distribution function which shows the cumulative probability, the complimentary cumulative distribution function which shows the complimentary cumulative probability of *ICTvec* and the histogram containing how many ships have a number of contacts.

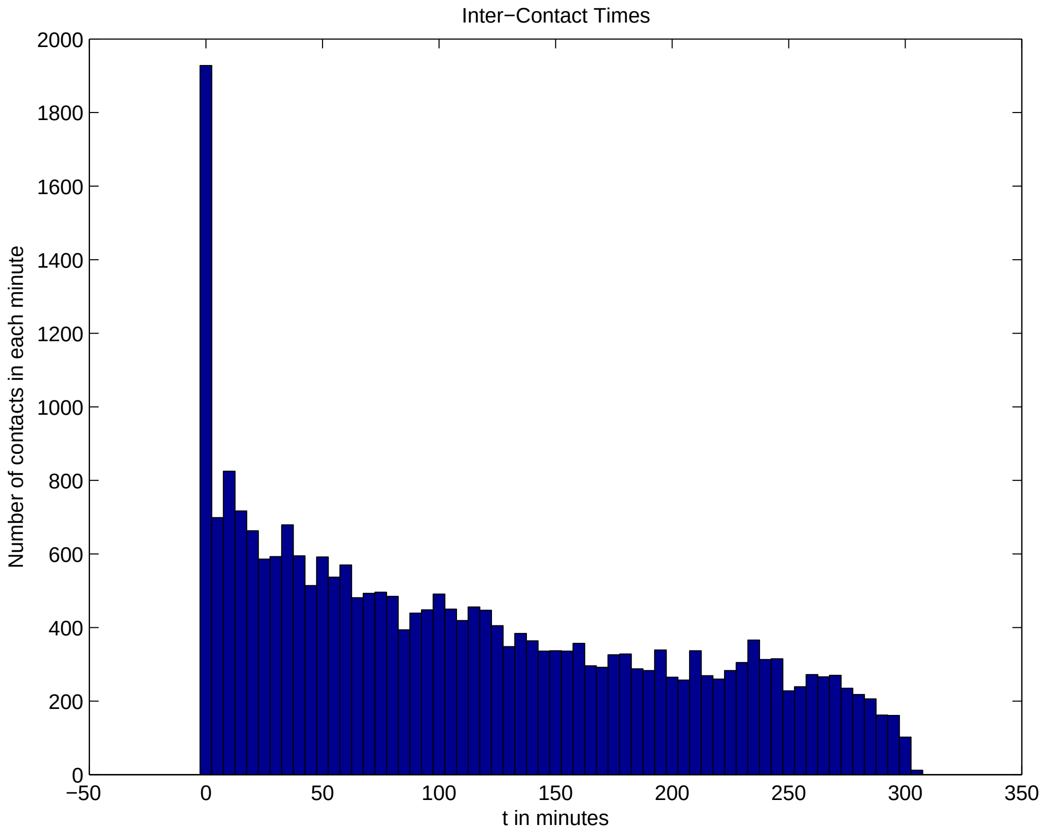


Figure 4.2: Histogram of inter - contact times day 1

The histogram of inter – any contact times shows that number of contacts in each minutes is a decreasing function in time. There is a peak at zero minutes and the most contacts occur at that time. The concentration of contacts is higher at zero can be explained by the fact that these contacts occur near ports by vessels entering or exiting the port, only meeting once and transmitting AIS data at the same minute. These vessels move with the smallest speeds, but their contacts is found by the simulator. As the inter – any contact time increases, a positive assumption can be made that vessels located in the middle of the sea, fulfil the specifications of time threshold and distance and they are suitable candidates for an opportunistic network in the sea.

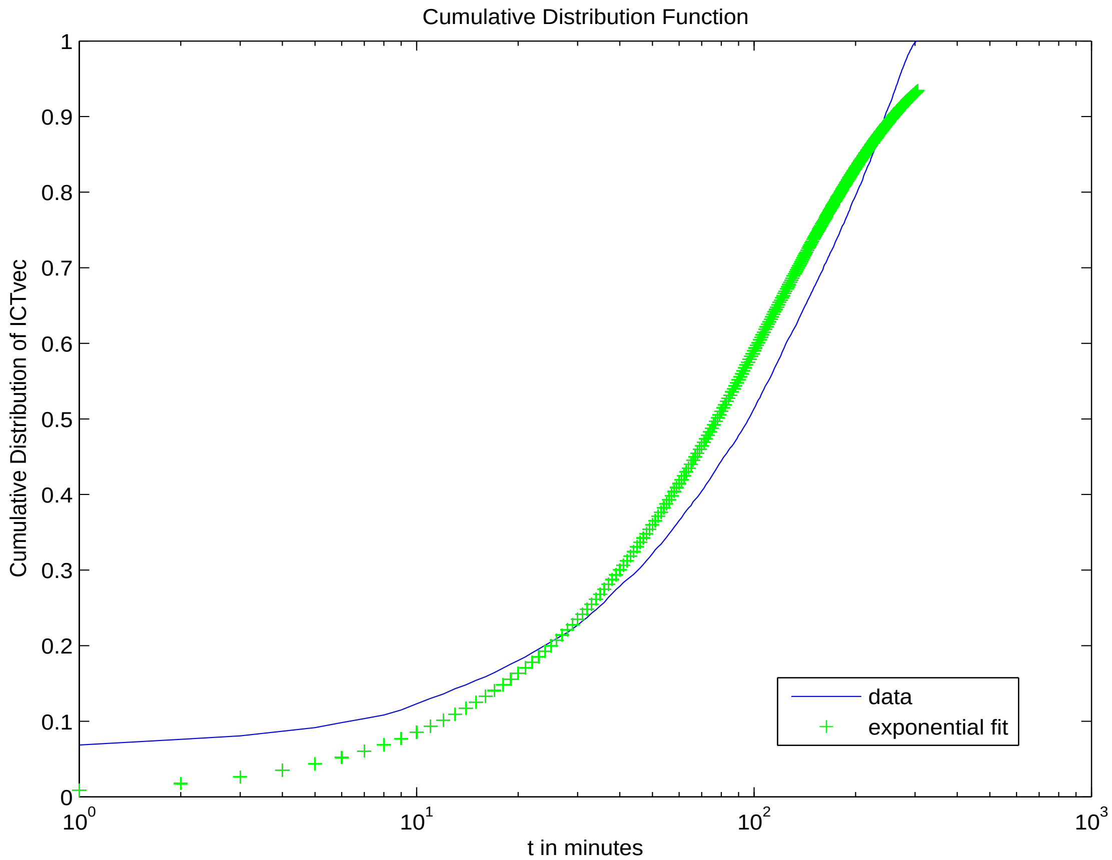


Figure 4.3: Cumulative Distribution Function of ICTvec

The cumulative distribution function follows an exponential distribution as shown in Figure 11. The blue line is the actual data, where the green line is the exponential fit with a mean value, $\mu = 112.1004$

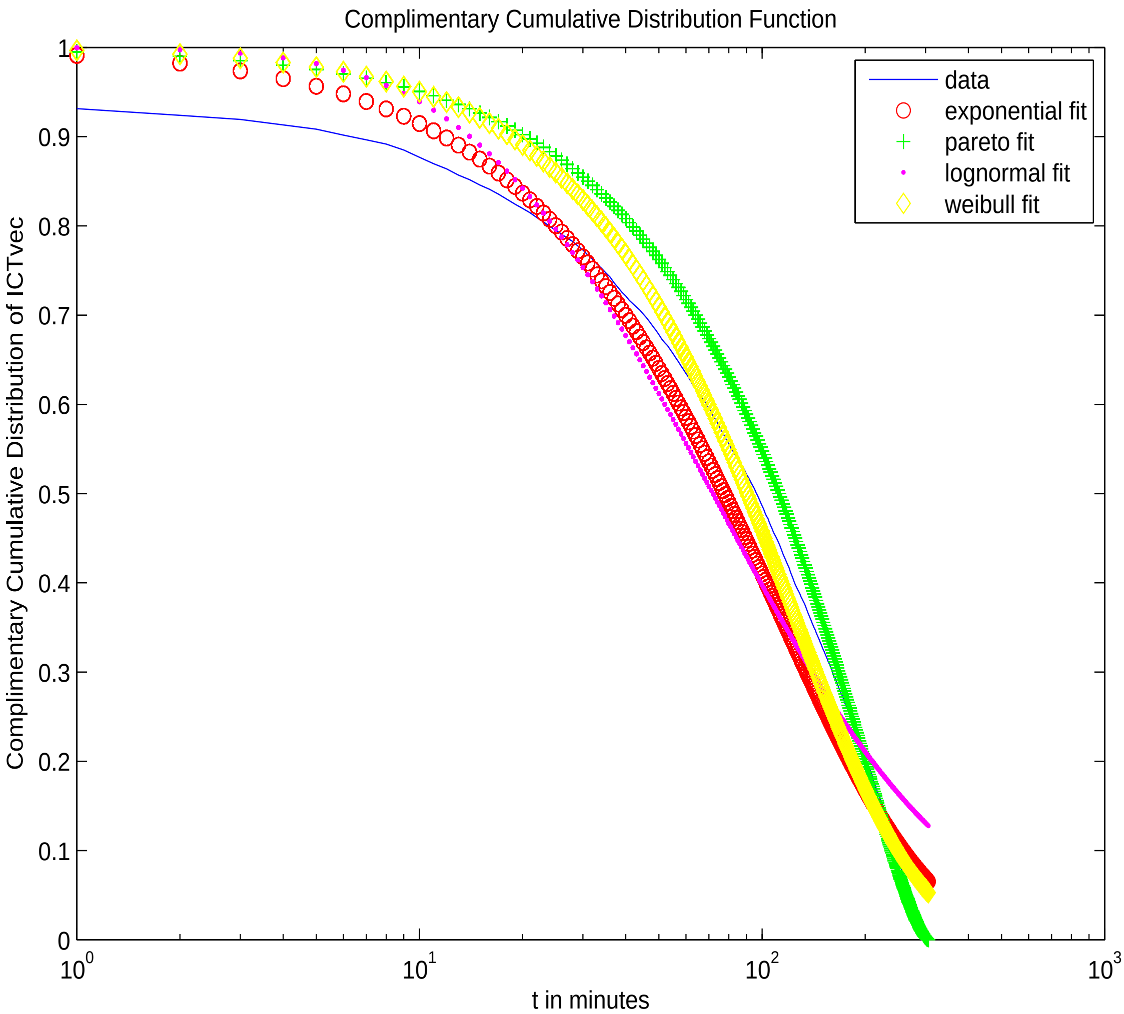


Figure4.4: Complimentary Cumulative Distribution of ICTvec with different fits day 1

Figure 12 shows the different distribution fits tried for the Complementary Cumulative Distribution of *ICTvec.* The distributions that were tried to fit the data of *ICTvec* were the exponential fit, the generalized pareto fit, the lognormal fit and the Weibull fit. For the exponential fit all values of ICTvec were use, including the zero values. For the rest of the fits the zero values were excluded, because of the assumption made before, that these contacts are either from vessels entering or exiting a port, or from vessels moving in a port. For both cases, there is no need in implementing a new network, because the vessels can exploit the terrestrial network really well, due to their proximity to the nodes of the network. The parameters of each of the fitting distribution are shown in the table that follows:

Distribution	**Parameters**		
exponential	$\mu = 112.1004$		
pareto	$\kappa = -0.0796765$	$\sigma = 1.64191$	$\vartheta = 0$
lognormal	$\mu = 0.38346$	$\sigma = 0.350482$	
Weibull	$a = 1.72298$	$b = 2.26995$	

Table 1: Parameters of different distributions for ICTvec day1

The exponential distribution gives the μ parameter, which is the mean value of the distribution. The Pareto distribution returns the parameters κ σ ϑ. κ is the shape parameter for the generalized Pareto distribution, specified as a scalar value. σ is the scale parameter for the generalized Pareto distribution, specified as a nonnegative scalar value and ϑ is the location parameter for the generalized Pareto distribution, specified as a scalar value. The lognormal distribution has two parameters. The μ is the mean value, while σ is the standard deviation of the distribution in a logarithmic scale and it is always a nonnegative value. Finally, the Weibull distribution parameters are: the scale parameter a, which takes a positive value and the shape parameter b, which is also positive.

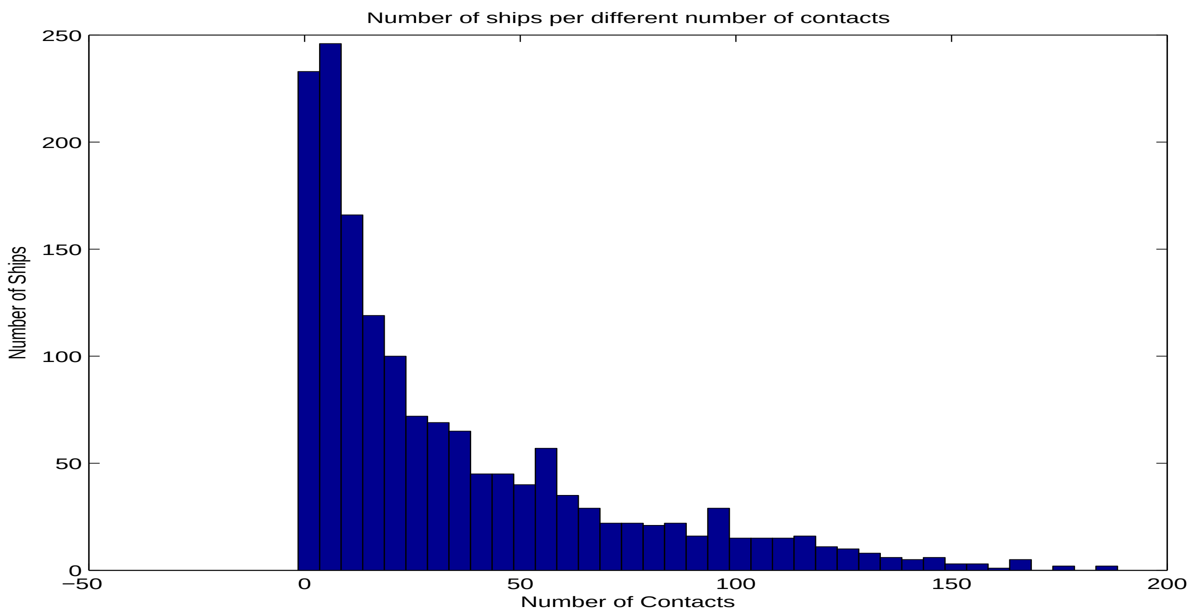


Figure 4.5: Number of ships per number of different contacts day1

The histogram shows that there is great concentration ships which have a few contacts. Most of the ships have ten of less contacts. For higher number of contacts, the concentration of the ships is decreasing and for the very large number of contacts, there are only a few ships that have so many contacts. So the concentration of the ship is a decreasing function of the number of contacts, which is a logical explanation. Many vessels can have only one or two contact, because of their movements in the sea. On the other hand, there are few vessels that have a lot of contacts, and that depends on the routes they follow are the ports the visit during their trips.

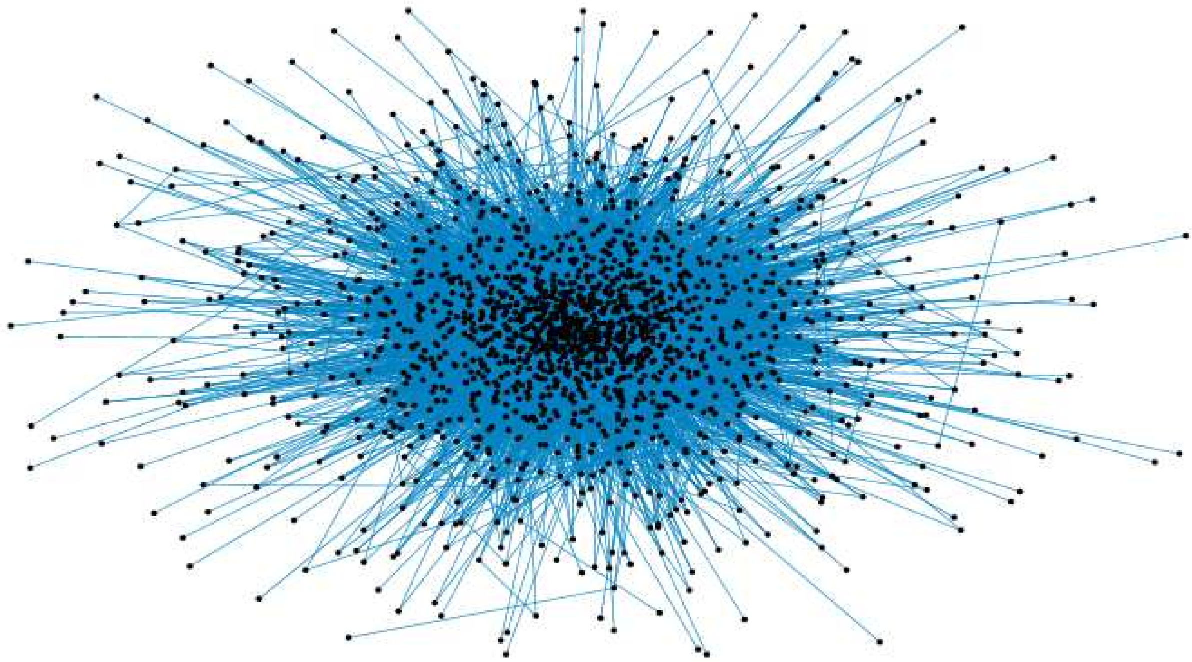


Figure 4.6: Contact graph created for the baseline scenario day1

The contact graph[5] is the final network characteristic to be discussed. The graph contains the contacts between two vessels as edges and the vessels themselves as nodes. As it is shown the graph is densely populated and reminds the graph of a social network. The fact the graph has this form is in fact correct, because social networks are based on the opportunistic networks.

[5] Contact graph was created with NodeXL extension for Office Excel, inserting the ships which have contacts as nodes

4.1.1.2 Day 2, Day 3 and Day 4 – Baseline Scenario

The simulator runs the baseline scenario for all the other three day existing in the data set. Day2 data set has 121380 entries. It starting time is on 10:30 in the morning and its finishing time is on 17:31 in the evening. The entries cover exactly seven hours of the day of the twentieth of August, 2012. The day2 data set is the largest of all. It also does not cover the whole day, but it covers a substantial part of it and safe conclusions can be drawn.

Day3 data set has 8659 entries. It has the least entries of all data sets. The entries span to a time space from 10:15 in the morning till 10:53 the same morning. The data set's duration is only thirty eight minutes. It cannot indicate the traffic existed in the sea on the thirtieth of August, 2012. Day3 data set is only studied, in order to examine the simulator's performance during extreme entries. In this case there are very little entries.

Day4 data set has 85055 entries. It is approximately the same size as the data set of day1. It starting time is at 11:34 in the morning and the entries have finishing time at 19:43. Its duration is seven hours. The duration of the data set is larger from this in day1. This can be explained by the date of the fourth data set. Data is collected on the twenty – eighth of October, 2012. It is a day in late autumn and traffic at the sea is expected to be less than in a day in the midst of summer.

- **Day2**

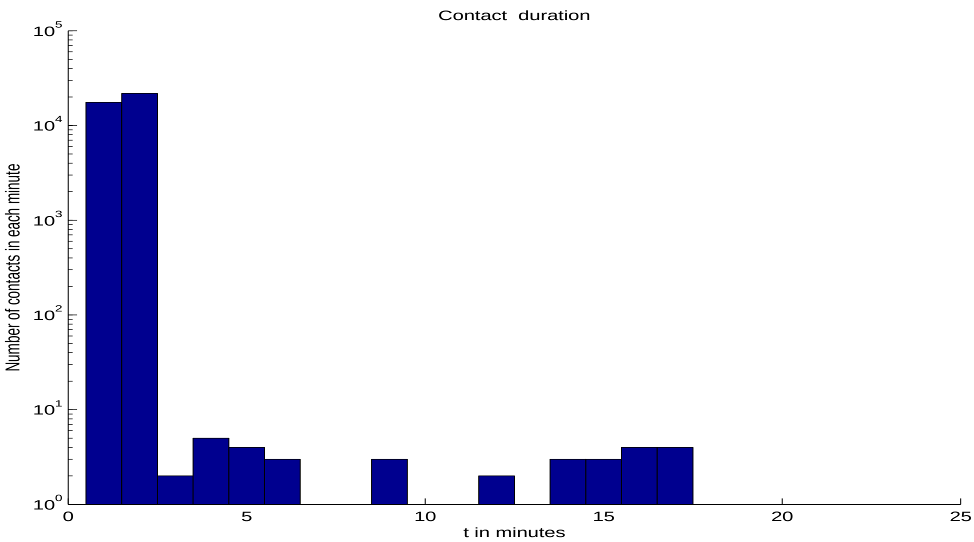


Figure4.7: Contact duration day2

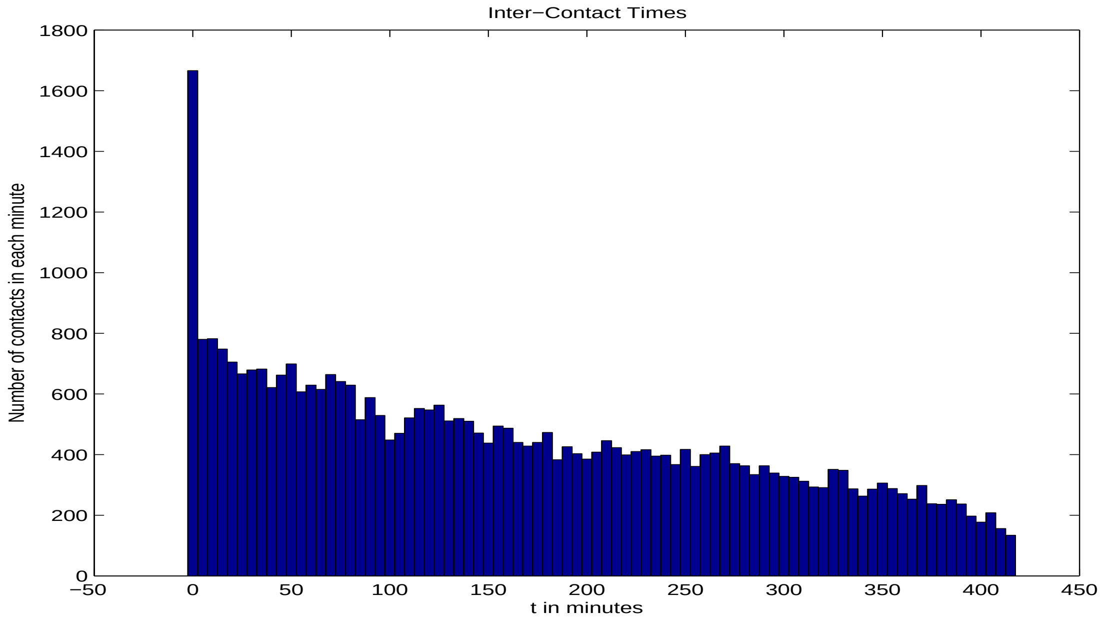


Figure 4.8: Histogram ICTvec day2

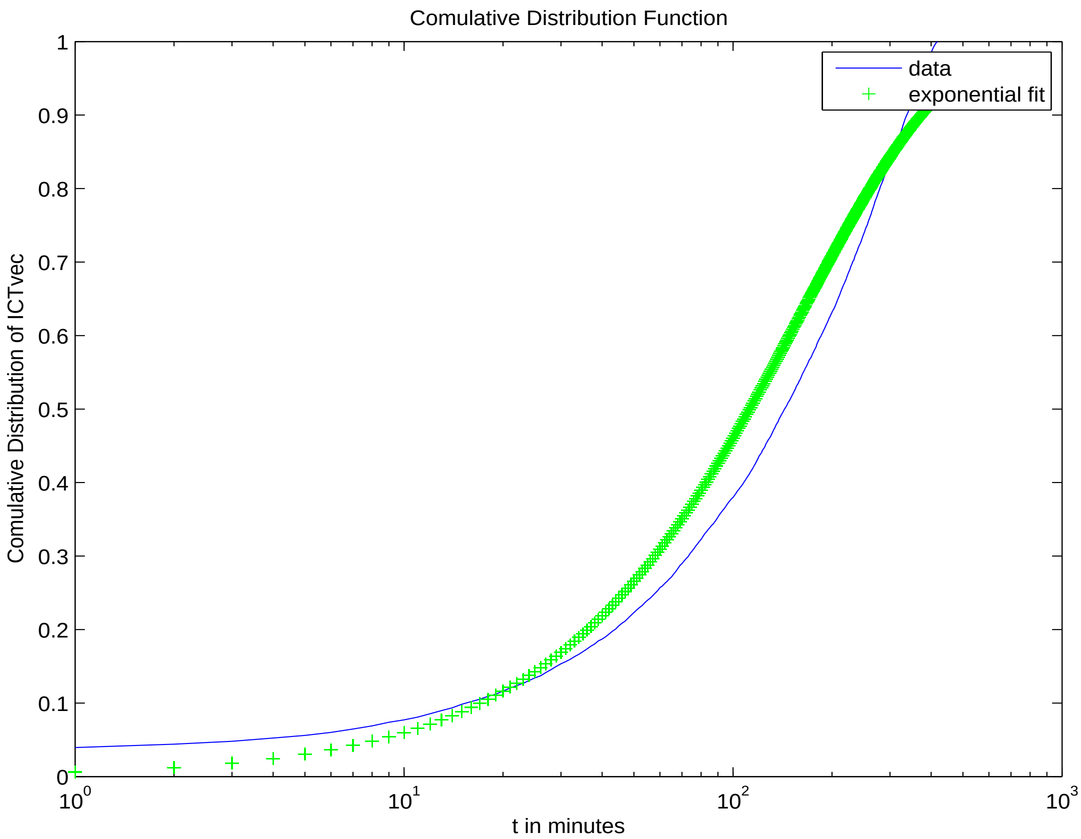


Figure 4.9: CDF ICTvec day2

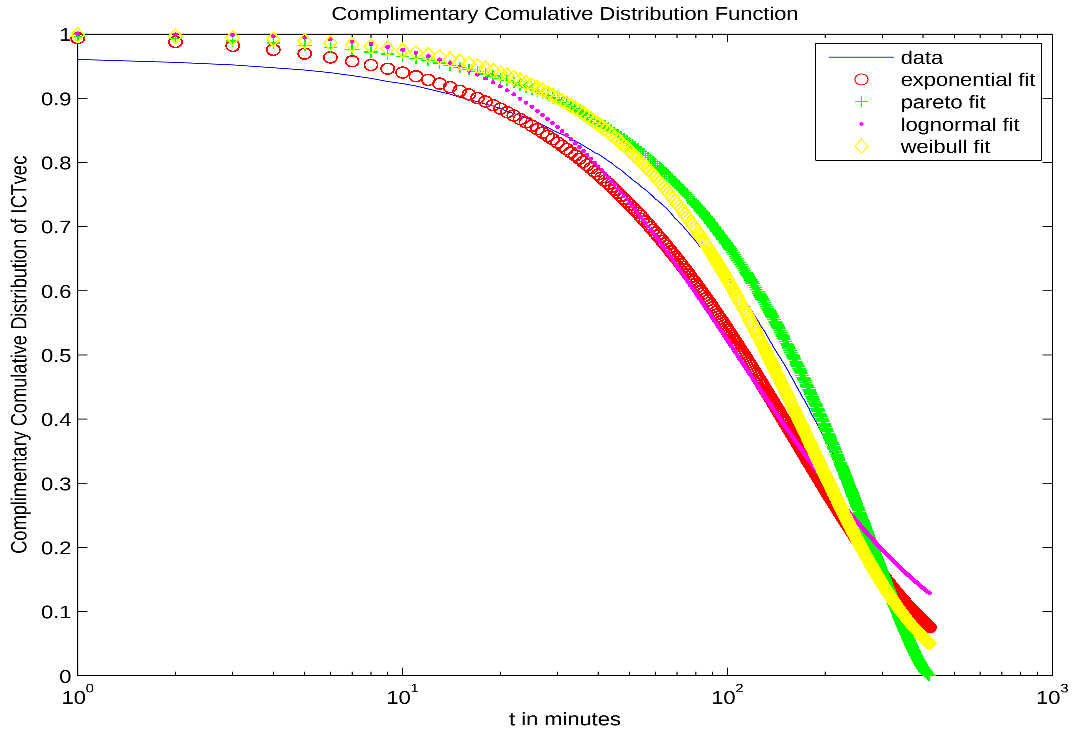


Figure 4.10: CCDF ICTvec day 2

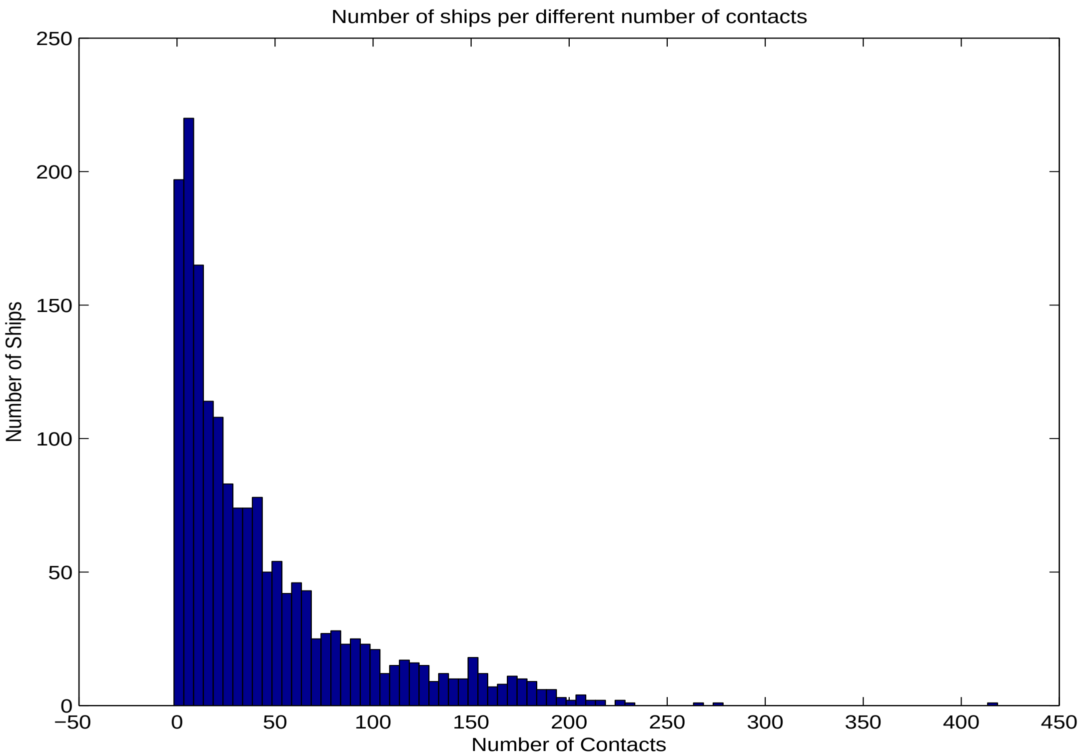


Figure 4.11: Number of ships per contact day 2

Distribution	Parameters		
exponential	$\mu = 166.15$		
pareto	$\kappa = -0.0771916$	$\sigma = 1.63798$	$\vartheta = 0$
lognormal	$\mu = 0.385709$	$\sigma = 0.34974$	
Weibull	$a = 1.7317$	$b = 2.35735$	

Table 2: Distributions Parameters

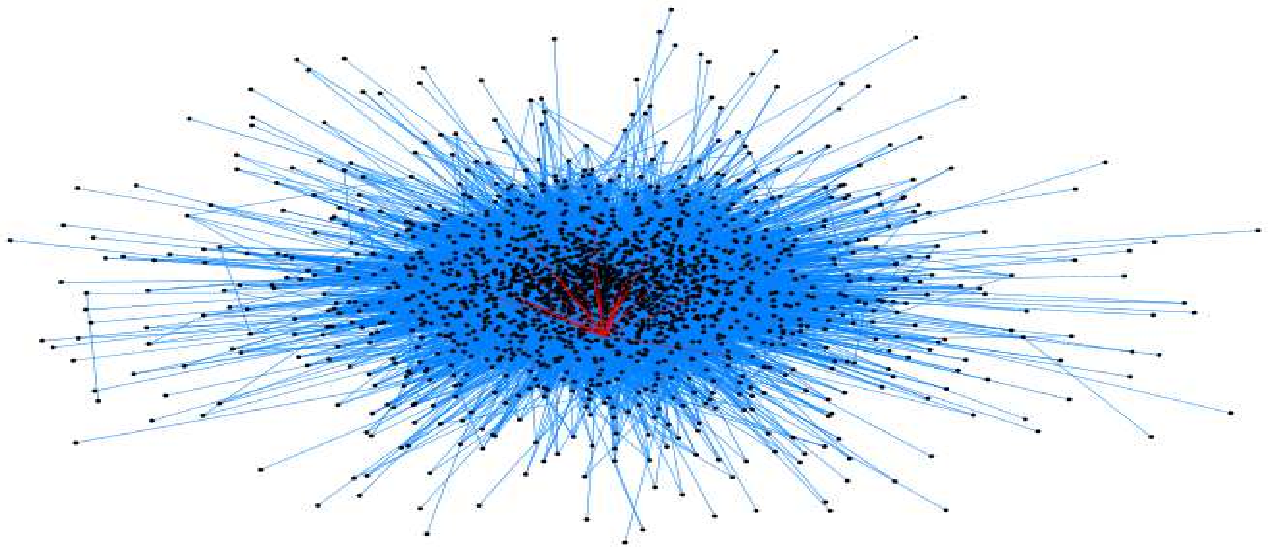


Figure 4.12: Contact graph day 2

Comparing the results with those obtain for day1, for the baseline scenario, one can see that there are more values for the contact duration, because there are more entries in the data set. From the histogram of *ICTvec,* there is high concentration of contacts for the values near zero, that means that many contact are very small, but the curve gradually drops and not as dramatically as in the first day. EDF and CCDF data and fitting curves, are similar in both cases. For the CCDF curves there is a concentration at 1 and the curve drops following the theoretical values. According to the histogram of ships per contact, there are more ships in the data set and the concentration for larger contact times becomes smaller. Parameters from the fittings are shown in Table3 and the contact graph is the one shown in Figure20. The graph is more dense and complicated, than the one from day1, and this happens because there are more entries in the data set meaning more contacts have been made during the period concerning the data set.

- **Day 3**

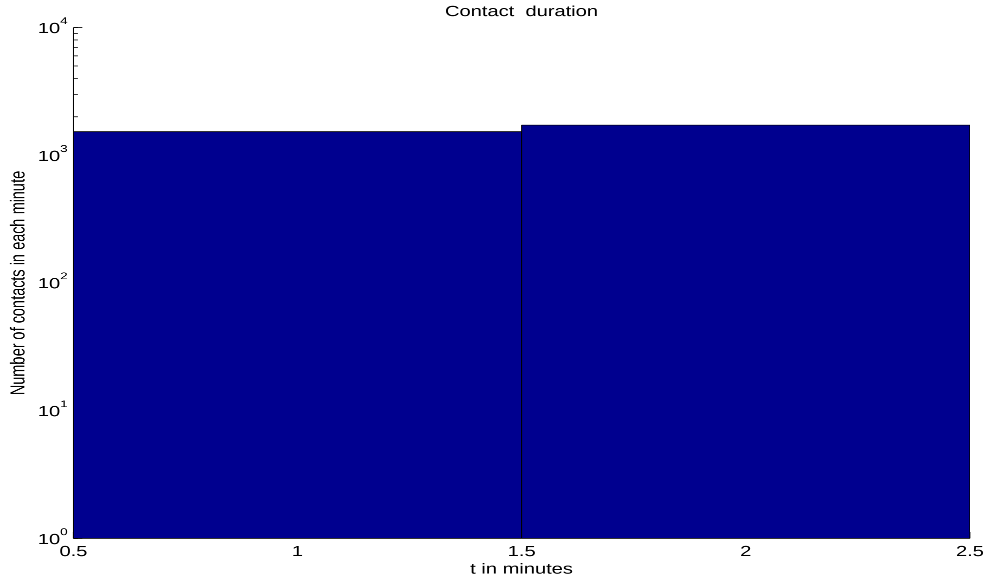


Figure 4.13: Contact duration day 3

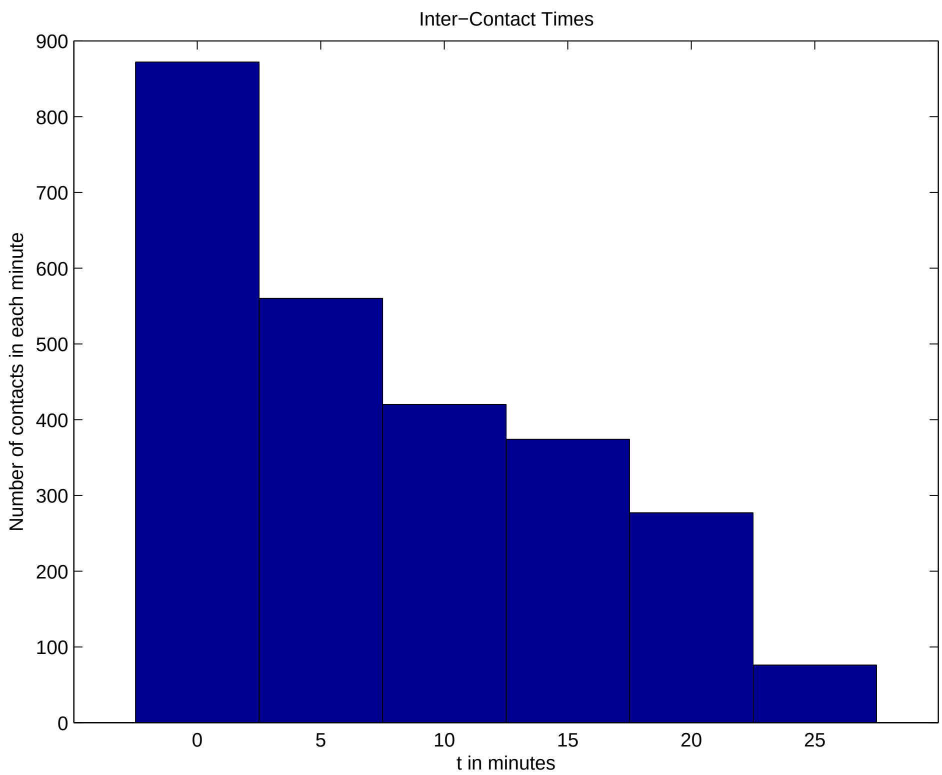


Figure 4.14: Histogram ICTvec day 3

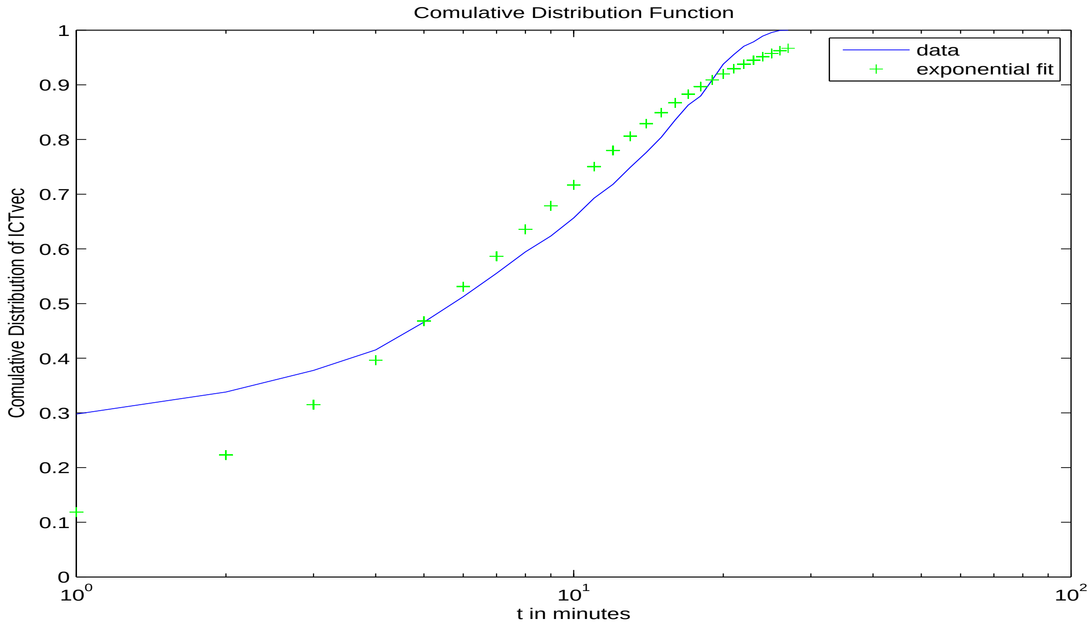


Figure 4.15: CDF ICTvec day 3

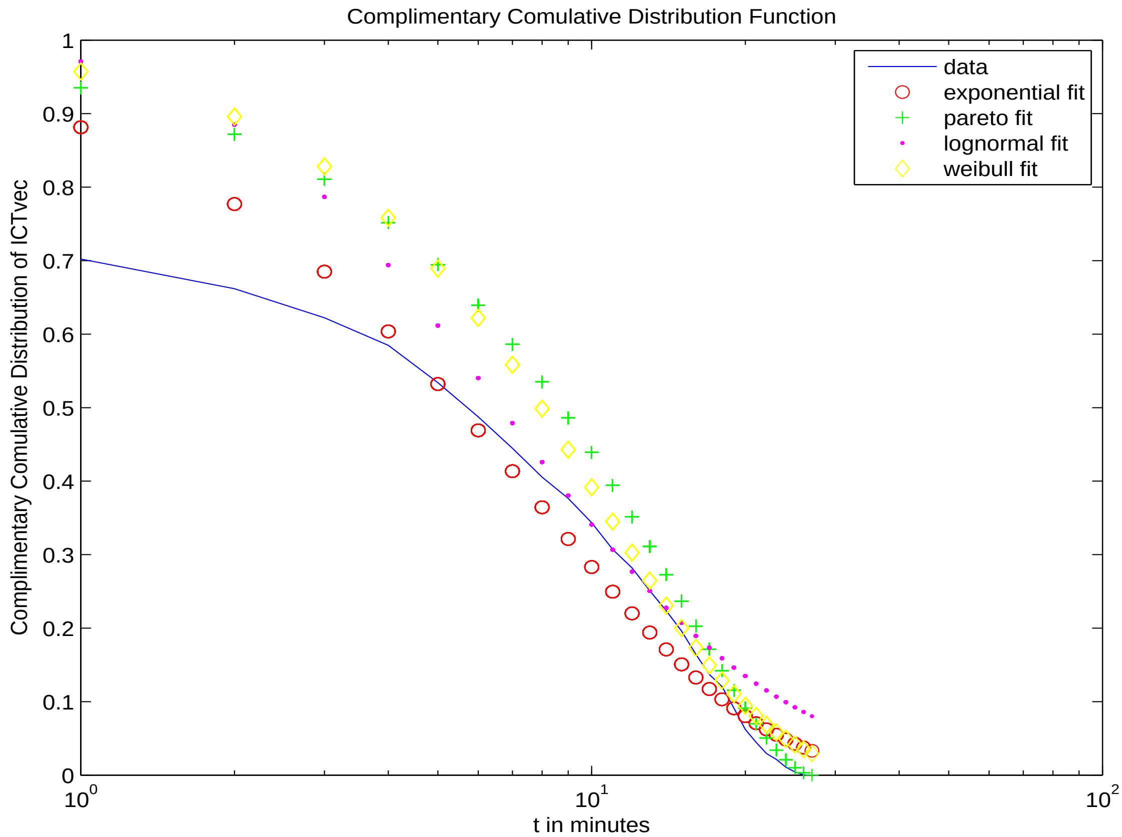


Figure 4.16: CCDF ICTvec day 3

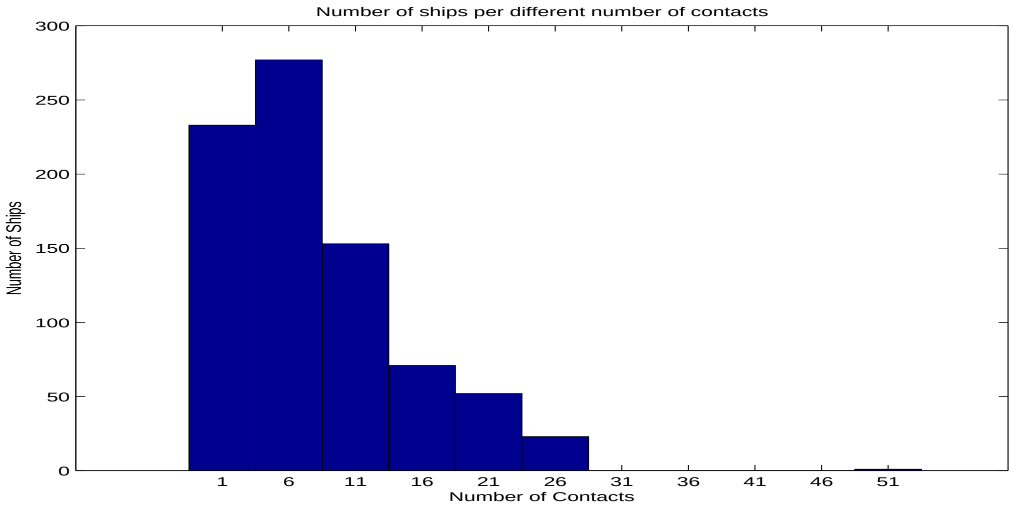


Figure 4.17: Number of ships per contact day 3

Distribution	Parameters		
exponential	$\mu = 9.68404$		
pareto	$\kappa = -2.03.194$	$\sigma = 4.03194$	$\vartheta = 0$
lognormal	$\mu = 0.367261$	$\sigma = 0.346009$	
Weibull	$a = 1.70931$	$b = 3.59219$	

Table 3: Distribution Parameters

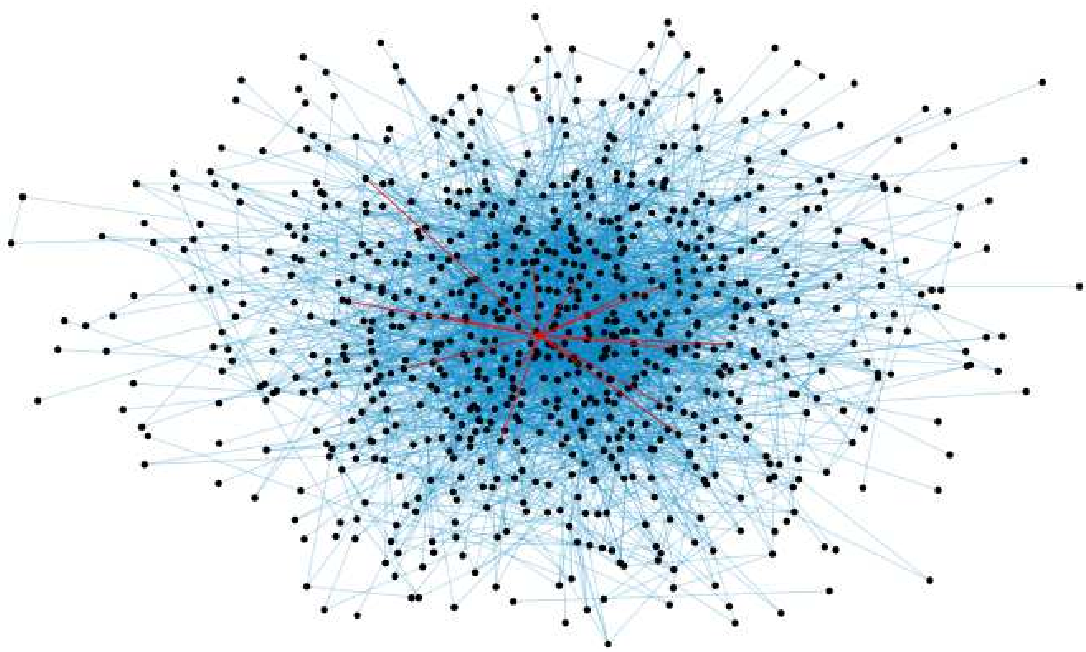


Figure 4.18: Contact graph day 3

For the third day's dataset and due to the fact that the entries in the data set is few, contact duration takes only two values, zero and one. The vessels that communicated during this data set, only spend these few minutes to send and receive from each other. The histogram of *ICTvec* shows inter – any contact times are also small, both in number and in values comparing the results with day1 and day2 data sets. The ECDF and CCDF curves and the corresponding fitting distributions can be shown on Figure23 and Figure24 and the parameters of the fitting curves on Table4. Due to the fact of the small size of the data set, the fitting curves can be shown better on the figures and the differences between them and the actual data is better shown in day3 data set. Finally, the histogram containing the ships per contact time shows that most ships have contacts that last only a few minutes and very little ships have contact with larger duration. The contact graph created is less dense than the ones emerging from data from day1 and day2, and it is actually revealing of the size of the opportunistic network existing from the vessels moving in the Mediterranean Sea.

- **Day 4**

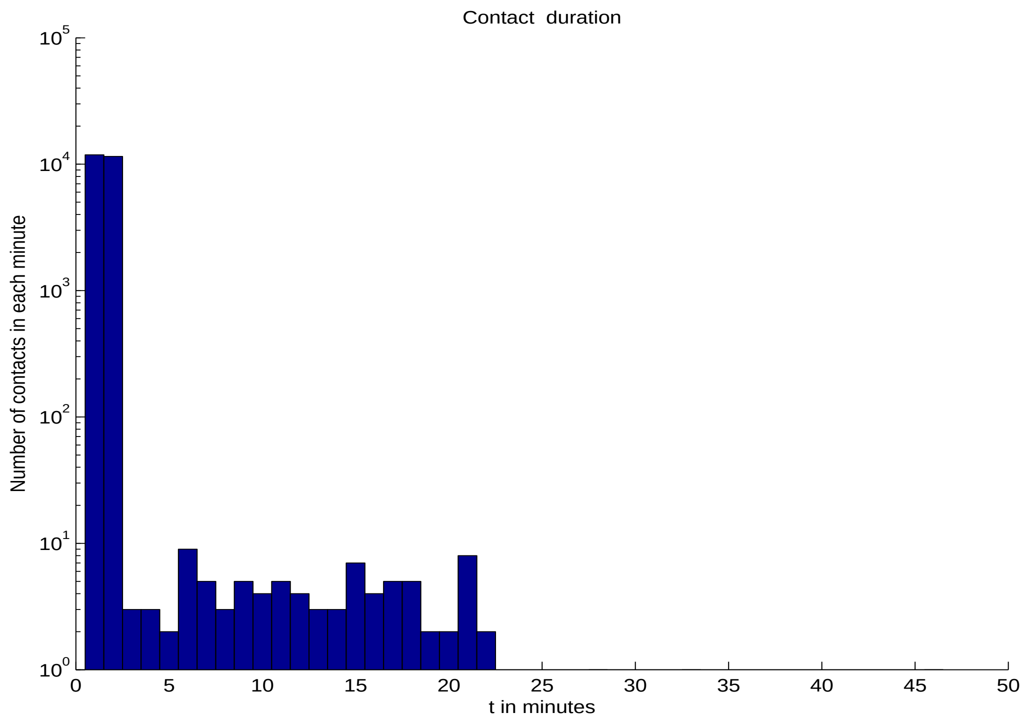


Figure 4.19: Contact duration day 4

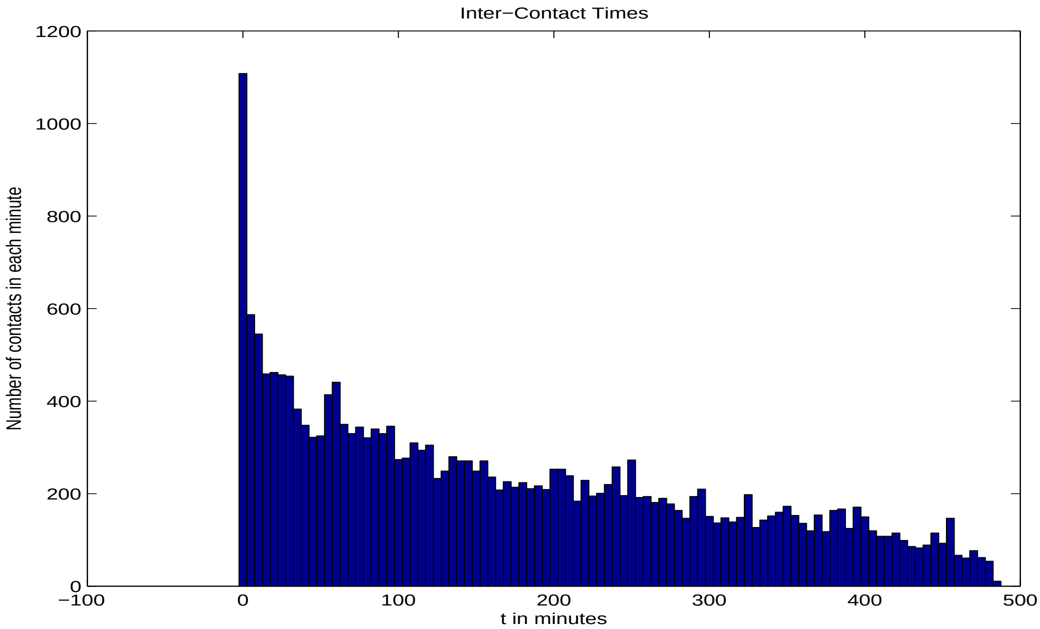


Figure 4.20: Histogram ICTvec day 4

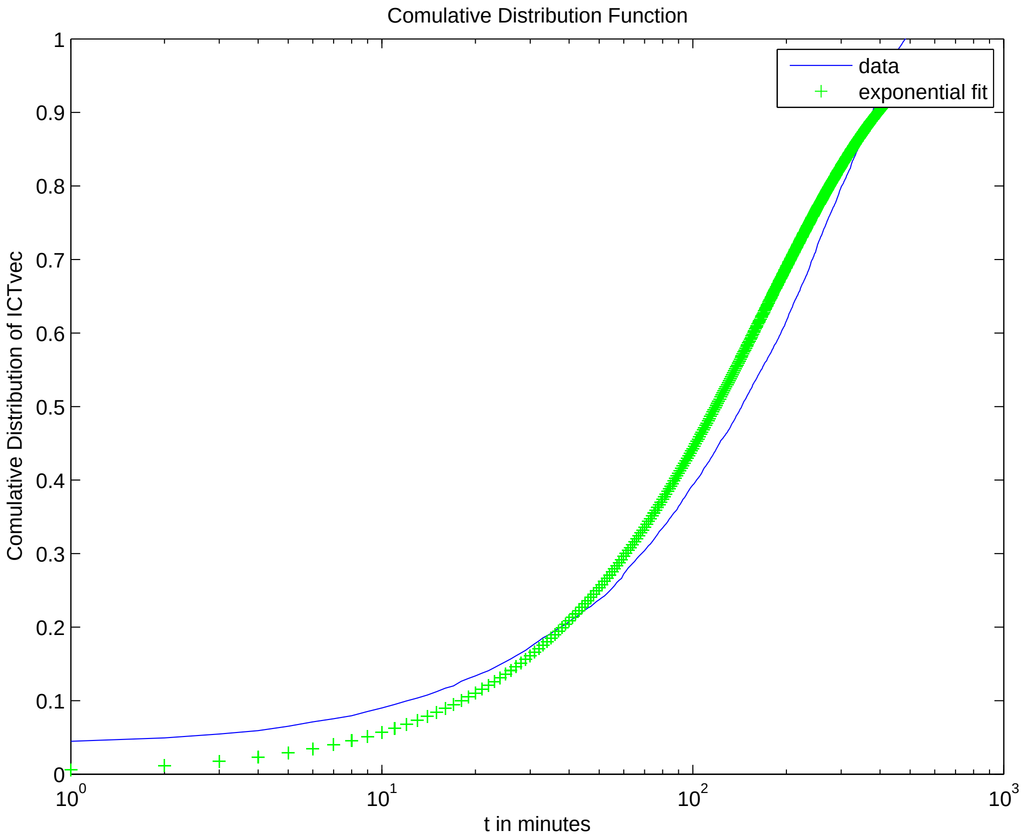


Figure 4.21: CDF ICTvec day4

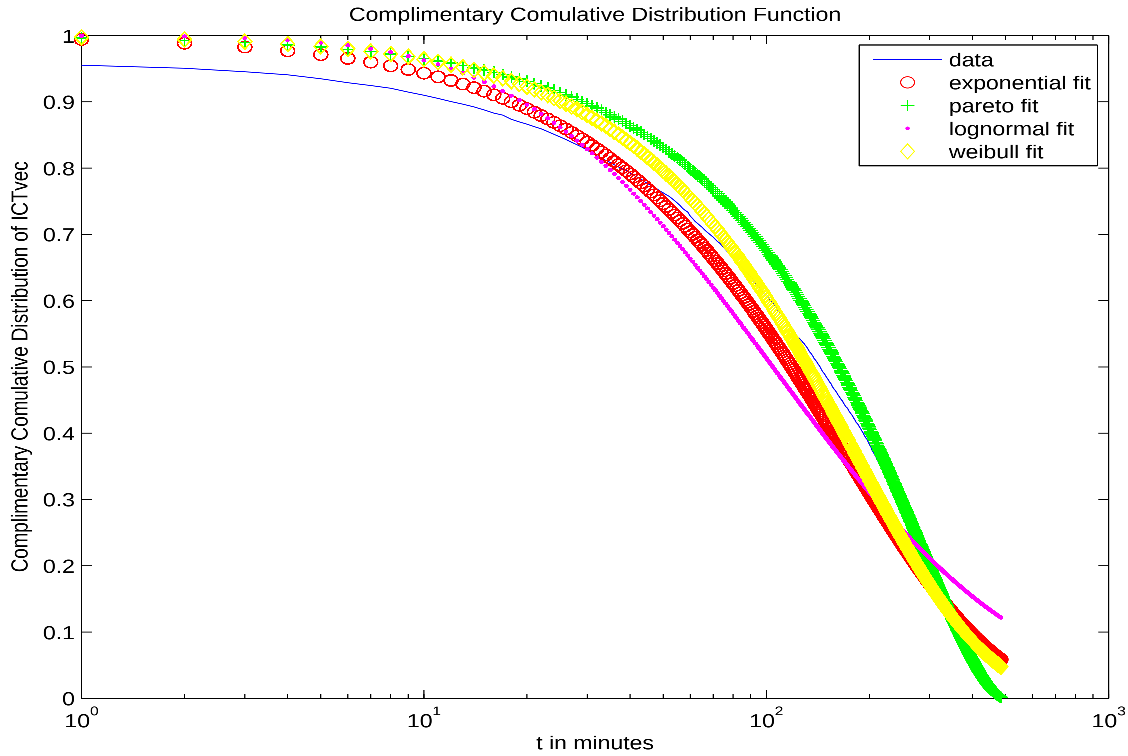


Figure 4.22: CCDF ICTvec day 4

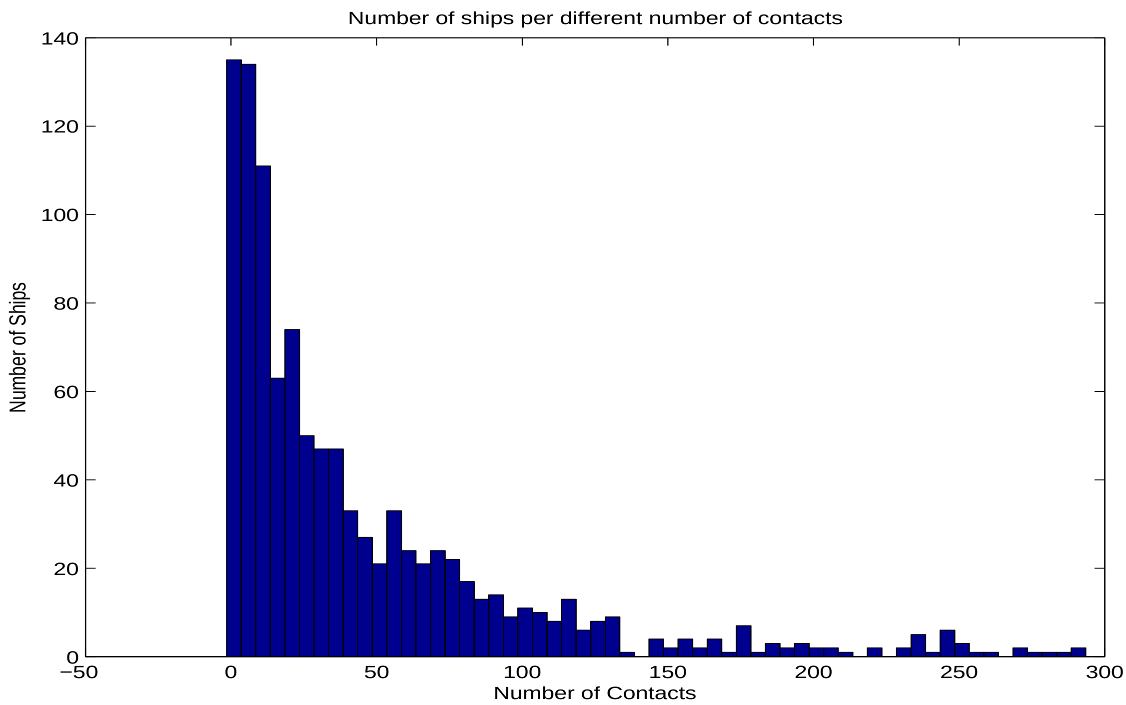


Figure 4.23: Number of ships per contact day 4

Distribution	Parameters		
exponential	$\mu = 175.193$		
pareto	$\kappa = -0.0298362$	$\sigma = 1.57213$	$\vartheta = 0$
lognormal	$\mu = 0.349076$	$\sigma = 0.370251$	
Weibull	$a = 1.71631$	$b = 1.76505$	

Table 4: Distribution Parameters

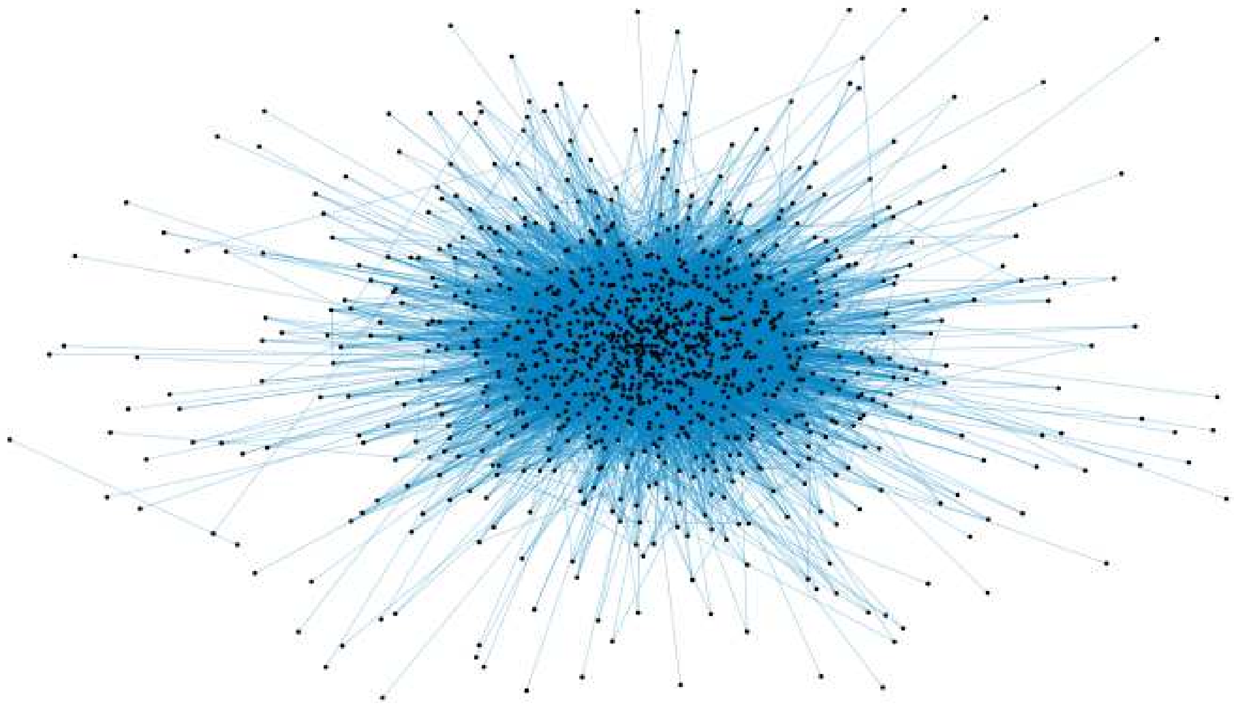


Figure 4.24: Contact graph day 4

Day4 data set has similar number of entries as the day1 one data set. Contact duration histogram shows that the values of contact duration of day4 are concentrated at the initial values and there are no missing values as it happens in the day1 data set. *ICTvec* histogram shows that there is a higher concentration at the initial values as expected from the result in day1, but the final values of ICTvec are much larger than the ones in day1. Figures 29 and 30 show the ECDF and CCDF of the actual data and their fitting curves. Table5 contains the parameters of the fitting curves. The histogram that contains the ships per contact is similar to the one from day1, but again there is bigger distribution and the final values are larger than in day1. Finally, the contact graph created is densely populated because of the size of the data set and is shown in Figure32.

4.1.2 Scenario 1 *dist=0.5* and *thresh_t=1*

Scenario 1 has a different distance, in the range of which two vessels consider to have a contact. The new distance is the half of the distance in the baseline scenario and it is *dist = 0. 5* km. In this case the results are expected to be different from the results derived from the baseline scenario. The scenario will be thoroughly studied for the second day, which has the largest data set, with the most entries, because the data sets for the first and the fourth day are similar to this one, and the data set for the third day, with the least of entries is an outlier. All results from the simulation will be cited in the Appendix for further evaluation and study. A comparative table with the changes of the fitting parameter for each fitting distribution and for each day will also be given.

Starting, the histogram of the contact duration for the data set of the second day is:

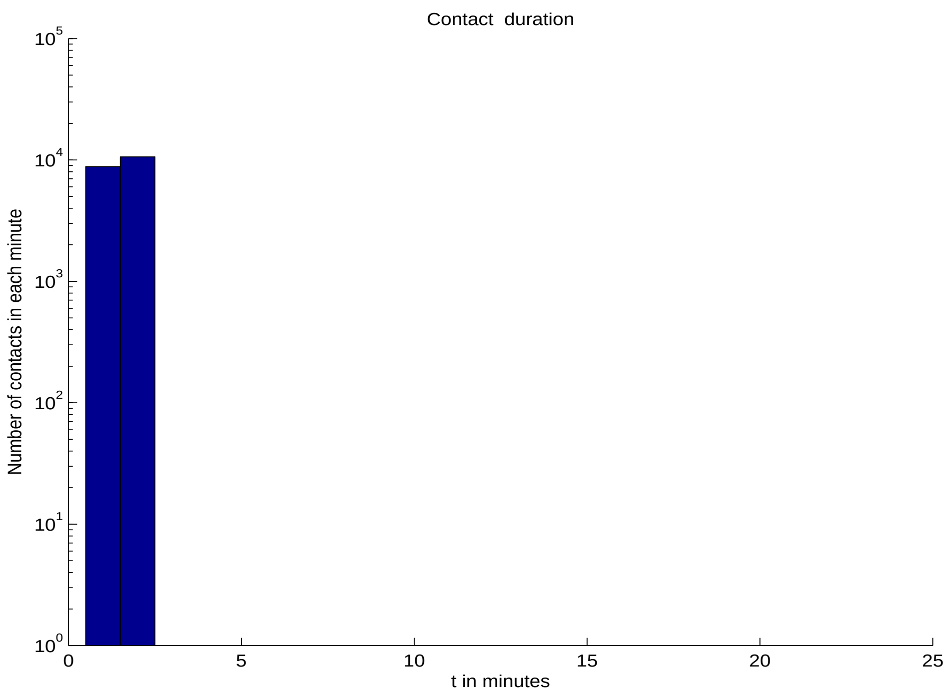


Figure 4.25: Contact duration – Scenario 1 –day 2

The results show that the duration of the contacts is only one or two minutes and there are no larger durations. It is an expected result because the distance is set at 500m, and vessels must be within that range in order to contact. The number of contact though is the same for the one or two minute contacts, with that represented in the baseline scenario for the second day. The distance upper bounds the duration of the contacts in this case.

Continuing with the histogram create from the variable *ICTvec:*

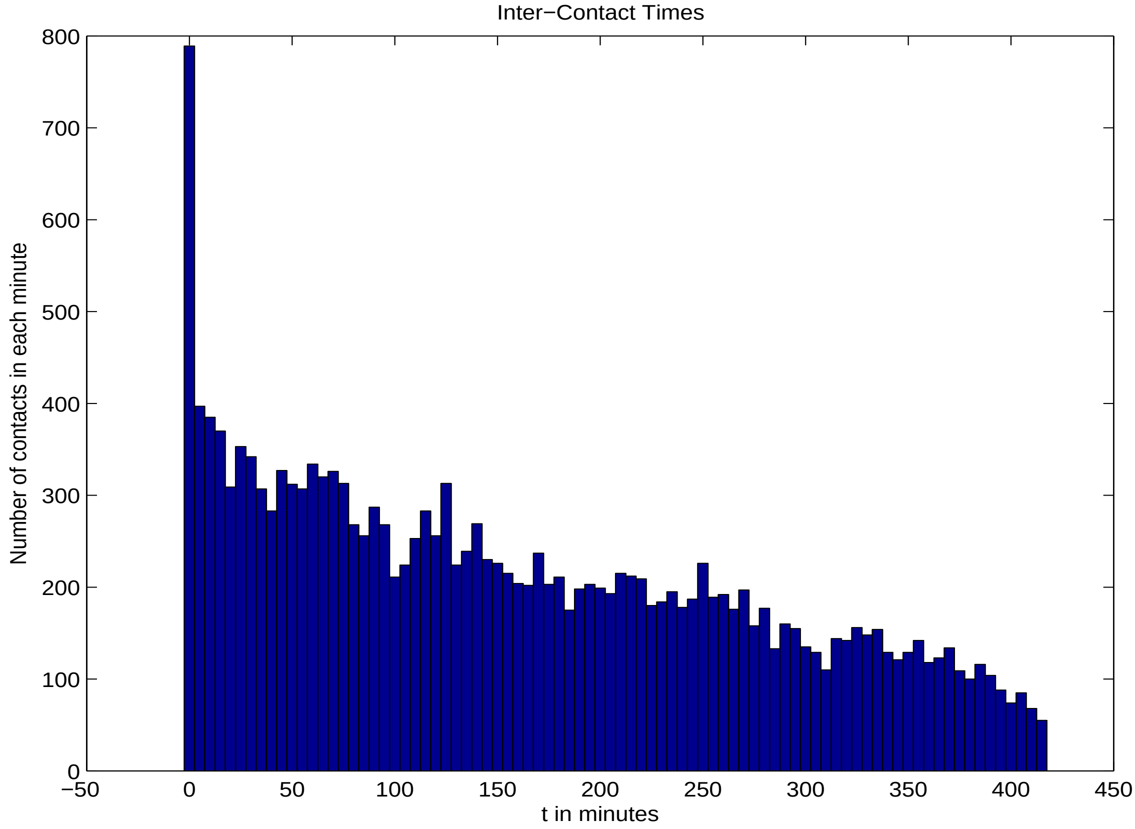


Figure4.26: Histogram ICTvec – Scenario 1 – day 2

The histogram has almost the same range in time starting from zero minutes to four hundred and twenty minutes for the contacts. It is also follows the same curve as it does in the baseline scenario for the second day. The curve is a decreasing function in time and for contacts with times less than five minutes the concentration is higher, but after that is drops almost in half and afterwards follows a smooth decreasing curve. The number of contacts for each time in the histogram is almost the half of the number of contacts in the histogram of the baseline scenario.

The following network characteristics are the curves of ECDF and CCDF for the variable *ICTvec.* The curves show the actual graph created from the data of the second day, as well as the curves of the fitting distributions. The chosen fitting distribution are: the exponential, the generalized pareto, the lognormal distribution and the Weibull distribution. They are considered for the reasons given in the baseline scenario. For all distribution, except the exponential, the zero values were chosen to be omitted.

The ECDF curves are:

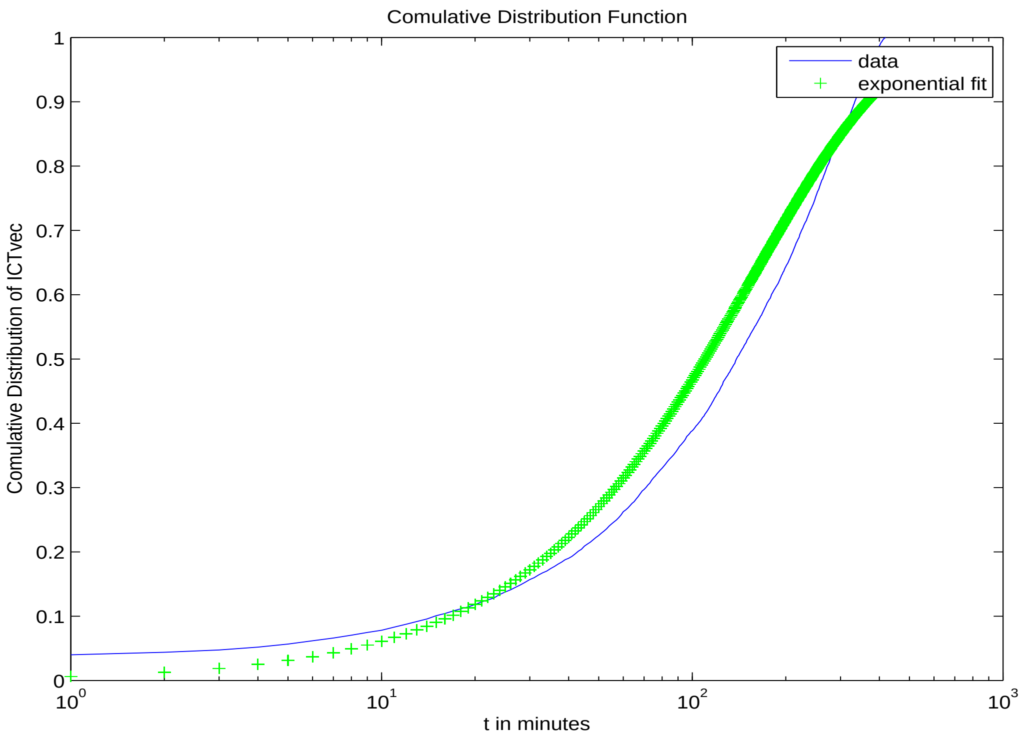


Figure 4.27: CDF ICTvec – Scenario 1 –day 2

The graph shows that the ECDF curve of the data has almost a constant value of probability near zero for the first nine minutes and after that follows the theoretical curve of the exponential distribution that fits the *ICTvec* data and becomes one approximately at time = 400 mins. At the first part of the curve the data is under – fitted by the exponential curve, and for the second part they are over – fitted. The graph is quite similar with the one derived from the baseline scenario, and this happens because the *ICTvec* variable presents the same distribution with the baseline scenario, even if the number of contacts is almost the half for each time. Also the mean of the exponential distribution is almost the same in both scenarios. For the baseline scenario, as it has already be stated the mean is $\mu = 166.16$ and for the scenario, where the distance is 500m the mean is$\mu = 163.018$. The hypothesis explaining the above fact it that most of the contacts in the baseline scenario happen in a distance of 500m and there are few contacts happening in distances between 500m and 1km. CCDF graph shows similar results. The curves have a constant value near one for the first nine minutes and the drop following the theoretical lines. This happens because CCDF is the complimentary value of ECDF.

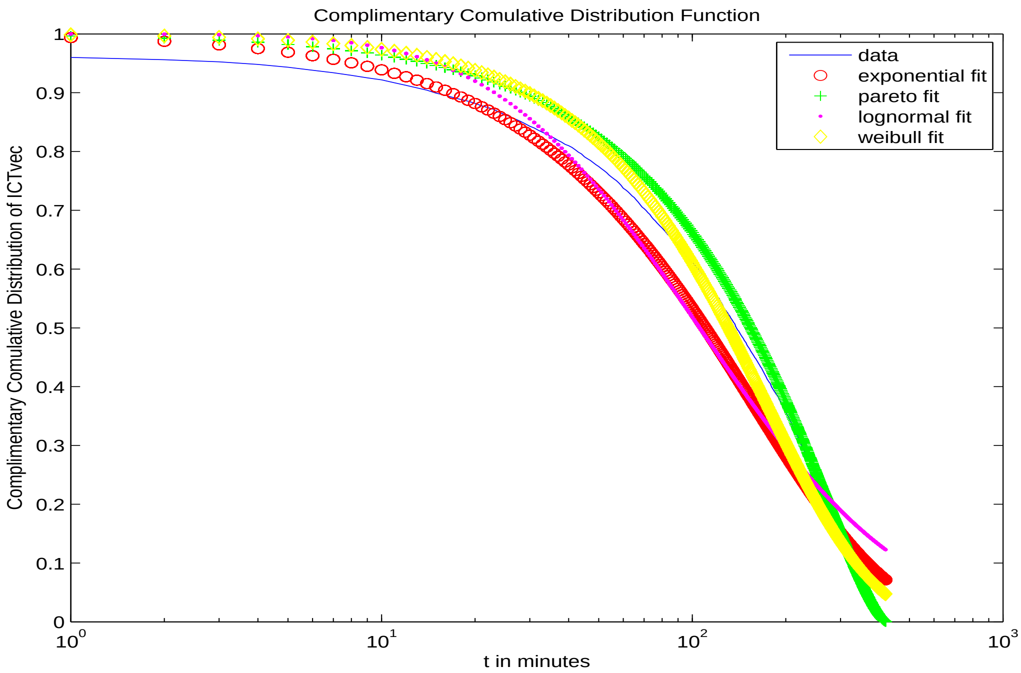


Figure 4.28: CCDF ICTvec – scenario 1- day 2

Finally, the histogram for ships per number of contacts is:

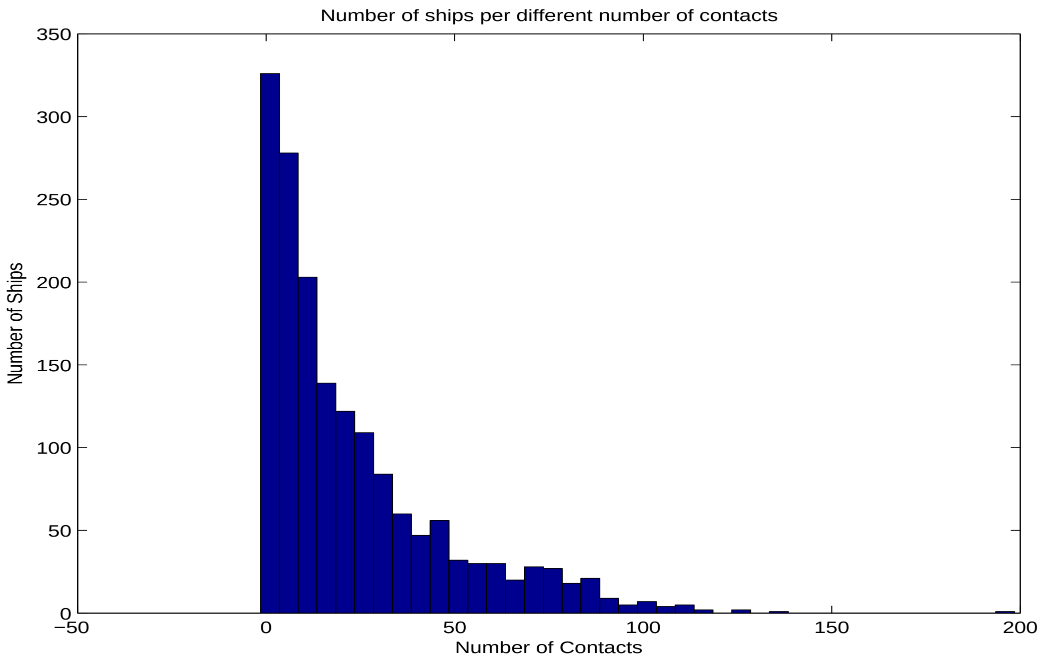


Figure 4.29: Number of ships per contact – Scenario1 – day 2

The histogram shows that the maximum number of ships per number of contacts is larger than the one on the baseline scenario. Also the curve drops in a smoother way than before, but the maximum number of contacts is less than half of what was the result in the baseline scenario.

The contact graph created for the date of the second day and *thresh_t = 1*, *dist = 0.5* is:

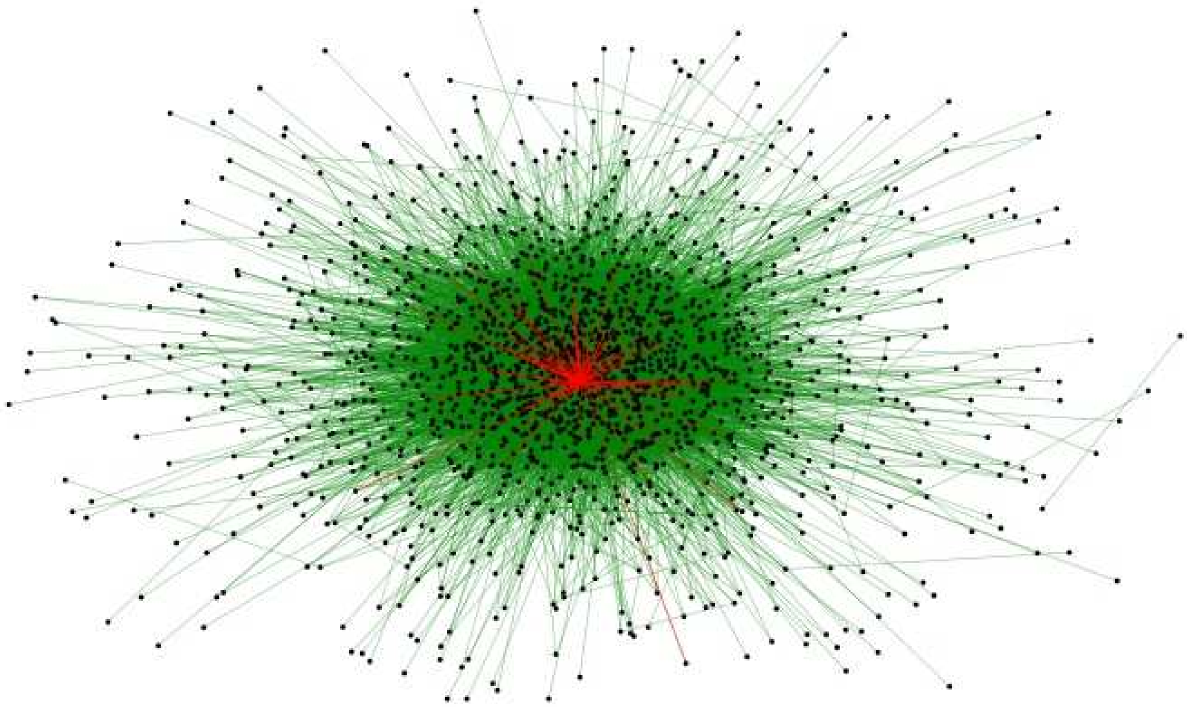


Figure 4.30: Contact graph – scenario 1- day 2

The graph is very densely populated especially in the centre of it and there is less density at the perimeter of the graph. The explanation is the size of the data set and the number of ships that participate in contacts are very large. Also the ship ids are sequential, as a result two different ships that participate in many contacts, and have ids that are two sequential numbers are located very closely in the center of the graph. The graph represent a true opportunistic network, as it is studied.

Finally, the comparative table containing all the parameters of the distribution functions is shown below:

Day 1			
Distribution	Parameters		
exponential	$\mu = 113.261$		
pareto	$\kappa = -0.0833885$	$\sigma = 1.61567$	$\vartheta = 0$
Lognormal	$\mu = 0.371049$	$\sigma = 0.347938$	
Weibull	$a = 1.70218$	$b = 2.57176$	
Day 2			
Distribution	Parameters		
exponential	$\mu = 163.018$		
pareto	$\kappa = -0.0758807$	$\sigma = 1.62068$	$\vartheta = 0$
lognormal	$\mu = 0.379502$	$\sigma = 0.347254$	
Weibull	$a = 1.71346$	$b = 2.58977$	
Day 3			
Distribution	Parameters		
exponential	$\mu = 8.99197$		
pareto	$\kappa = -1.79186$	$\sigma = 3.58372$	$\vartheta = 0$
lognormal	$\mu = 0.36414$	$\sigma = 0.346247$	
Weibull	$a = 1.70473$	$b = 3.571$	
Day 4			
Distribution	Parameters		
exponential	$\mu = 170.618$		
pareto	$\kappa = -0.0668802$	$\sigma = 1.53592$	$\vartheta = 0$
lognormal	$\mu = 0.321162$	$\sigma = 0.352729$	
Weibull	$a = 1.63748$	$b = 2.21254$	

Table 5: Concentrated Table with Distribution Parameters

The comparative table shows that the three days which have similar data set sizes have similar parameters for each distribution. The third day with the smallest data set is an outlier and the parameters of each distribution are significantly different from the same parameters of the same distribution in every other day.

4.1.3 Scenario 2 *dist=2* and *thresh_t=1*

The second scenario examined is the scenario where time remains the same, while the distance in which to vessels can communicate and the communication to be considered a contact grows at 2 km. Before, implementing the scenario a hypothesis can be done: more contacts are to be expected because the vessels find more vessels further apart, but these contact will logically be sparse. Again, day2 data set is presented in detail, but all data set from all day involved in the application and their corresponding results, against the scenario, are cited in the Appendix. The first metric to be presented is the duration of the contacts. A histogram is made from the simulator and it is show below:

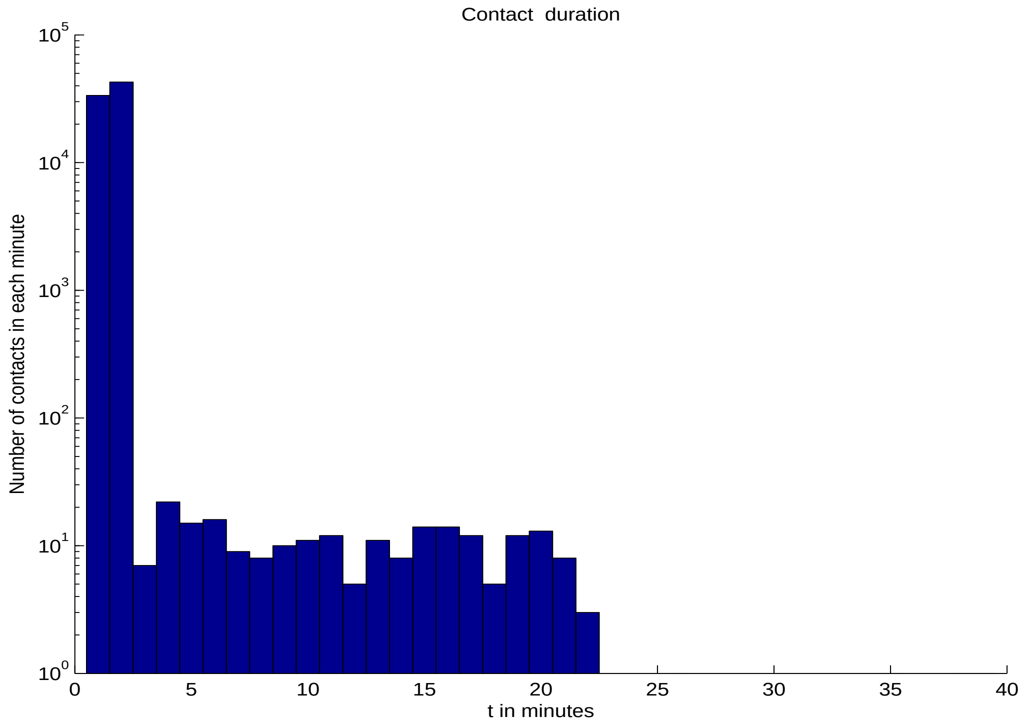


Figure4.31: Contact duration – scenario 2 –day 2

The histogram shows that the density of the number of contacts is the highest in the first two bins which are almost the same as in the baseline scenario, but the rest of the bins till the time of contacts is twenty two minutes is full and most of the bins show the contacts to be around ten for each one. There is also one bin further apart at thirty six minutes that has only one contact. This means that one contact lasts thirty six minutes.

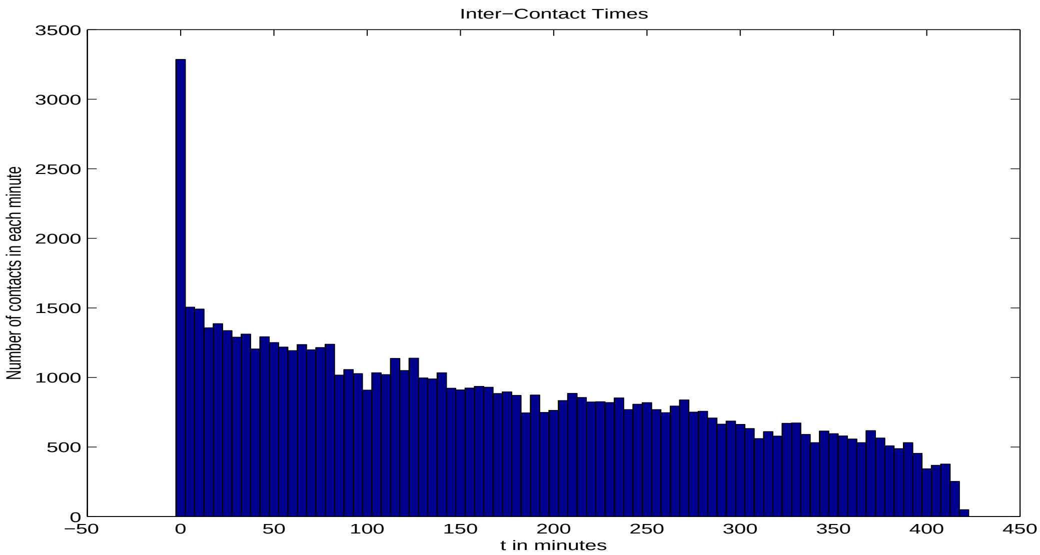


Figure4.32: Histogram ICTvec –scenario2 –day 2

The histogram corresponding to the variable *ICTvec* shows that the highest concentration of contacts is at zero minutes. At that time there are 3400 contacts. Afterwards the contacts drop at 1500 and they continue to decrease as time passes. The contacts is a decreasing function in time. Comparing with the baseline scenario, the function is smoother and the number of contacts that exist for scenario 2 are almost one and a half times more and the minutes of the contacts are over 400 hundred minutes. The ECDF and CCDF graphs with the corresponding fitting distributions are given below:

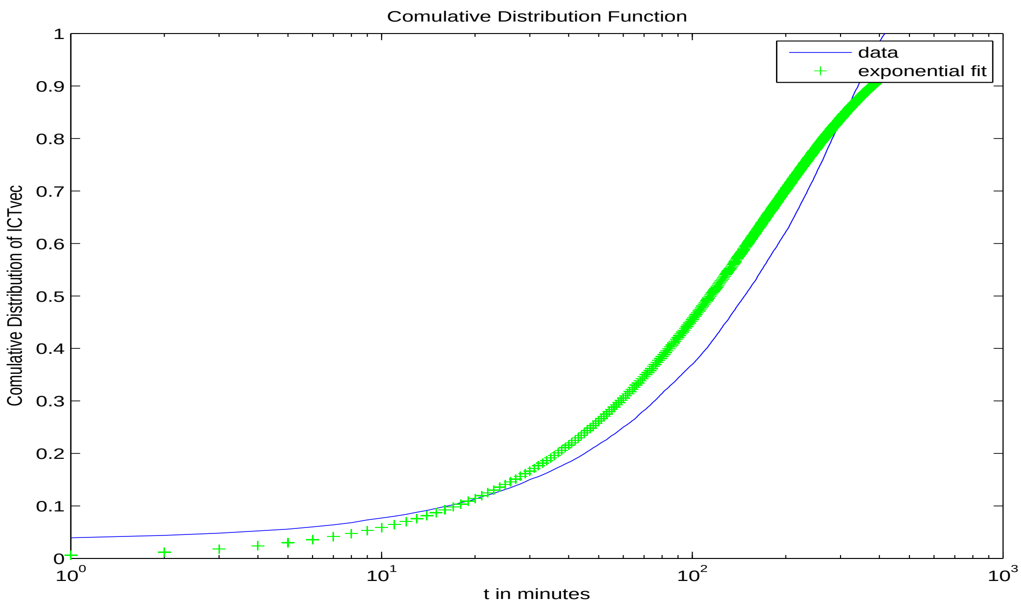


Figure4.33: CDF ICTvec – scenario 2 –day2

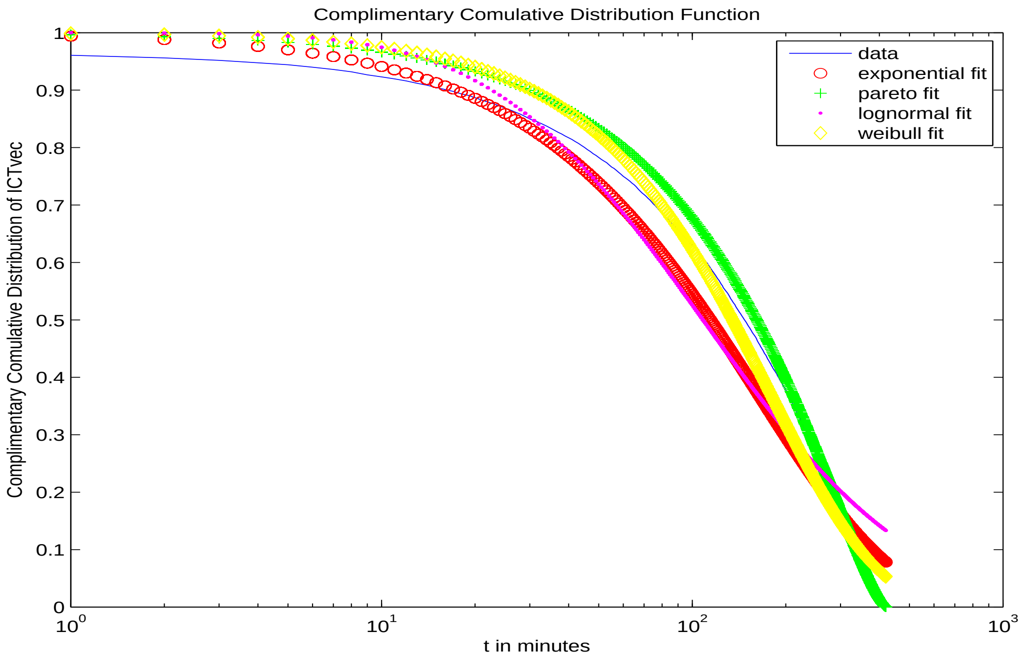


Figure 4.34: CCDF ICTvec –scenario2 –day2

For both ECDF and CCDF curves the comments made for scenario 1 can be applied and for scenario 2. The final metric is the histogram that shows the number of ships per number of contacts

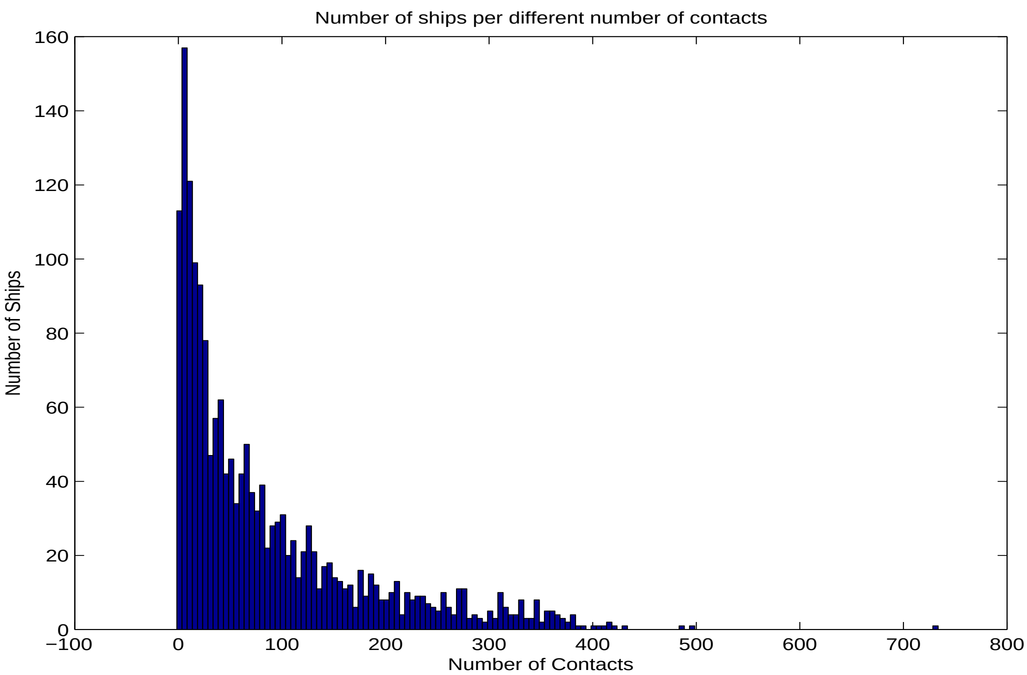


Figure 4.35: Number of ships per contact – scenario 2 – day 2

It is shown that the concentrations are higher for the initial values of the contacts, and then the curve dramatically drops. There are also ships that have many hundreds of contacts, but they are very few. The contact graph is also densely populated and complicated as in scenario 1 for the same reasons.

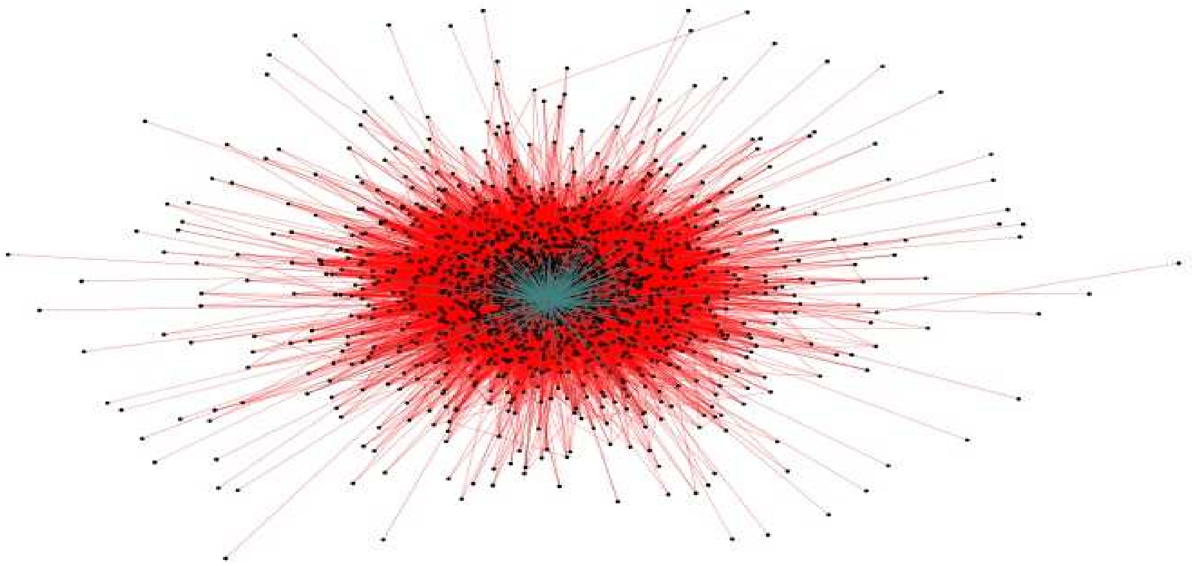


Figure 4.36: Contact graph – scenario 2 – day 2

Also the table with all the parameters of the distribution function is given:

Day 1			
Distribution	Parameters		
exponential	$\mu = 119.878$		
pareto	$\kappa = -0.0610995$	$\sigma = 1.64986$	$\vartheta = 0$
Lognormal	$\mu = 0.391467$	$\sigma = 0.357329$	
Weibull	$a = 1.76091$	$b = 2.023$	
Day 2			
Distribution	Parameters		
exponential	$\mu = 168.445$		
pareto	$\kappa = -0.0404183$	$\sigma = 1.63512$	$\vartheta = 0$

lognormal	$\mu = 0.393543$		$\sigma = 0.360324$
Weibull	$a = 1.77092$		$b = 1.93752$
Day 3			
Distribution	Parameters		
exponential	$\mu = 9.83819$		
pareto	$\kappa = -3.3131$	$\sigma = 6.62619$	$\vartheta = 0$
lognormal	$\mu = 0.369041$		$\sigma = 0.345871$
Weibull	$a = 1.71192$		$b = 1.93752$
Day 4			
Distribution	Parameters		
exponential	$\mu = 175.963$		
pareto	$\kappa = -0.009185$	$\sigma = 1.65258$	$\vartheta = 0$
lognormal	$\mu = 0.37679$		$\sigma = 0.409759$
Weibull	$a = 1.823201$		$b = 2.48729$

Table 6: Concentrated Table of Distribution Parameters

4.1.4 Scenario 3 *dist=1* and *thresh_t=2*

Scenario 3 changes the time threshold to two minutes comparing with one that existed in the baseline scenario. Considering two minutes of uncertainty between the transmissions of AIS from the different vessels is not a leap. Most of the vessels in the data set do not transmit AIS data each minute. There is a mean time space of three minutes between the transmissions. Choosing the two minutes time gives the opportunity to ships to find the contacts that in any other occasion would be missed in the simulation. The expected results are to be more contacts during that times, and various concentrations of contact duration, ICTvec and the ECDF/CCDF curves.

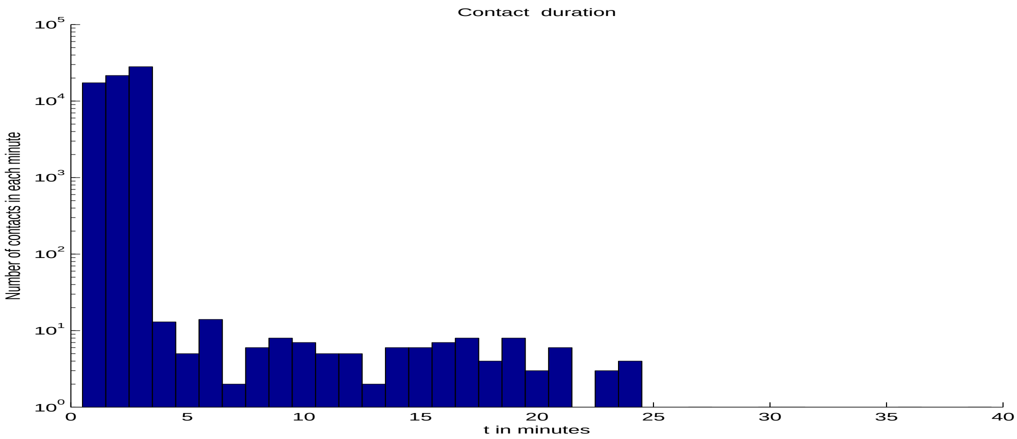


Figure 4.37: Contact duration – scenario 3- day2

Contact duration histogram shows that the values of contact duration are more scattered in scenario 3, based on the results taken on the baseline scenario. Most of the contacts last three minutes and the variation of the values is showing on the histogram. The second histogram is the one provided from the ICTvec variable:

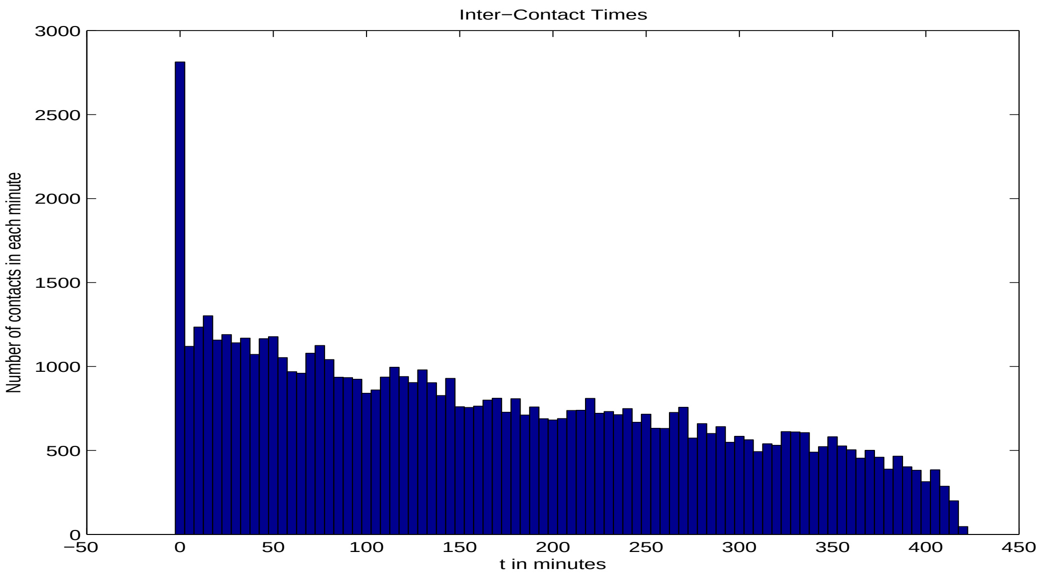


Figure 4.38: Histogram ICTvec – scenario 3 –day 2

The histogram shows a high concentration at zero, but afterwards the drop is smooth and the values are comparable. The higher the time the curve goes more to a straight line. The ECDF curves are:

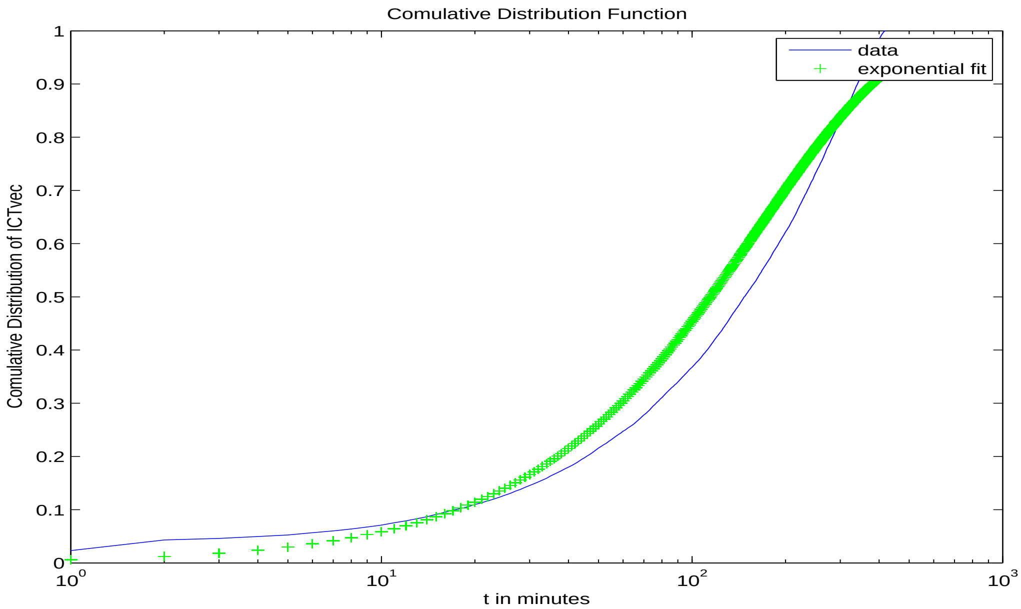


Figure 4.39: CDF ICTvec – scenario 3 – day 2

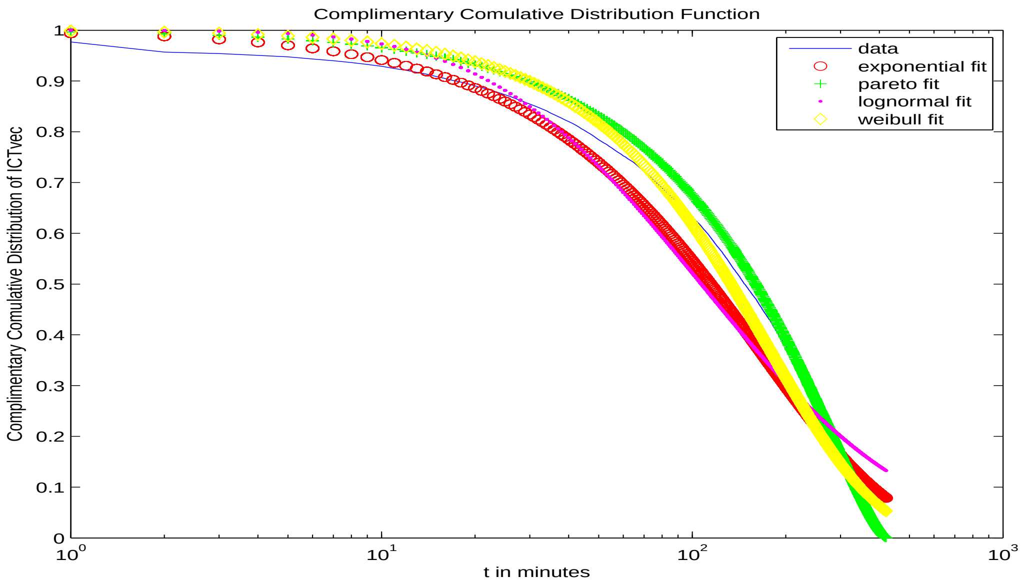


Figure4.40: CCDF ICTvec – scenario 3 – day 2

Finally, the histogram containing the number of ships per number of contacts is:

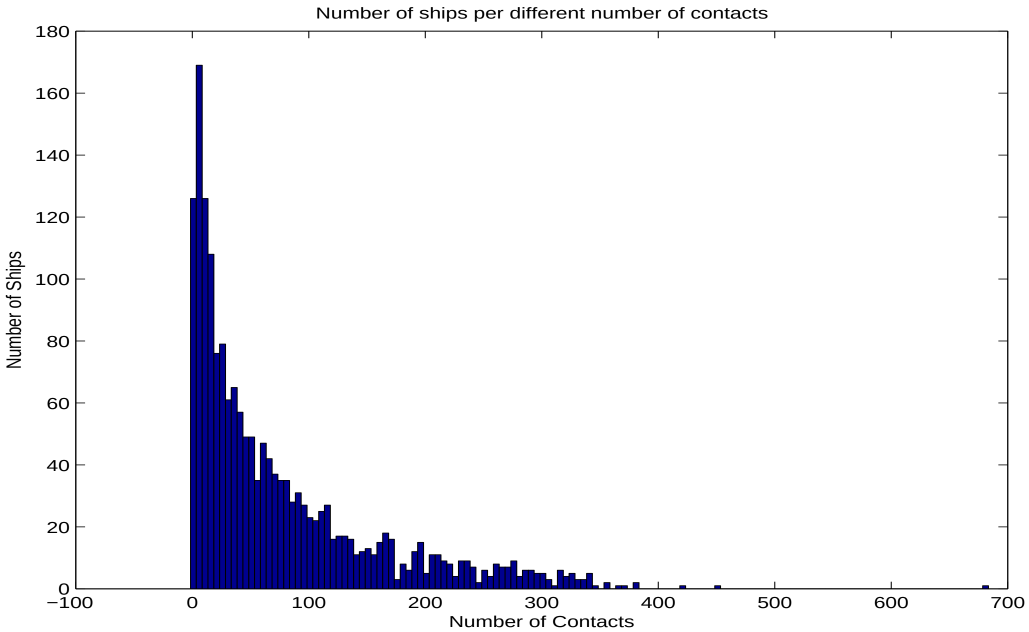


Figure4.41: Number of ships per contact – scenario 3 – day 2

And the contact graph:

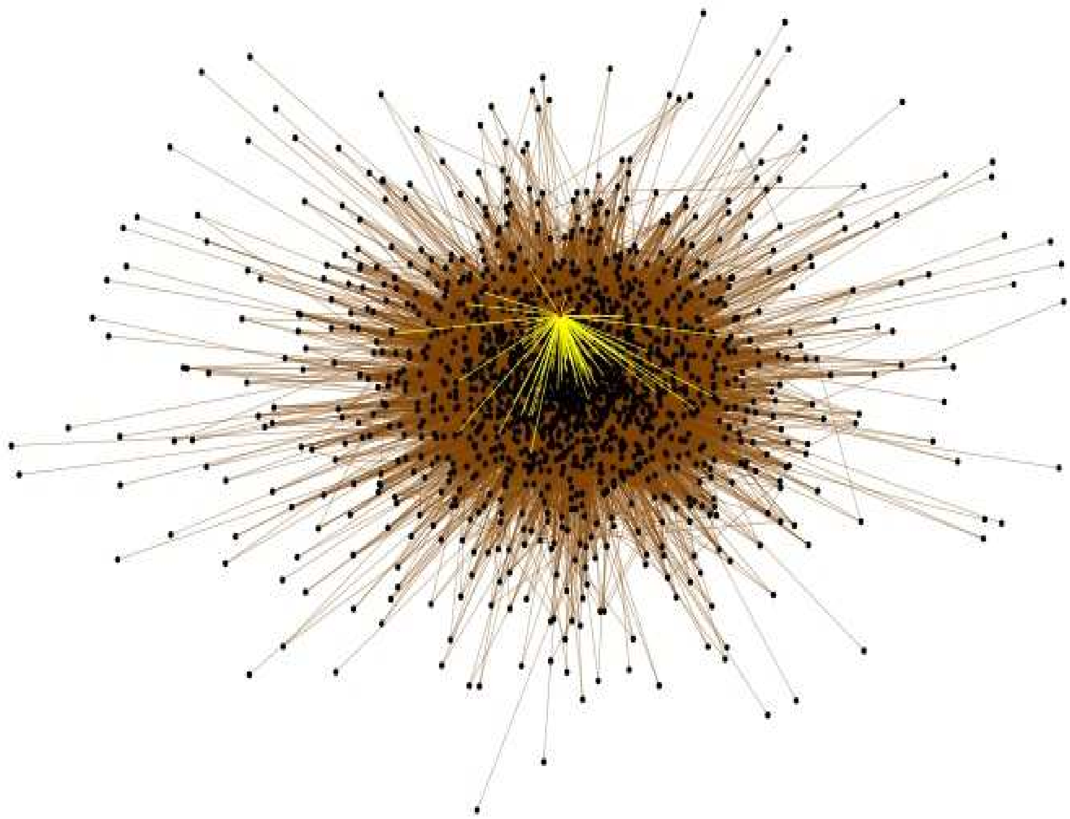


Figure 4.42: Contact graph – scenario 3- day 2

The comparative table which gives all the parameters of the distributions functions is:

Day 1			
Distribution	Parameters		
exponential	$\mu = 116.584$		
pareto	$\kappa = -0.095798$	$\sigma = 2.32094$	$\vartheta = 0$
Lognormal	$\mu = 0.688427$	$\sigma = 0.446417$	
Weibull	$a = 2.4406$	$b = 2.2302$	
Day 2			
Distribution	Parameters		
exponential	$\mu = 167.595$		
pareto	$\kappa = -0.0573236$	$\sigma = 2.26238$	$\vartheta = 0$
lognormal	$\mu = 0.687251$	$\sigma = 0.446144$	
Weibull	$a = 2.4365$	$b = 2.15158$	
Day 3			
Distribution	Parameters		
exponential	$\mu = 9.60035$		
pareto	$\kappa = -1.88876$	$\sigma = 5.66629$	$\vartheta = 0$
lognormal	$\mu = 0.69869$	$\sigma = 0.444987$	
Weibull	$a = 2.46613$	$b = 3.10935$	
Day 4			
Distribution	Parameters		
exponential	$\mu = 174.828$		
pareto	$\kappa = -0.0474749$	$\sigma = 2.3428$	$\vartheta = 0$
lognormal	$\mu = 0.694879$	$\sigma = 0.48182$	
Weibull	$a = 2.53255$	$b = 1.72958$	

Table 7: Concentrated Table of Distribution Parameters

4.1.5 Scenario 4 *dist=0.5* and *thresh_t=2*

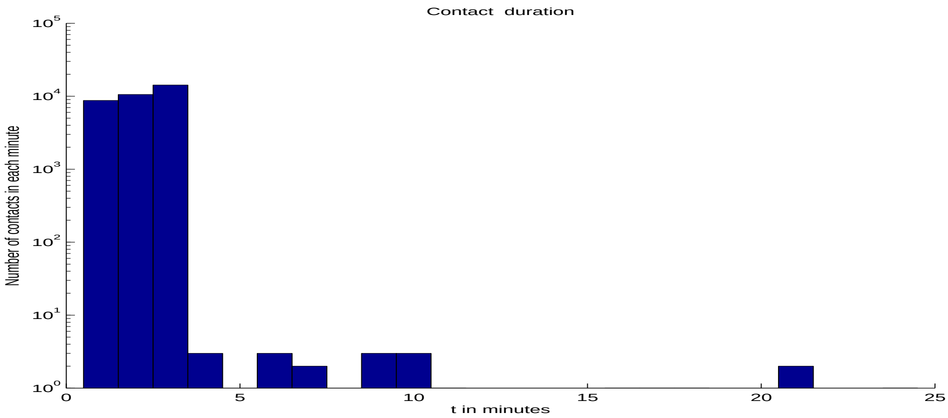


Figure 4.43: Contact duration – scenario 4 – day2

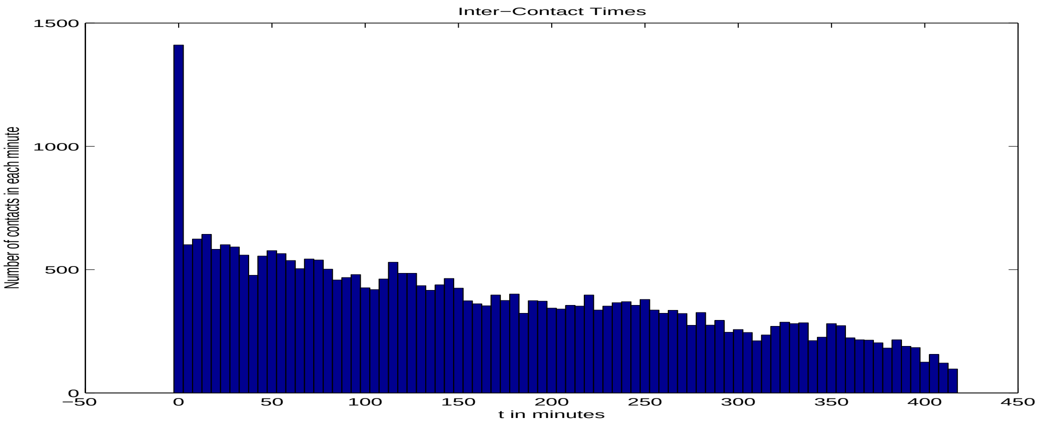


Figure 4.44: Histogram ICTvec – scenario 4 – day 2

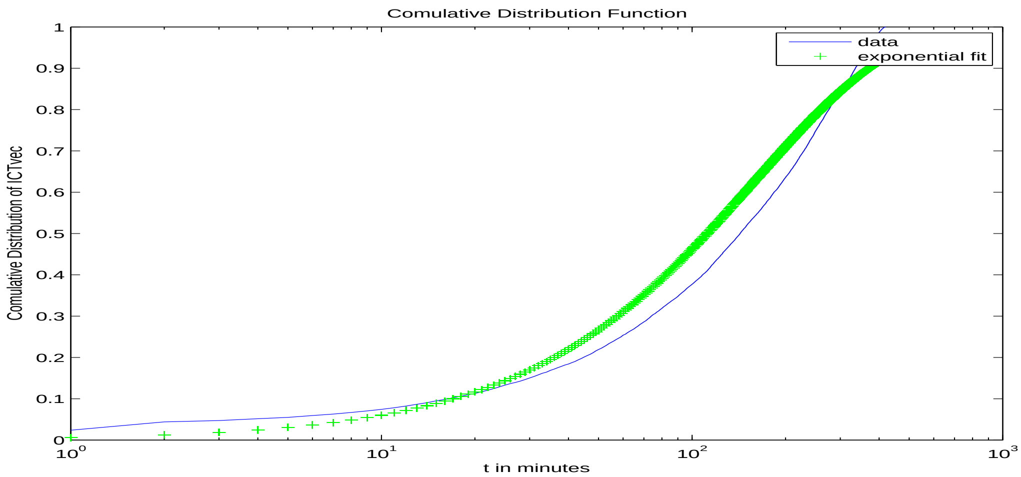


Figure 4.45: CDF ICTvec – scenario 4 – day 2

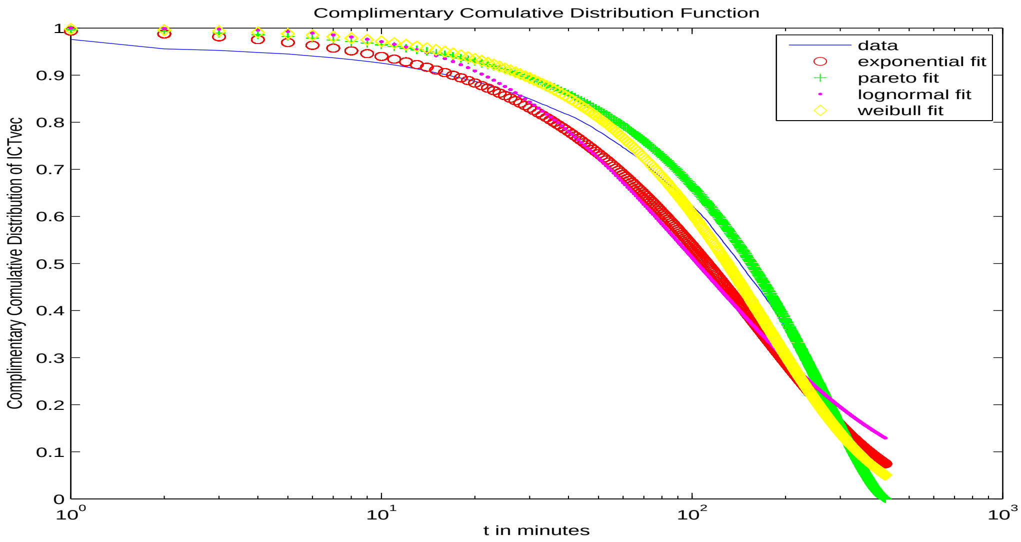


Figure4.46: CCDF ICTvec – scenario 4 –day 2

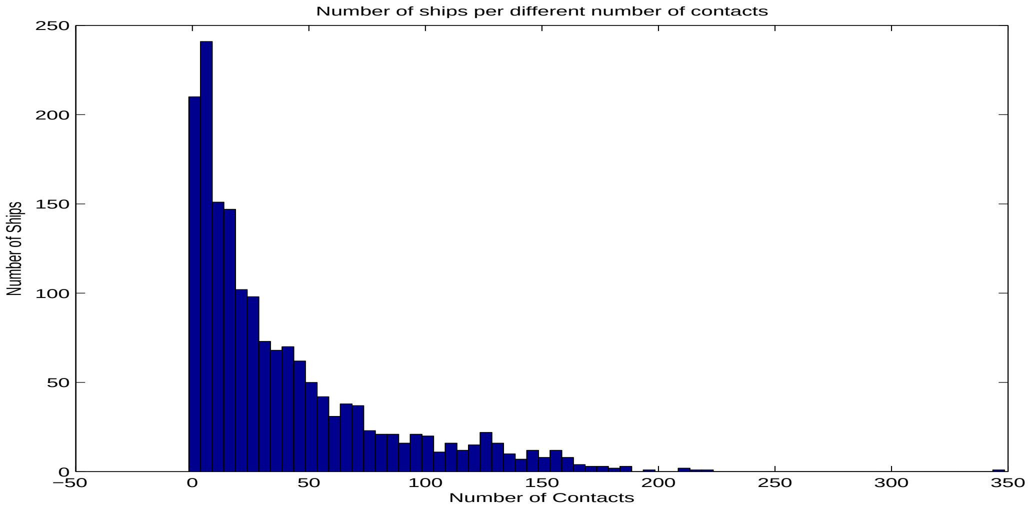


Figure 4.47: Number of ships per contact – scenario 4 –day 2

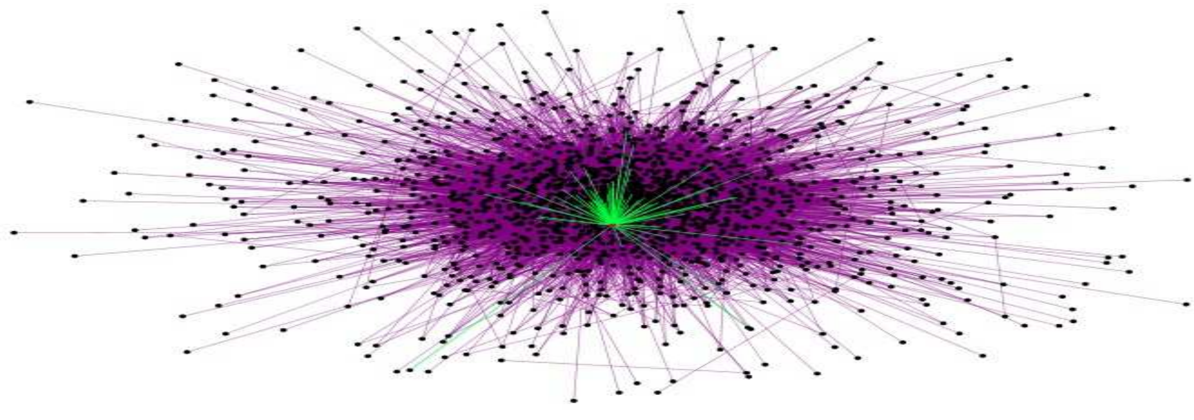


Figure 4.48: Contact graph – scenario 4 –day 2

Finally, the table with all the parameter of the distribution used in the fitting curves is:

Day 1			
Distribution	Parameters		
exponential	$\mu = 116.584$		
pareto	$\kappa = -0.095798$	$\sigma = 2.32094$	$\vartheta = 0$
Lognormal	$\mu = 0.688427$	$\sigma = 0.446417$	
Weibull	$a = 2.4406$	$b = 2.2302$	
Day 2			
Distribution	Parameters		
exponential	$\mu = 167.595$		
pareto	$\kappa = -0.0573236$	$\sigma = 2.26238$	$\vartheta = 0$
lognormal	$\mu = 0.687251$	$\sigma = 0.446144$	
Weibull	$a = 2.4365$	$b = 2.15158$	
Day 3			
Distribution	Parameters		
exponential	$\mu = 9.60035$		
pareto	$\kappa = -1.88876$	$\sigma = 5.66629$	$\vartheta = 0$
lognormal	$\mu = 0.69869$	$\sigma = 0.444987$	
Weibull	$a = 2.46613$	$b = 3.10935$	
Day 4			
Distribution	Parameters		
exponential	$\mu = 174.828$		
pareto	$\kappa = -0.0474749$	$\sigma = 2.3428$	$\vartheta = 0$
lognormal	$\mu = 0.694879$	$\sigma = 0.48182$	
Weibull	$a = 2.53255$	$b = 1.72958$	

Table 8: Concentrated Table of Distribution Parameters

Due to the fact that the distances is smaller then than in the baseline scenario, there are fewer contacts and the durations of the contacts are concentrated in the beginning of the histograms. There are few higher values, which exist because the time threshold is set at two minutes. From the table, the variation of the distribution functions parameters is obvious.

4.1.6 Scenario 5 *dist=2* and *thresh_t=2*

The final scenario tested. In this scenario there is the largest time threshold at two minutes and the largest distance in which two vessels can communicate. Based on all previous scenarios, it is expected larger contact durations, as well as larger distribution of contacts at the initial times. Also, times going further apart and larger number of contact are expected. The number of ships per number of contact are expected to present a peak in the initial values of the numbers of contacts and the drop, but few ships are expected to have a lot of contacts. ECDF and CCDF curves probably follow the theoretical curves of different distribution function and their parameters will be shown in the comparative table. The contact graph is expected to be densely populated, especially in the centre, while in the far sides of it, the population is decreased. All the above conclusions are shown for the data set of the second day, which is the largest data set, while the rest is cited in the Appendix.

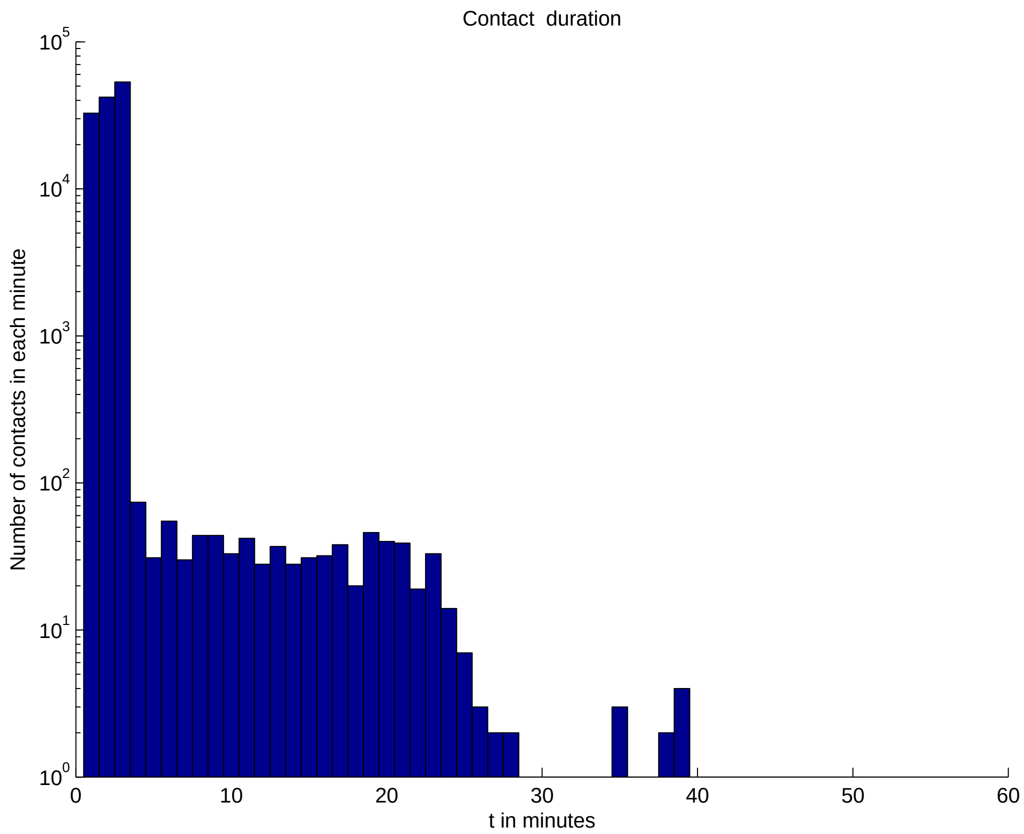


Figure 4.49: Contact duration – scenario 5 – day 2

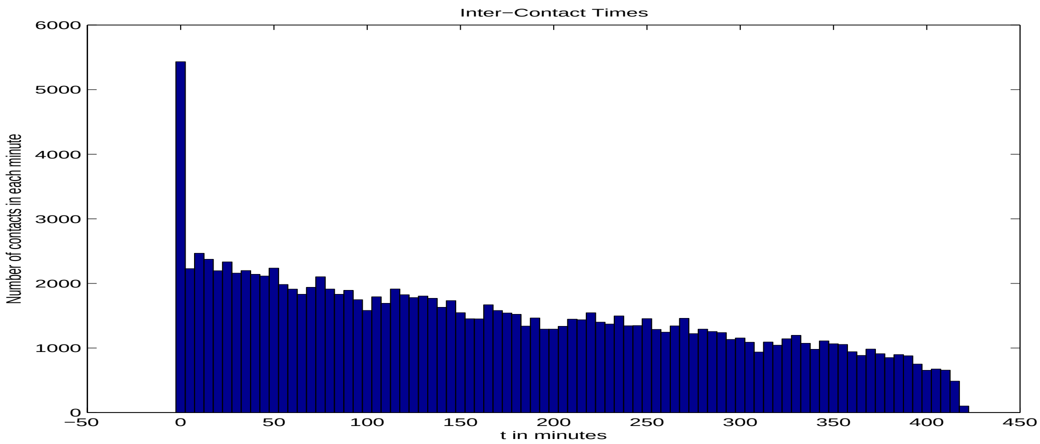


Figure 4.50: Histogram ICTvec – scenario 5 – day 2

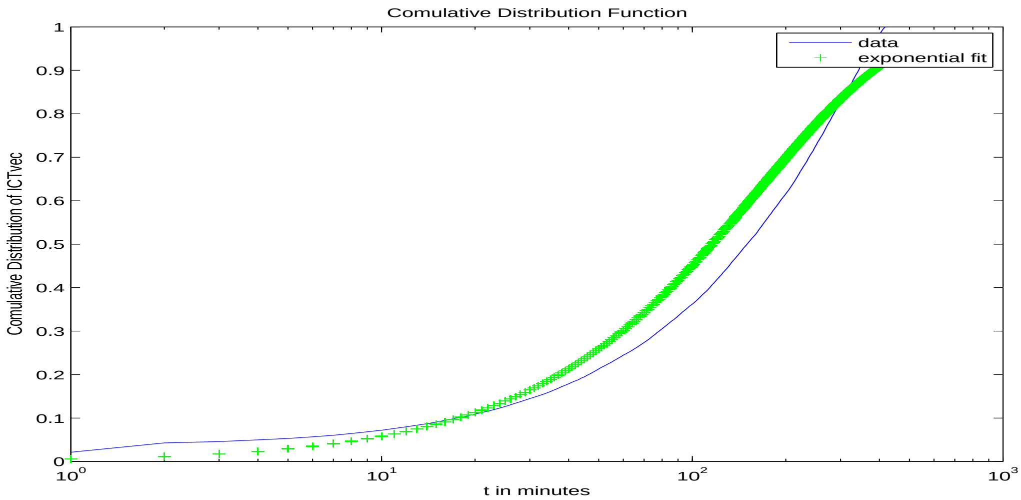


Figure 4.51: CDF ICTvec – scenario 5 – day 2

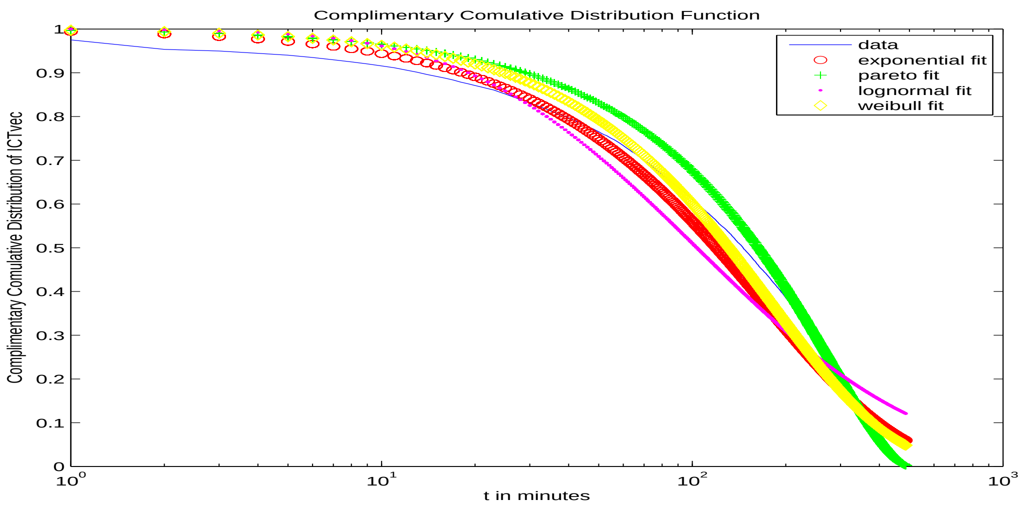


Figure 4.52: CCDF ICTvec – scenario 5 – day 2

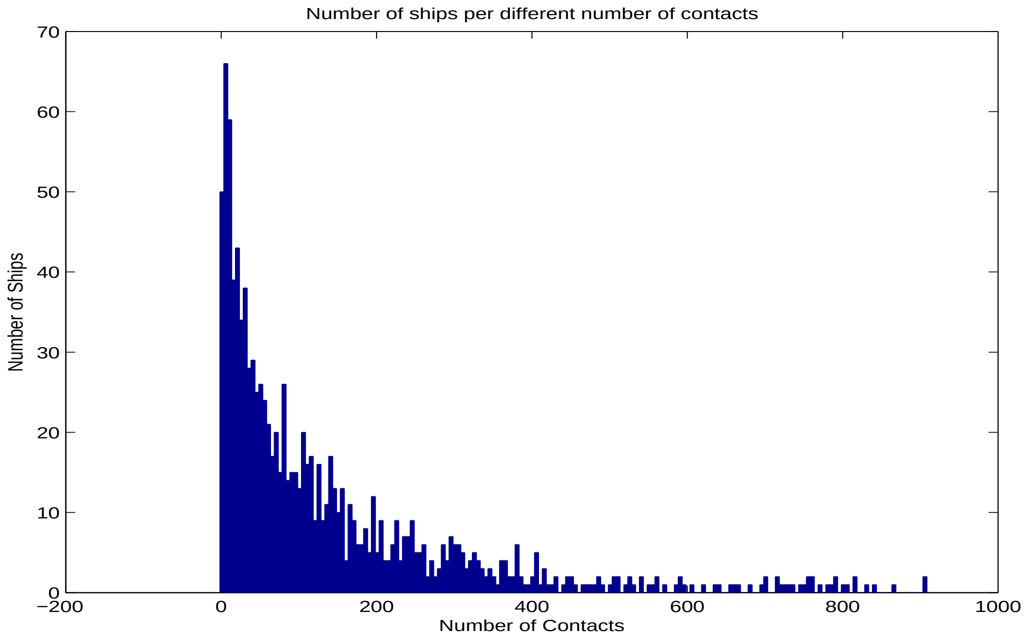


Figure 4.55: Number of ships per contact – scenario 5 –day2

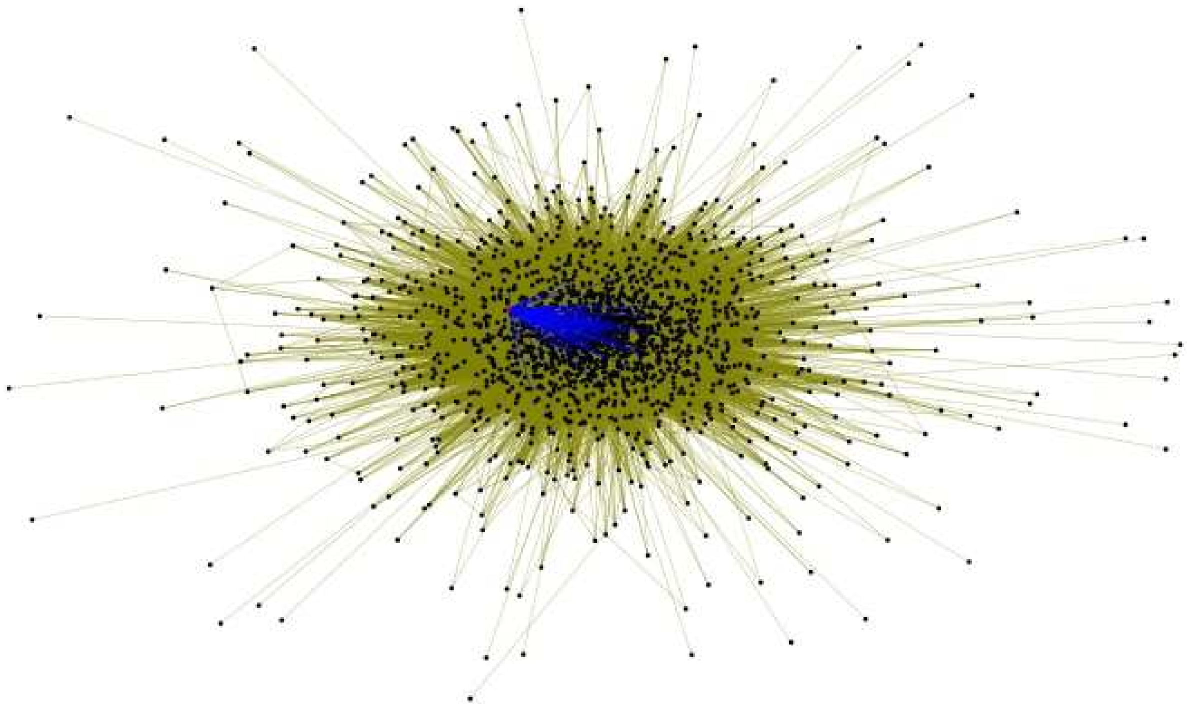


Figure 4.54: Contact graph – scenario 5 – day2

Finally, the table containing all the parameters derived from the distribution function for each day is:

Day 1			
Distribution	Parameters		
exponential	$\mu = 119.438$		
pareto	$\kappa = -0.0581037$	$\sigma = 2.30148$	$\vartheta = 0$
Lognormal	$\mu = 0.690605$	$\sigma = 0.454923$	
Weibull	$a = 2.47537$	$b = 1.90237$	
Day 2			
Distribution	Parameters		
exponential	$\mu = 169.017$		
pareto	$\kappa = -0.0453913$	$\sigma = 2.30502$	$\vartheta = 0$
lognormal	$\mu = 0.695608$	$\sigma = 0.459347$	
Weibull	$a = 2.50059$	$b = 1.82018$	
Day 3			
Distribution	Parameters		
exponential	$\mu = 9.8713$		
pareto	$\kappa = -1.39219$	$\sigma = 4.17658$	$\vartheta = 0$
lognormal	$\mu = 0.689834$	$\sigma = 0.444456$	
Weibull	$a = 2.44599$	$b = 3.07779$	
Day 4			
Distribution	Parameters		
exponential	$\mu = 175.187$		
pareto	$\kappa = -0.00858352$	$\sigma = 2.45817$	$\vartheta = 0$
lognormal	$\mu = 0.724876$	$\sigma = 0.523742$	
Weibull	$a = 2.71646$	$b = 1.42344$	

Table 9: Concentrated Table of Distribution Parameters

4.2 Distributions Comparison

Scenarios with changing values for the distance and the time threshold have already been test and the results are shown through graphical representation. There are histograms for contact duration and inter – contact times, as well as ECDF and CCDF curves for their values and the fitting distribution. Graphs show a rough estimate of which fitting distribution matches better the data. In order to exist an explicit match, the mathematical expression of the goodness of fit of each distribution should be measured. The goodness of fit is a measure that reveals how well a distribution fits to the data. It is a number with values between 0 and 1.

Chi – square goodness – of – fit test was used against the distributions. The chi-square goodness-of-fit test determines if a data sample comes from a specified probability distribution, with parameters estimated from the data**.** The alternative hypothesis is that the data does not come from such a distribution. The result h is 1 if the test rejects the null hypothesis at the 5% significance level, and 0 otherwise. p - value of the test, returned as a scalar value in the range [0,1]. p is the probability of observing a test statistic as extreme as, or more extreme than, the observed value under the null hypothesis. Small values of p cast doubt on the validity of the null hypothesis.

Tests results are show at Table 10, below. Generally the more entries exist in the data set, the more accurate is the distribution fit. Results from day 3 confirm the rule. Also, larger distance and time threshold give better results. A final conclusion that may be drown from the mathematical results is the lognormal and Weibull distribution are better fits for inter – contact times, and pareto distribution is a better fit for contact duration characteristic. It is important to mention, that all the distribution are close fits and the designer should make the final decision to use one of them.

Baseline Scenario						
ICTvec					**Contact duration**	
Day 1	exponential	pareto	lognormal	Weibull	exponential	pareto
Prob.	0	5.4395e-157	0	0	0	0
Day 2	exponential	pareto	lognormal	Weibull	exponential	pareto
Prob.	0	1.5411e-194	0	0	0	0
Day 3	exponential	pareto	lognormal	Weibull	exponential	pareto
Prob.	2.6442e-78	1.1406e-06	4.4796e-117	2.3562e-41	1.5786e-277	0
Day 4	exponential	pareto	lognormal	Weibull	exponential	pareto
Prob.	0	1.3902e-138	0	0	1.1244e-130	NaN

Scenario 1 *dist =0.5* and *thresh_t=1*						
ICTvec					**Contact duration**	
Day 1	exponential	pareto	lognormal	Weibull	exponential	pareto
Prob.	5.136e-315	3.7683e-67	0	3.6227e-215	0	0
Day 2	exponential	pareto	lognormal	Weibull	exponential	pareto
Prob.	0	6.1317e-38	0	2.5059e-310	0	0
Day 3	exponential	pareto	lognormal	Weibull	exponential	pareto
Prob.	3.1457e-26	1.9853e-03	2.0106e-37	4.0606e-13	3.7391e-120	0
Day 4	exponential	pareto	lognormal	Weibull	exponential	pareto
Prob.	4.5363e-200	6.7913e-56	0	5.1111e-152	1.6291e-305	1.6914e-287
Scenario 2 *dist=2* and *thresh_t=1*						
ICTvec					**Contact duration**	
Day 1	exponential	pareto	lognormal	Weibull	exponential	pareto
Prob.	0	6.7330e-292	0	0	0	0
Day 2	exponential	pareto	lognormal	Weibull	exponential	pareto
Prob.	0	3.0851e-176	0	0	0	0
Day 3	exponential	pareto	lognormal	Weibull	exponential	pareto
Prob.	1.5984e-227	1.9474e-41	0	1.5268e-162	0	0
Day 4	exponential	pareto	lognormal	Weibull	exponential	pareto
Prob.	0	1.3824e-307	0	0	0	NaN
Scenario 3 *dist = 1* and *thresh_t=2*						
ICTvec					Contact duration	
Day 1	exponential	pareto	lognormal	Weibull	exponential	pareto
Prob.	0	5.1387e-309	0	0	0	0
Day 2	exponential	pareto	lognormal	Weibull	exponential	pareto
Prob.	0	1.6246e-189	0	0	0	0
Day 3	exponential	pareto	lognormal	Weibull	exponential	pareto
Prob.	1.3027e-200	5.5754e-36	1.1754e-275	1.924e-129	0	0
Day 4	exponential	pareto	lognormal	Weibull	exponential	pareto
Prob.	0	6.0944e-267	0	0	0	0
Scenario 4 *dist = 0.5* and *thresh_t=2*						
ICTvec					Contact duration	
Day 1	exponential	pareto	lognormal	Weibull	exponential	pareto
Prob.	0	5.6867e-168	0	0	0	0
Day 2	exponential	pareto	lognormal	Weibull	exponential	pareto
Prob.	0	1.5979e-77	0	0	0	0
Day 3	exponential	pareto	lognormal	Weibull	exponential	pareto
Prob.	3.4908e-55	4.5724e-10	2.5247e-80	3.3282e-34	0	0
Day 4	exponential	pareto	lognormal	Weibull	exponential	pareto
Prob.	0	6.3850e-110	0	8.2590e-284	0	0
Scenario 5 *dist=2* and *thresh_t=2*						
ICTvec					Contact duration	
Day 1	exponential	pareto	lognormal	Weibull	exponential	pareto
Prob.	0	0	0	0	0	0
Day 2	exponential	pareto	lognormal	Weibull	exponential	pareto
Prob.	0	0	0	0	0	0
Day 3	exponential	pareto	lognormal	Weibull	exponential	pareto
Prob.	0	3.2164e-104	0	0	0	0
Day 4	exponential	pareto	lognormal	Weibull	exponential	pareto
Prob.	0	0	0	0	0	0

Table 10: Distributions mathematical comprison

5. Epilogue

The dissertation was a research undertaken during the winter semester of 2015, of a Master of Science degree in Information and Communications Technology Systems, for the International Hellenic University. The goal of this dissertation was the study of the trajectories and the places of the vessels that move across the Mediterranean Sea, for the study of an opportunistic network. Vessels transmit AIS data while travelling the sea and the goal of the dissertation is to study if during the transmissions, the vessels can be transformed to contact traces for opportunistic networks. Characteristics and metrics taken through a simulator developed in Matlab and they are studied against the expected theoretical values of an opportunistic network.

5.1 Conclusions

Maritime traffic across the Mediterranean Sea consists one of the major vessel traffic in the world. Vessels can be either small, like rowing boats or huge cruisers. Passenger capacity is also varied, from few passengers to thousands. All vessels travel in regions that the cellular network coverage is problematic, due to the distance from the land. In the submitted research, the addressed problem is how to create a network exploiting the vessels' AIS transmissions.

The dissertation studies the data transmitted from vessels in the Mediterranean Sea. From this data, through an application developed in Matlab, network characteristics and metrics are derived. The first step of the research was the data acquisition and its pre – processing in order to be in a form accessible through Matlab. The second step was the simulator's development in order to measure the network created from the vessels. The application finds the contacts between two vessels, counts the number of contacts each ship has, as long the data set last. It also measures the duration of every contact and inter – contact times from the vessels involved in the network. The final also finds the continuous contacts and differentiates them from the new ones.

Different scenarios are imposed for sensitivity analysis. There are two variables the time threshold and the distance that change during the scenarios. The data sets are separated in four different days and each day is teste for each scenario. For each one of the scenarios and for both parameters the empirical cumulative probability for inter – contact times and contact duration is computed, as well as the complementary cumulative probability. The probability curves are tested against different distributions, known to be compliant with the opportunistic networks. These distributions are called fitting distribution and there are parameters derived from the data, that show how much each data set in each scenario approximates the network's performance. The used distributions were: exponential, pareto, lognormal and Weibull.

Analysing the scenarios results for all the parameters, the best suiting parameters are found in order an opportunistic network to be created. The conclusion that can be drown is a network created from the vessels moving in the Mediterranean Sea is possible and approximates the ideal opportunistic network with great accuracy. Even though, the collected data is scarce and does not cover a whole day, but few hours in each day, the conclusion is still valid.

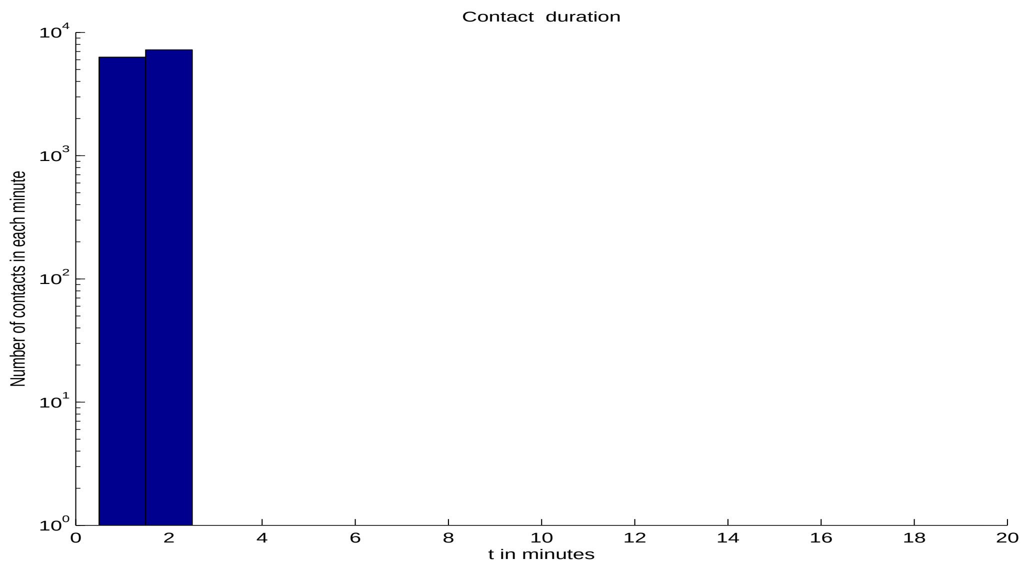
5.2 Further Work

Further work is implied from the research for the dissertation. The presented research covers only the characterization of a marine opportunistic network. The following steps should be towards the direction of the simulation of network, using the already collected data. The simulation is necessary for testing network protocols and finding which is the most suitable to implement. Though network protocols are well documented, the uncharacteristic nature of the means of implementation, namely vessels travelling in the sea, requires extensive simulation and observation. Finally, after the simulation, the last step is the final implementation on vessels. While the implementation takes place, observations are made and the final documentation is presented.

Appendix

In the appendix, results of data sets of different days taken from the scenarios ran in the application are selectively presented.

Scenario 1 – Day 1



FigureA.1: Contact duration - scenario 1 day 1

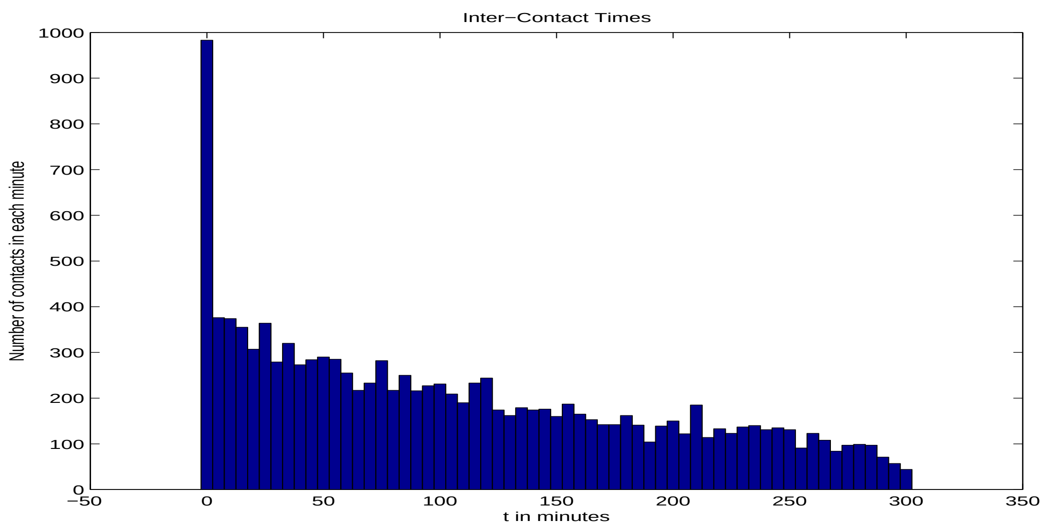


Figure A.2: Histogram ICTvec - scenario 1 -day1

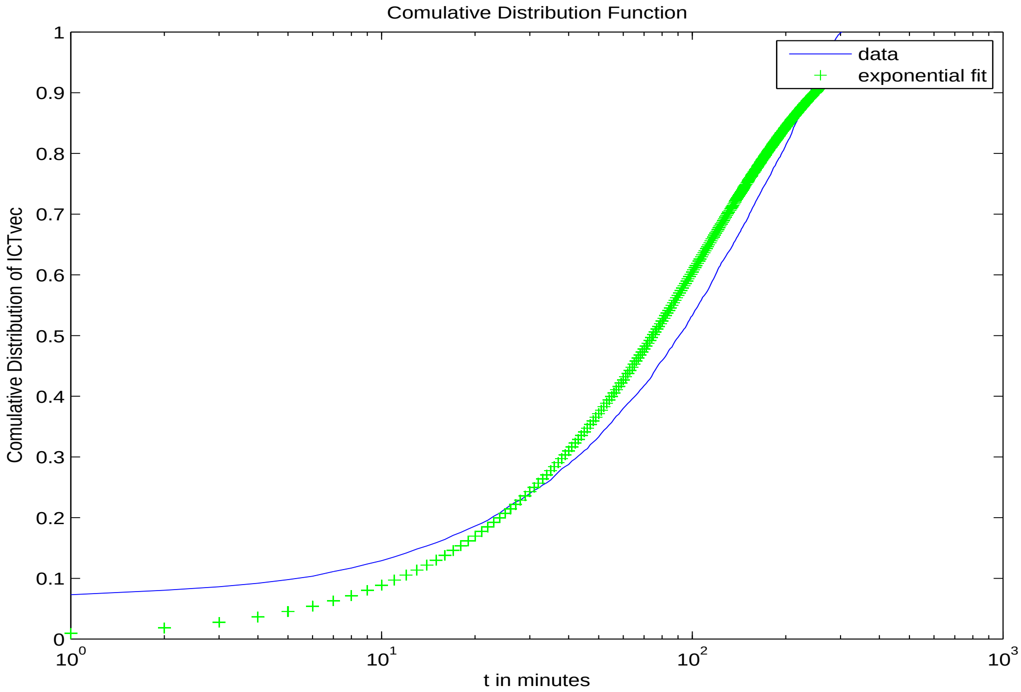


Figure A.3: CDF ICTvec - scenario 1 -day1

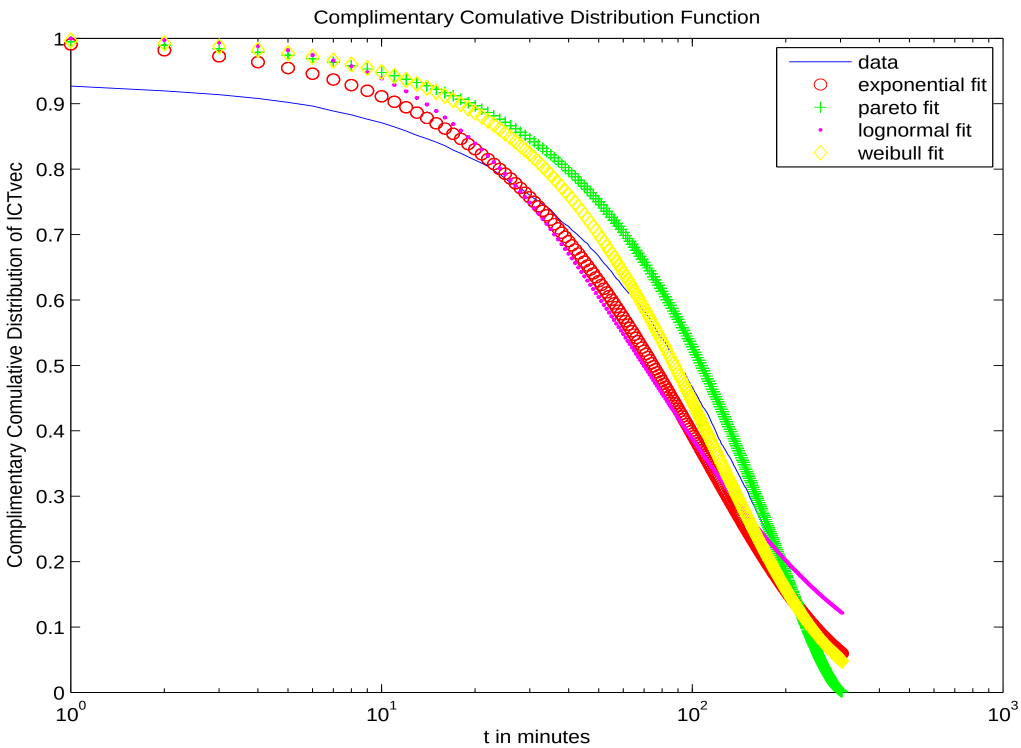


Figure A.4: CCDF ICTvec -scenario 1 -day 1

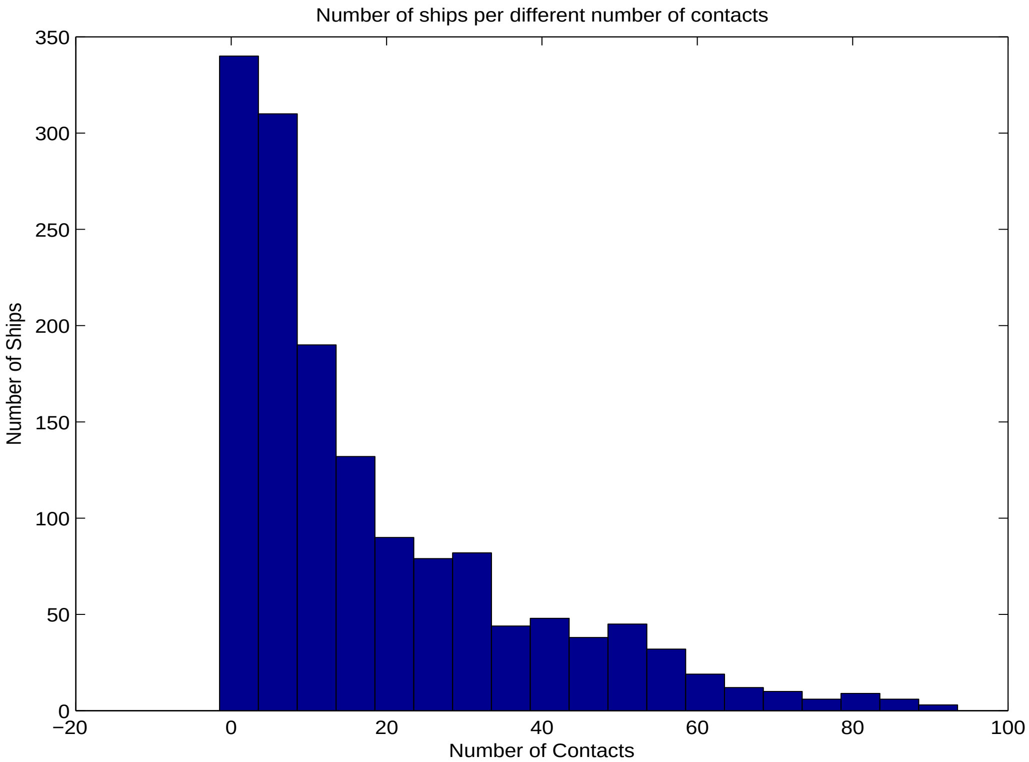


Figure A.5: Number of ships per contacts - scenario 1 -day1

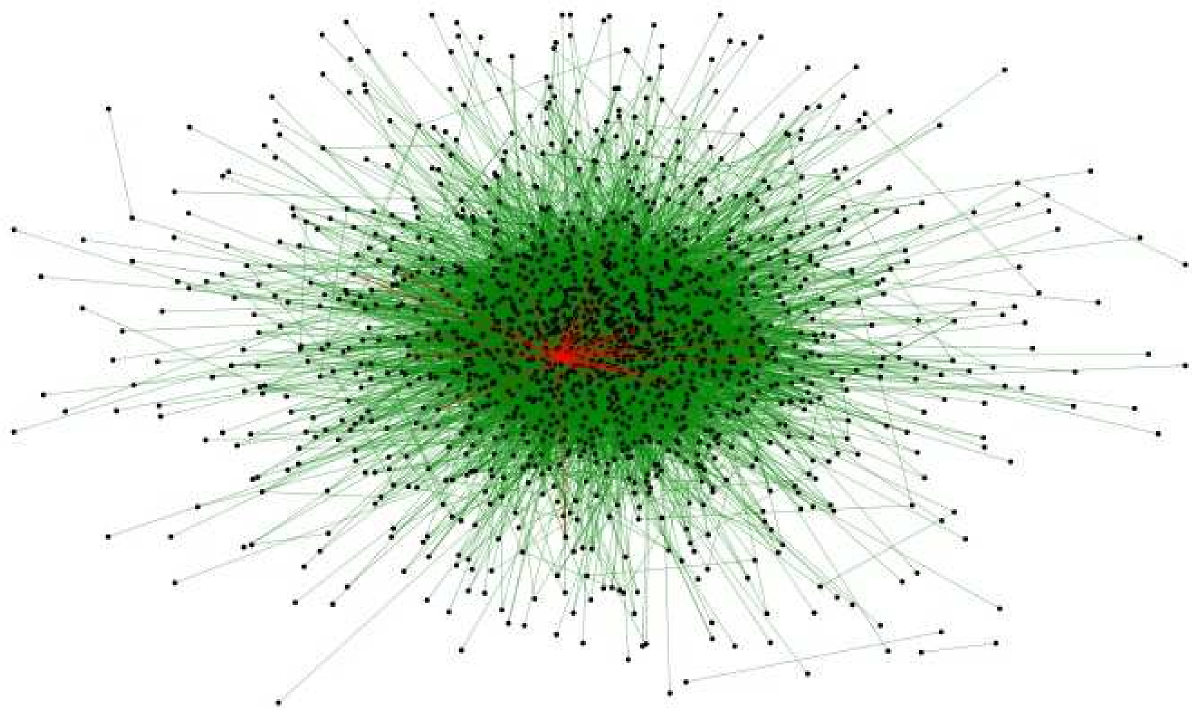


Figure A.6: Contact graph - scenario 1 - day 1

Scenario 5 – Day 3

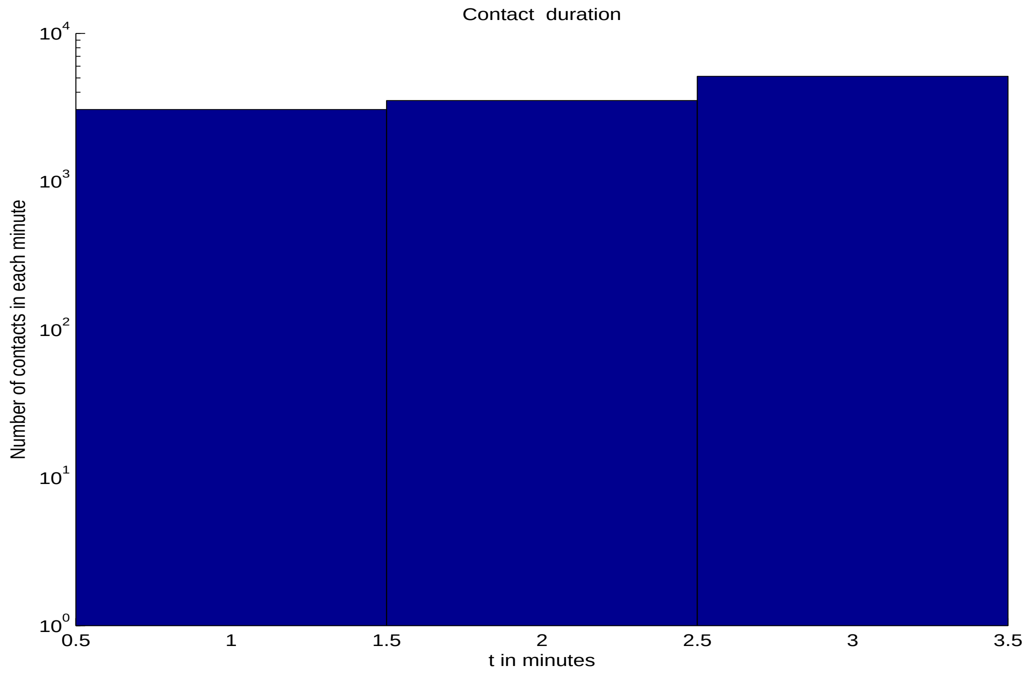


Figure A.7: Contact duration - scenario 5 - day 3

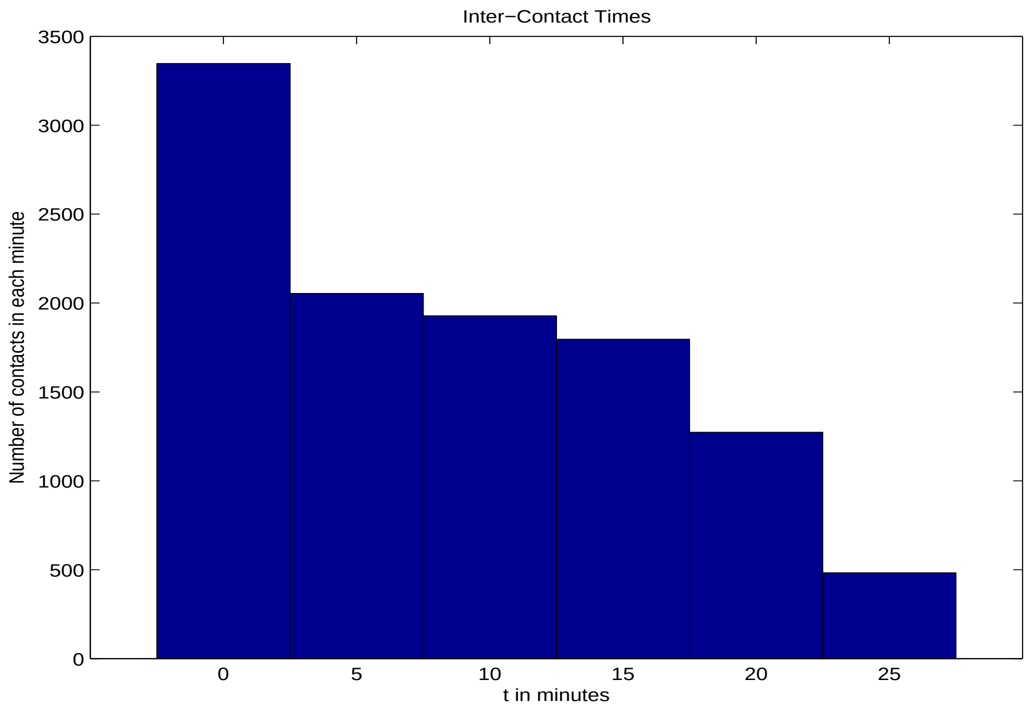


Figure A.8: Histogram ICTvec - scenario 5 - day 3

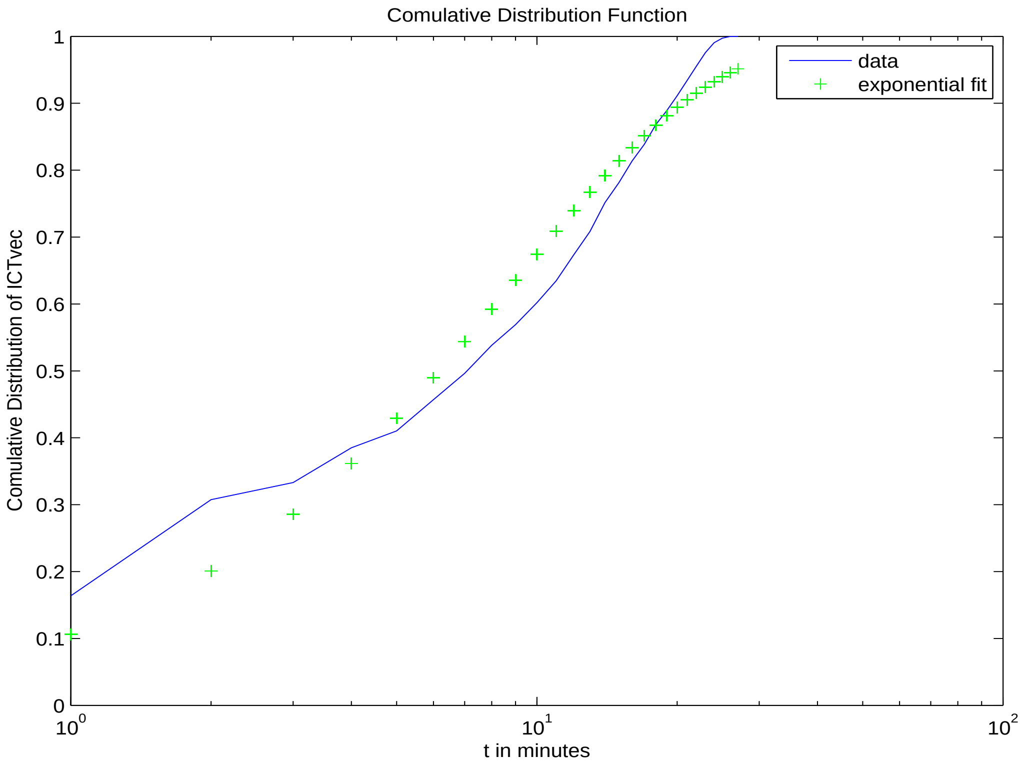


Figure A.9: CDF ICTvec - scenario 5 -day 3

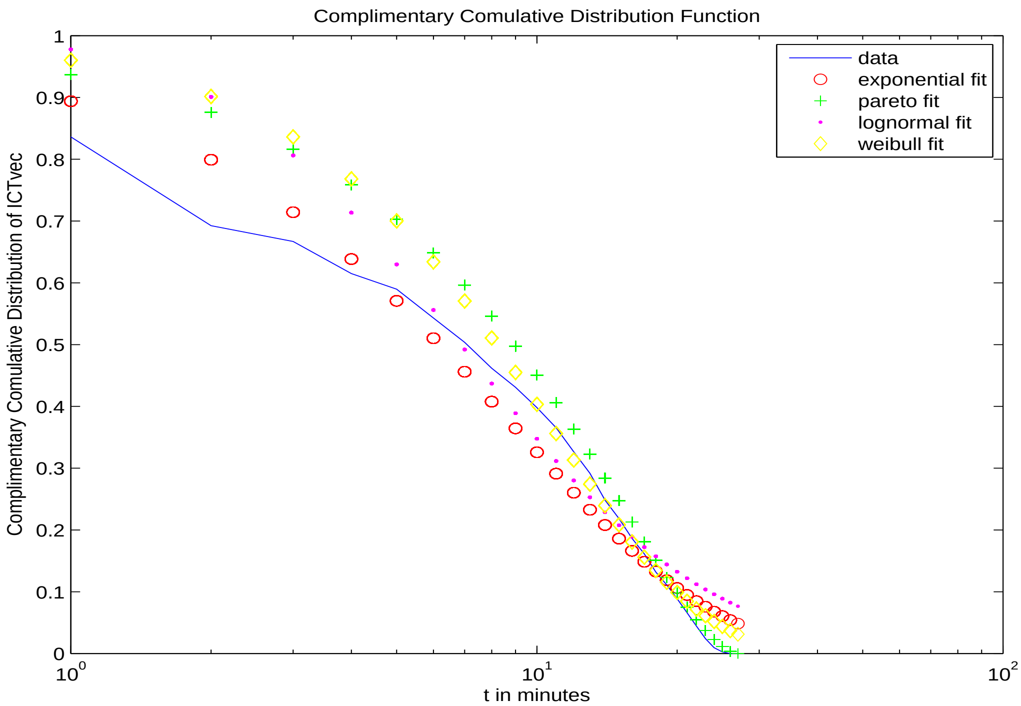


Figure A.10: CCDF ICTvec - scenario 5 - day 3

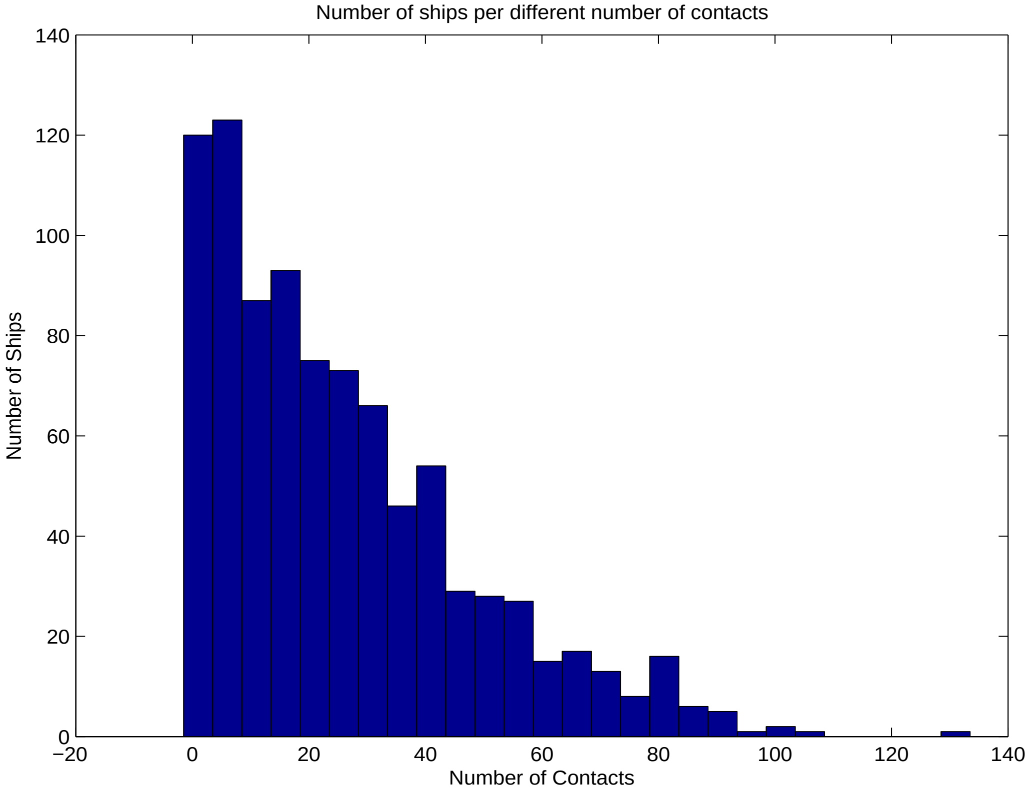


Figure A.11: Number of ships per contact - scenario 5 -day 3

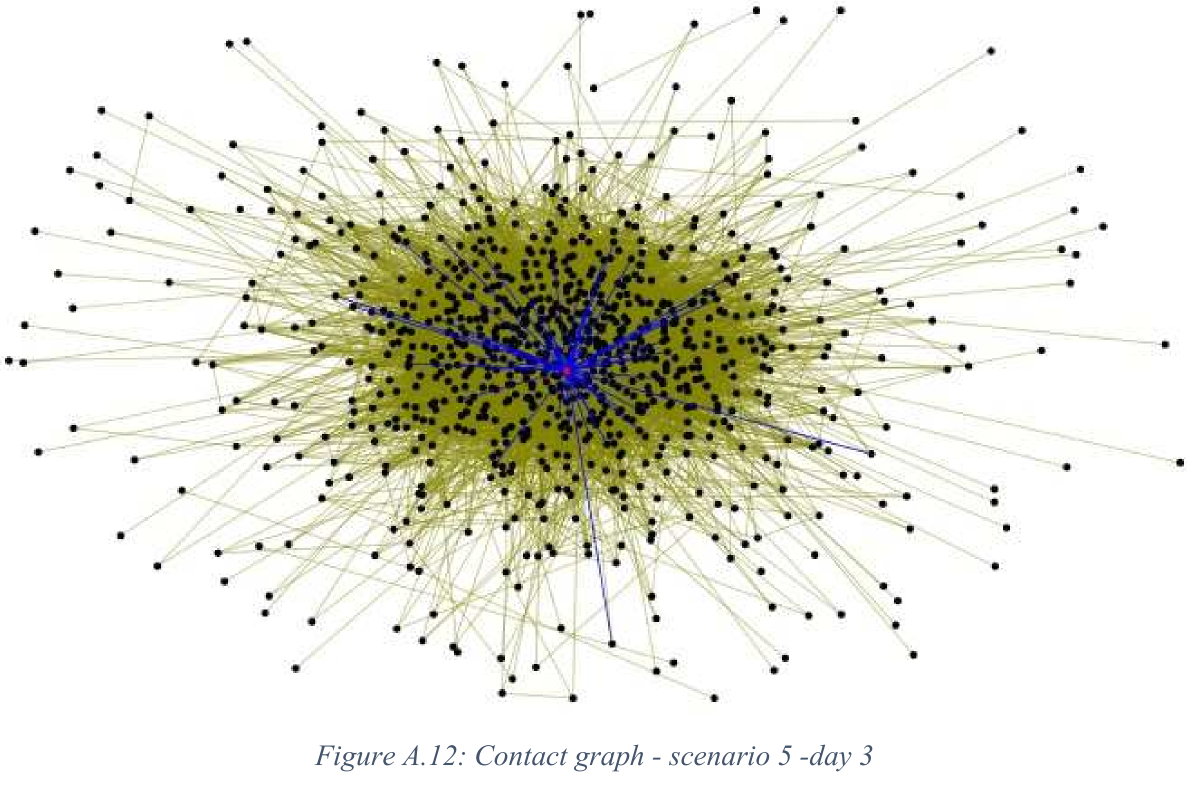


Figure A.12: Contact graph - scenario 5 -day 3

References

Bernardin, P., Yee, M.F. & Ellis, T., 1998. Cell radius inaccuracy: a new measure of coverage reliability. *IEEE Transactions on Vehicular Technology*, 47(4), pp.1215–1226.

Boldrini, C., Conti, M. & Passarella, A., 2011. From pareto inter-contact times to residuals. *IEEE Communications Letters*, 15(11), pp.1256–1258.

Buchsbaum, A.L. et al., 2006. Rectangular Layouts and Contact Graphs. , V(October), p.28. Available at: http://arxiv.org/abs/cs/0611107.

Calcev, G. & Bonta, J., 2009. OFDMA Cellular Networks with Opportunistic Two-Hop Relays. *EURASIP Journal on Wireless Communications and Networking*, 2009(1), p.702659. Available at: http://jwcn.eurasipjournals.com/content/2009/1/702659 [Accessed August 25, 2015].

Chaintreau, a et al., 2006. Impact of Human Mobility on the Performance of Opportunistic Forwarding Algorithms. *Infocom*, 6(6), pp.606–620.

Groenevelt, R., Nain, P. & Koole, G., 2005. The message delay in mobile ad hoc networks. *Performance Evaluation*, 62(1-4), pp.210–228.

Hay, D. & Giaccone, P., 2009. Optimal routing and scheduling for deterministic delay tolerant networks. *2009 Sixth International Conference on Wireless On-Demand Network Systems and Services*.

Huang, C.M., Lan, K.C. & Tsai, C.Z., 2008. A survey of opportunistic networks. *Proceedings - International Conference on Advanced Information Networking and Applications, AINA*, pp.1672–1677.

Jain, S., Fall, K. & Patra, R. 2004. Routing in a delay tolerant network. *SIGCOMM Comput. Commun.* Rev. 34, 4.pp.145-158

Karagiannis, T., Le Boudec, J.-Y. & Vojnović, M., 2007. Power law and exponential decay of inter contact times between mobile devices. *Proceedings of the 13th annual ACM international conference on Mobile computing and networking MobiCom 07*, (March), pp.183–194. Available at: http://dl.acm.org/citation.cfm?id=1287853.1287875.

Karaliopoulos, M. & Rohner, C. 2012. Trace-based performance analysis of opportunistic forwarding under imperfect node cooperation. *Proceedings of the 31st Annual IEEE International Conference on Computer Communications (IEEE INFOCOM 2012).*

Kostadinova, R. & Adam, C., Performance Analysis of the Epidemic Algorithms. , pp.1–5.

Kotsiantis, S.B., Kanellopoulos, D. & Pintelas, P.E., 2006. Data Preprocessing for Supervised Leaning. , 1(2), pp.111–117.

Li, N., Hou, J.C. & Sha, L., 2005. Design and analysis of an MST-based topology control algorithm. *IEEE Transactions on Wireless Communications*, 4(3), pp.1195–1206. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1427709.

Lindgren, A., Doria, A. & Schelén, O., 2003. Probabilistic routing in intermittently connected networks. *ACM SIGMOBILE Mobile Computing and Communications Review*, 7(3), p.19.

Mishra, N. et al., 2007. Clustering Social Networks. *Lecture Notes in Computer Science*, 4863, pp.1–12. Available at: http://www.springerlink.com/index/lrr012003t3757x5.pdf.

Passarella, A. & Conti, M., 2011. Characterising aggregate inter-contact times in heterogeneous opportunistic networks. *Networking 2011*, pp.301–313.

Passarella, A. & Conti, M., 2013. Analysis of individual pair and aggregate intercontact times in heterogeneous opportunistic networks. *IEEE Transactions on Mobile Computing*, 12(12), pp.2483–2495.

Rahm, E., Data Cleaning : Problems and Current Approaches.

Ramanathan, R. et al., 2007. Prioritized epidemic routing for opportunistic networks. *Proceedings of the 1st international MobiSys workshop on Mobile opportunistic networking - MobiOpp '07*, pp.62–66. Available at: http://portal.acm.org/citation.cfm?doid=1247694.1247707.

Spyropoulos, T., Psounis, K. & Raghavendra, C.S., 2005. Spray and Wait : An Efficient Routing Scheme for. *Direct*, pp.252–259. Available at: http://portal.acm.org/citation.cfm?id=1080139.1080143.

Standard, I., 2004. INTERNATIONAL STANDARD ISO. , 2004.

Viterbi, A.J., Viterbi, A.M., Gilhousen, K.S. & Zehavi, E., 1994. Soft Handoff Extends CDMA Cell Coverage and Increases Reverse Link Capacity. IEEE Journal on Selected Areas in Communications, 12(8), pp.1281–1288.

Wu, S., 2013. A review on coarse warranty data and analysis. *Reliability Engineering and System Safety*, 114, pp.1–11. Available at: http://dx.doi.org/10.1016/j.ress.2012.12.021.

Zhang, X. et al., 2007. Performance modelling of epidemic routing. *Computer Networks*, 51(10), pp.2867–2891.