NATIONAL CENTER FOR FREIGHT & INFRASTRUCTURE RESEARCH & EDUCATION

CFIRE

Mining Automatic Identification Systems (AIS) Data for Improved Vessel Trip Analysis Capabilities

CFIRE RI-4
July 2016

National Center for Freight & Infrastructure Research & Education
Department of Civil and Environmental Engineering
College of Engineering
University of Wisconsin–Madison

Authors:
James P. Dobbins, Ph.D., Vanderbilt University
Samir Dahr, University of Toledo

Principal Investigator:
Peter Lindquist and Richard Martinko, University of Toledo
Jimmy Dobbins, Vanderbilt University

This page intentionally left blank.

Technical Report Documentation

1. Report No. CFIRE RI-04	2. Government Accession No.	3. Recipient's Catalog No. **CFDA 20.701**
4. Title and Subtitle Mining Automatic Identification Systems (AIS) Data for Improved Vessel Trip Analysis Capabilities		5. Report Date **July 2016** 6. Performing Organization Code
7. Author/s James P. Dobbins, Ph.D., Vanderbilt University and Samir Dahr, University of Toledo		8. Performing Organization Report No. **CFIRE RI-04**
9. Performing Organization Name and Address National Center for Freight and Infrastructure Research and Education (CFIRE) University of Wisconsin-Madison 1415 Engineering Drive, 2205 EH Madison, WI 53706		10. Work Unit No. (TRAIS) 11. Contract or Grant No. T002688
12. Sponsoring Organization Name and Address The Office of the Secretary of Transportation – Research US Department of Transportation 1200 New Jersey Avenue, SE Washington, DC 20590		13. Type of Report and Period Covered **Final Report [7/1/2012–5/30/2016]** 14. Sponsoring Agency Code
15. Supplementary Notes Project completed by CFIRE with support from the Wisconsin Department of Transportation.		
16. Abstract A methodology for processing AIS data from multiple sites in near real-time was developed as well as a capability to support ad-hoc data queries. Such analyses can identify high-risk locations, generate better travel time estimates, detect vessel arrivals, identify key traffic areas for investment and enhancement, and in general lead to a better understanding of vessel traffic within a given area. Key research tasks included installing additional shore-based AIS antennas at strategically selected Great Lakes and inland waterway locations, and developing the server and database management system that can quickly receive and process the data from multiple sites. Benefits of AIS automation include increased levels of information regarding boat location and trajectory supporting safer waterway operations, more efficient lock operation and lead times, and increased shipping efficiencies as waterway traffic is better understood and managed.		
17. Key Words Automatic Identification Systems, Freight Transportation, Marine Transportation, Best Practices, Global Positioning System, Information Dissemination, Inland Waterways; Locks (Waterways), Statistical Analysis	18. Distribution Statement **No restrictions. This report is available through the Transportation Research Information Services of the National Transportation Library.**	

19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price
Unclassified	**Unclassified**	121	**-0-**

Form DOT F 1700.7 (8-72) Reproduction of form and completed page is authorized.

Report Contents

Inland Rivers Component (13 pages)

Great Lakes Component (107 pages)

Mining Automatic Identification Systems (AIS) Data for Improved Vessel Trip Analysis Capabilities

Inland Rivers Component

Final Report

James P. Dobbins, Ph.D.
Research Associate Professor
Vanderbilt University

June 6, 2013

Introduction and Background

Automatic identification systems (AIS) provide communication between vessels and assist vessel traffic control functions in congested ports, locks and waterways. Vessels digitally broadcast information, including position, call sign, vessel name, course, speed, and navigation status. Publicly available waterway trip data is aggregated to protect business confidentiality of shippers. However, the level of aggregation renders the data unusable for a variety of applications, including waterway congestion, port demand modeling and risk analysis. CFIRE researchers at the University of Toledo (UT) and Vanderbilt University (VU) have independently conducted research into collecting and analyzing AIS data on the Great Lakes and inland waterways, respectively. The work completed by these researchers has demonstrated the feasibility of using shore-based AIS receivers to acquire, archive and analyze data on vessel movements. The integration of relational database management systems (RDBMS), geographic information systems (GIS) and custom software routines is a powerful combination of technology that is capable of leveraging large amounts of AIS data.

This project was designed to develop a methodology for processing AIS data from multiple sites in near real-time as well as develop a capability to support ad-hoc data query. Such analyses can identify high-risk and high-traffic locations, generate better travel time estimates, detect vessel arrivals, identify key traffic areas for investment and enhancement, enable terminal operators to better manage their operations, and in general lead to a better understanding of vessel traffic within a given area. Benefits of AIS automation include increased levels of information regarding boat location and trajectory which supports safer waterway operations, more efficient lock operation and lead times, and increased shipping efficiencies as waterway traffic is better understood and managed. Key research tasks included the installation of additional AIS antennas at strategically selected Great Lakes and inland waterway locations, and development of server and database management systems that can quickly receive, process, analyze and publish the data from multiple sites. The research was completed in two components; an inland rivers component (lead by Vanderbilt University) and a Great Lakes component (lead by the University of Toledo). This project report documents the inland rivers component.

In 2011, Vanderbilt University completed a 6-month study of inland AIS data received at an Ingram Barge Company facility in Paducah, Kentucky (Dobbins and Langsdon, 2012). Despite the complex river port configuration (Paducah is located at the meeting of the Cumberland, Ohio, and Tennessee Rivers), researchers successfully built and deployed models to generate useful trip data as well as detect vessel operations such as fleeting, docking and lockages. AIS is not required to be carried above Baton Rouge, LA on the inland waterways. One third of vessels transiting the Paducah area between March and September 2011 were found to have transmitting AIS equipment on board. This was found by comparing the AIS trip information with lock performance data maintained by the U.S. Army Corps of Engineers (USACE) during the same time period, as the case study area was bound by 4 locks: Lock and Dam 52, Smithland Lock and Dam, Kentucky Lock and Barkley Lock. The Paducah AIS receiver's 17-mile range easily captured vessel events at all of the aforementioned locks. Data manipulation of this project took place once the archive of data was completed as the goal was to see if archived AIS data could be manipulated to generate reliable waterway trip information.

This work led to an additional research project in 2012 where researchers' goal was to develop software routines to process the data as it arrives from the receiver in Paducah as well as an additional receiver in Reserve, LA. The additional receiver location was chosen due to mandatory AIS carriage in that area, as well as the mix of deepwater and inland marine traffic. This project was successfully completed in early 2013.

Research Objectives

The accomplishments of the earlier projects related to AIS data analysis and manipulation support the objectives of this project, which are listed and briefly described below.

1. *Development of a normalized relational database design* – a normalized design will be generated to efficiently store the processed AIS messages. This design will be converted to a data warehouse during the project to efficiently archive the raw data and enable swift queries of the large data repository. The database platform will be Microsoft SQL Server 2012.
2. *Develop code to process data immediately after being received from the AIS receivers in Paducah and Reserve* – as mentioned before, previous analyses was done on an archive of data; this effort will sample the data as it is received and only stored what is necessary to provide useful information, for both vessel traffic visualization purposes and from an analysis standpoint.
3. *Relate real-time AIS data to U.S. Coast Guard static vessel databases* – the AIS vessel data collected in Paducah and Reserve will be joined to the USCG vessel database. Static particulars about the data (including length, beam, horsepower, vessel type, etc.) will then be included in real-time traffic views of the AIS case study areas.
4. *Spatially-enable live data in GIS* – create spatial database views based upon the most recent AIS data received and convert them to a GIS layer. This layer will be based upon routines inside the Microsoft SQL Server 2012 database and Spatial Database Engine (SDE) instance.
5. *Serve the vessel layers using Internet GIS in conjunction with additional real-time vessel traffic (lockage), river and weather data sources* – additional real-time weather, river condition and vessel lockage data will be integrated with the AIS data, enabling a complete picture of the waterways and current conditions. All data sources will be spatially-enabled and served via Internet GIS software (ArcGIS Server). The map service and ad-hoc query services will be password-protected.

Dataflow and Database Design

The AIS receivers in Paducah and Reserve are connected to a PC and continuously log AIS messages in binary format to a text file. A new file is created every 4 hours, and upon completion, these files are uploaded to a Vanderbilt server. As there are 27 different message types (USCG Website, 2013), ranging from position reports to equipment status reports, not all of these are of interest in a vessel traffic application. Custom code was written (in Visual Basic .NET) to filter out message types 1 (position reports), 4 (base station reports) and 5 (static and voyage related data). Type 1 messages contain vessel position information, including latitude/longitude, rate of turn, heading, course, speed, and a time stamp (seconds). Type 4 messages are interspersed throughout the logfiles and contain the time information (month, day, year, hour, minute). Seconds from type 1 messages are appended to the previous type 4 message. Finally, type 5 messages contain the vessel name, ship type, IMO number and call sign. Type 1 and 5 messages are related by a Maritime Mobility Service Identity (MMSI) number, which is a unique number given to all AIS-equipped vessels.

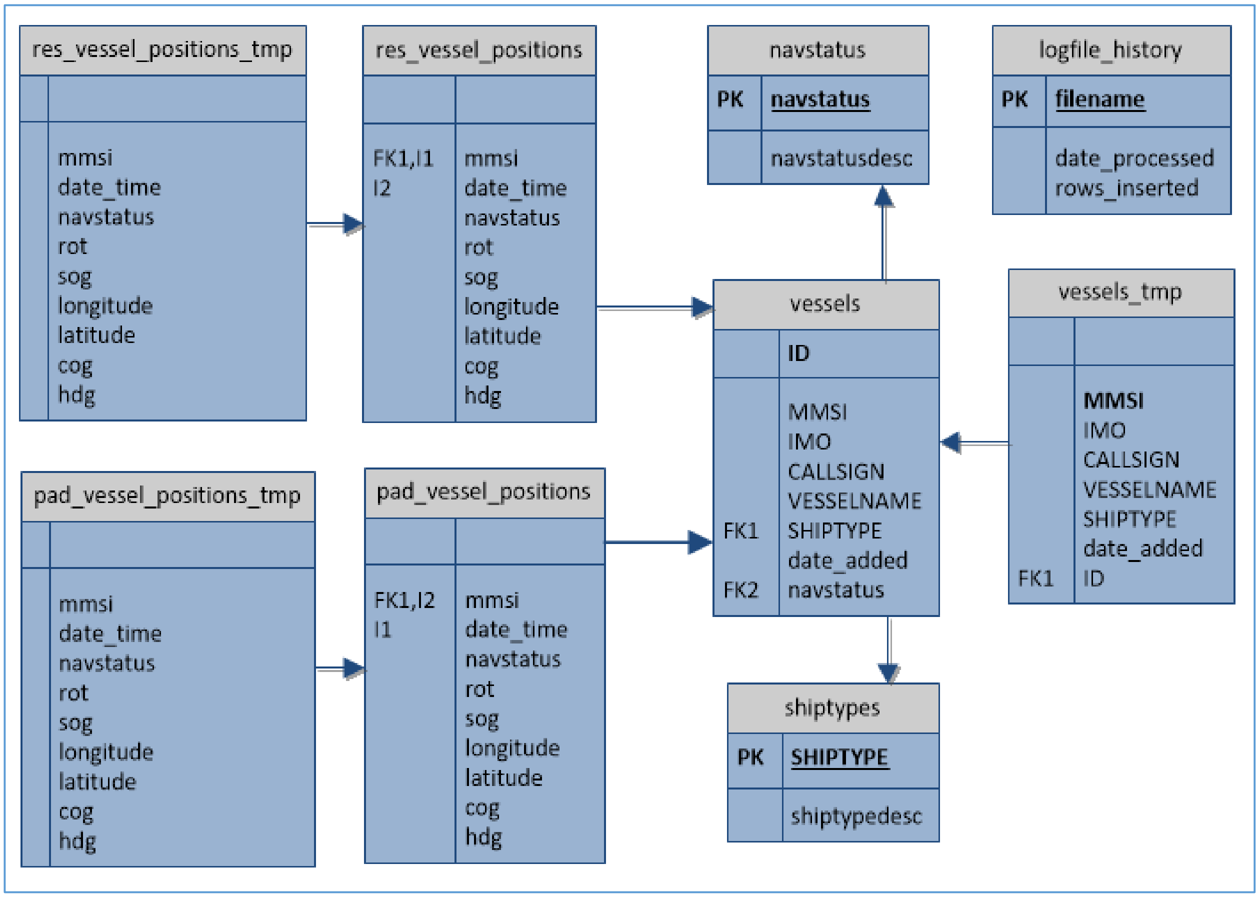


Figure 1. Database Diagram for Inland River Data Center.

Figure 1 shows the database diagram for the inland rivers data center. Note the three tables that are named with a "_tmp" suffix: pad_vessel_positions_tmp, res_vessel_positions_tmp and vessels_tmp. For example, with the vessels table, there are occasions where the vessels are renamed and there are also instances where a typo in the AIS unit will incorrectly name the vessel in the type 5 messages it transmits. Since there is no way of knowing which of these is the case in the event of a new name for a given MMSI number, the executable loads vessel data into a temporary table (vessels_tmp). Following each load, a database trigger automatically inserts new records (only those that do not currently exist) into the vessels database. The trigger also updates the date_added field so it is possible to determine when the vessel was first observed by the AIS receivers. The vessels table has an identity column that automatically assigns a new ID number to each vessel added. This way, we know the most recent name information for a given MMSI number. This was necessary since there are times that a MMSI number will have vessel names slightly different within the same 4 hour log file. Similarly, the executable samples (every ten minutes) and filters records to only load unique vessel positions into the pad_vessel_positions_tmp and res_vessel_positions_tmp tables. However, most vessels transmit their positions at least once every ten seconds. For a given minute, there may be 7 or more positions. A database trigger only inserts the most recent position (maximum second value) for each vessel in each sampled minute (for the following minutes: 00, 10, 20, 30, 40, and 50) into the pad_vessel_positions and res_vessel_positions tables.

The codes in the navstatus and shiptypes tables translate numeric codes into text values (e.g., ship type code of 31 is a towing vessel and a navigation status of 0 represents a vessel under way using her engine). Finally, the logfile_history table is a repository of the file name, time processed, and number of vessel_position records (the count inserted into the "_tmp" tables). The executable sends an email out to the principal investigator of the project every time a log file is processed. This email includes the filename, number of records and the AIS receiver location from which the data was received. Emails every 4 hours are confirmation indicators of normal operations. The database is compacted and backed up nightly, and the principal investigator receives email confirmation of these activities.

Join AIS Data to Static USCG Vessel Database

The original intent of the project was to increase the value of the data by joining it to static USCG vessel databases as well as those maintained by the USACE. This could have made additional attributes available, including year built, length, vessel horsepower, company name, among others. Unfortunately, there are some data issues that limited the number of successful matches between observed vessels and those in the USCG vessel database. The common fields in both tables are IMO number, call sign and vessel name.

The USCG vessel documentation was received from the Marine Information System for Safety and Law Enforcement (MISLE) database and contains 355,873 vessels (USCG MISLE database, 2013). In this database, only 6,627 of the vessels have their IMO number populated, so joining on that field was not an option. Since vessel name by itself is not a credible means of relating the tables, the call sign was used. Call sign is populated for 82,207 vessels and this information is nearly as constant as the IMO number. Successful matches were made on just 913 of the 4,114 observed AIS vessels (in both Reserve and Paducah) as of June 6, 2013. For this reason, the database overhead of indexing the 335,000+ record vessel database table with the AIS data was deemed more detrimental to performance than the benefit gained by including the extra attributes.

Spatial Enabling of the AIS Data

While traditional database views of the Reserve and Paducah data which show the most recent positions of vessels in the area are relatively easy to create, a couple of extra steps are required to spatially enable the data. Unfortunately, due to the nature of the data arrival at the Vanderbilt server (once every 4 hours in bulk), depending on the time of query the database will store position data that is between 15 minutes and 4 hours old at all times. For this reason, it is impossible to have a query return positions in the last 30 minutes, etc. A view was created to return the most recent position for each vessel in each area. Each vessel's latitude and longitude are stored in decimal degrees in the pad_vessel_positions and res_vessel_positions tables. However, these fields need to be converted to a special geography data type known as STPointFromText (MSDN, 2013) before they may be consumed directly by a GIS with a conversion step inbetween. ArcSDE (SDE = spatial database engine) is the middleware used in the project to make such a conversion. A view was created that is recognized by ArcSDE and as a result, the layer is readable by ArcGIS products (including ArcGIS Server, see the next section for the final results discussion) without the need for exporting and importing a layer and manually designating the latitude and longitude fields.

Web Application Development

The final deliverable of this research effort is a password-protected web application that showcases the most recent AIS vessel positions in Paducah, KY and Reserve, LA and has other real-time vessel traffic databases, weather information and river gage information. This combination of data sources in a single application showcases the strengths of GIS as a data management, analysis and most importantly, a visualization tool. This section contains a description of the data sources, the interface and the technology used to bring these data sources together.

In previous research projects completed by Vanderbilt University, the Corpslocks XML service (www.corpslocks.com) was identified as a valuable source of real-time inland marine transportation performance data. For 189 locks operated by the USACE, data is freely available in real-time on vessel lockages through an XML data service. For each lockage, the vessel name, USCG official number, direction (upbound/downbound), number of barges, and key times (arrival, start of lockage, end of lockage) are recorded. Vanderbilt has been continuously mining this data hourly since 2011 with an interruption in the second half of 2012 while the system was transitioned to a new platform. Valuable performance metrics such as queue length, lockage processing times and average tow sizes (number of barges and horsepower figures) are easily calculated and served using ArcSDE and ArcGIS Server 10.1. All of the information is readily available in an easy to use JavaScript web GIS interface (Figure 2).

Figure 2. Screenshot of Internet GIS Site of Lock Performance Data.

The lock layer with performance metrics (upbound and downbound wait time, queue length and tow size) was imported into the AIS application. The locks are very relevant in Paducah (as the area is bounded by 4 locks as mentioned previously) as well as Reserve, where the New Orleans locks (Algiers, Inner Harbor and Harvey Locks) have a significant impact on traffic and delay in the area.

Weather layers are maintained by the National Weather Service (NWS) using an ArcGIS Server interface (NOAA website, 2013). The following NWS weather layers were used in the application:

1. RIDGE Radar image layer – This is a mosaic of NWS radar antennas throughout the continental U.S. The resolution is 1km by 1km, so the images are not as detailed when zooming in to the city level or closer, but the composite image shows precipitation and what parts of the inland waterway network might be affected by severe weather events (RIDGE RADAR Map Server, 2013). Figure 3 shows a screenshot of the RIDGE Radar mosaic overlaid on the inland waterway network and USACE locks layers.

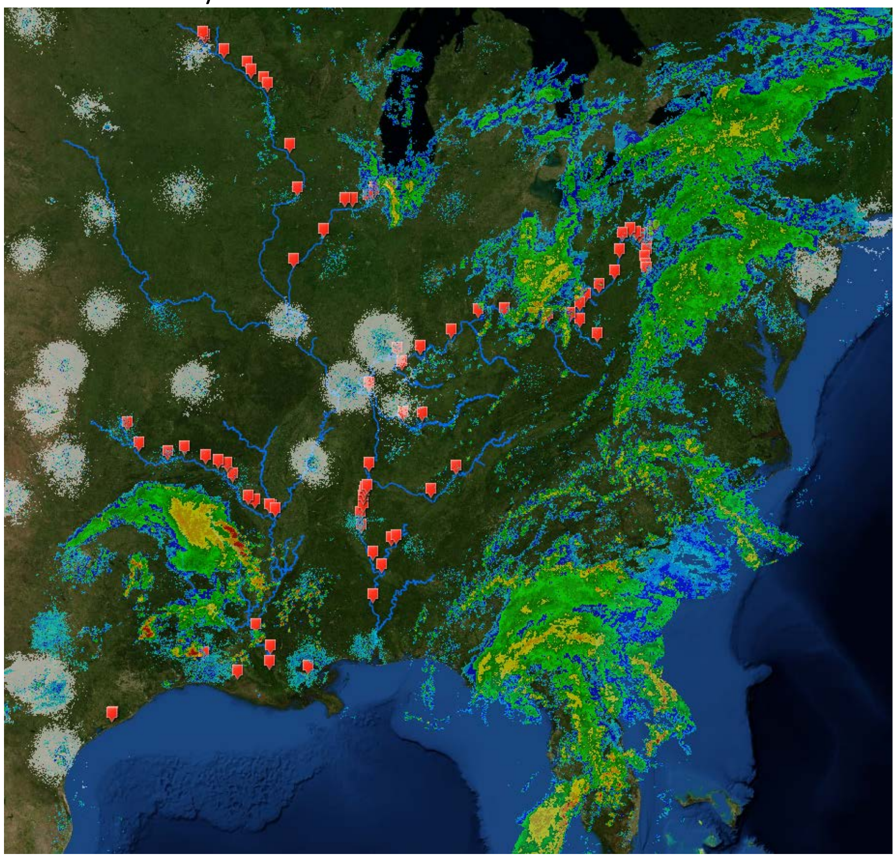


Figure 3. Screenshot of the RIDGE Radar Composite Image.

2. Watches and warnings polygon layer – This layer is a polygon layer that contains all of the active weather watches, alerts and advisories throughout the U.S. Watches are issued when the risk of a hazardous weather or hydrologic event has increased significantly, warnings are issued when hazardous weather or hydrologic events are imminent and/or occurring, and advisories are issued when hazardous weather or hydrologic events are imminent or conditions are present that could lead to situations that might threaten life or property. (NWS Watch Warn Map Service, 2013). Figure 4 shows a screenshot of this layer within the application. The green polygons represent flood advisories along the Lower Mississippi, the polygons in Alabama represent dense fog advisories and the pink polygons in Florida are fire weather watches. Clicking on the polygons brings up information about the condition, as well as a link to the page that contains the official NWS statement regarding the condition.

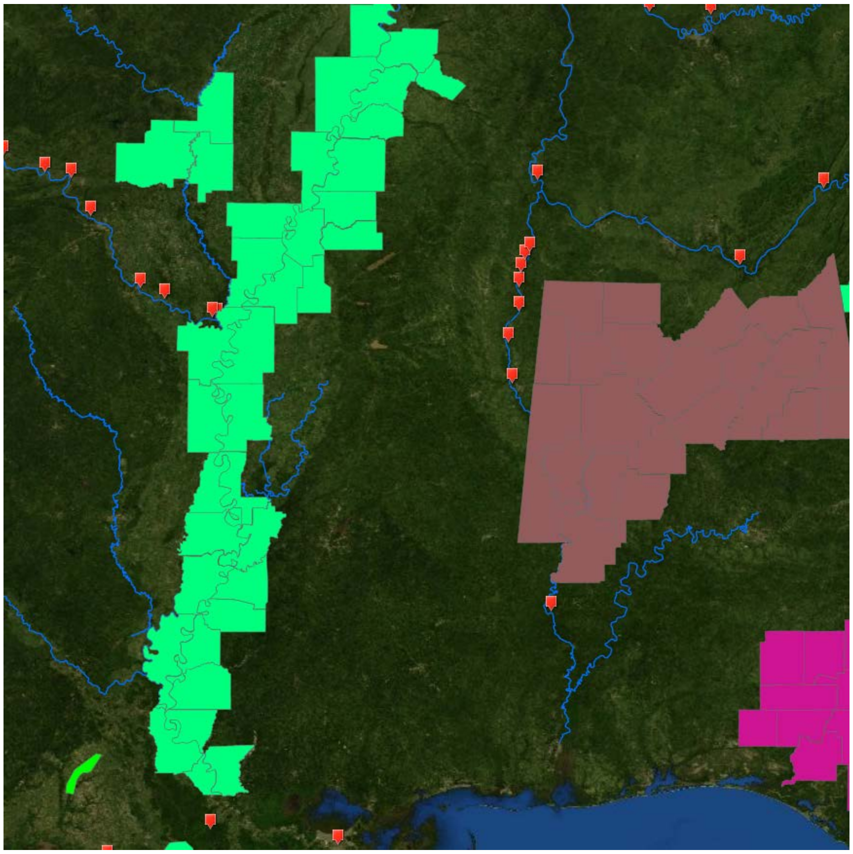


Figure 4. Screenshot of the Watches and Warning Map Service.

3. Observed and forecast river stages – These layers are from the Advanced Hydrologic Prediction Service (AHPS), a suite of tools that predicts weather and hydrologic conditions, such as floods. There are 5,952 gages in the layer, but these are filtered to show only those that currently observe or forecast (within the next 9 days) moderate or major flooding. Clicking on any of the gages will take the user to a page showing a graph of the most recent and forecast readings. (AHPS Map Service, 2013).

Two layers are derivatives of the layers above: the current flooding heat map and the forecast flooding heat map. Each of these are heat maps created on the fly that compute clustering of moderate and major flooding occurring in the U.S.

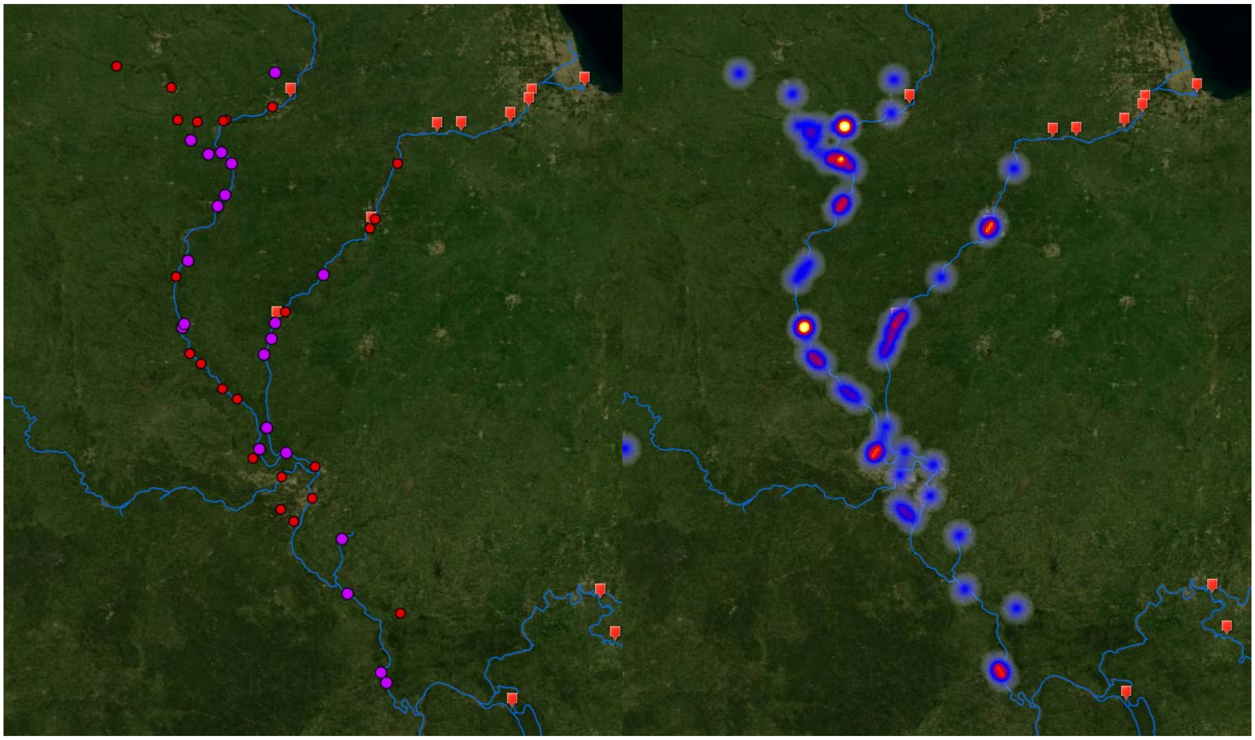


Figure 5. Observed River Stages (left) and Current Flooding Heat Map (right) on June 6, 2013.

Figure 5 shows the heat map calculated on June 6, 2013, a time of flooding along the Upper Mississippi. Each point in the map on the left represents a gage that is currently at moderate or major flood stage, and the heat map derivative map is shown in the picture on the right. This quickly provides a visualization of where flooding is potentially currently impacting vessel traffic and the forecast layer gives an indication of where flooding may impact navigation within the next 9 days.

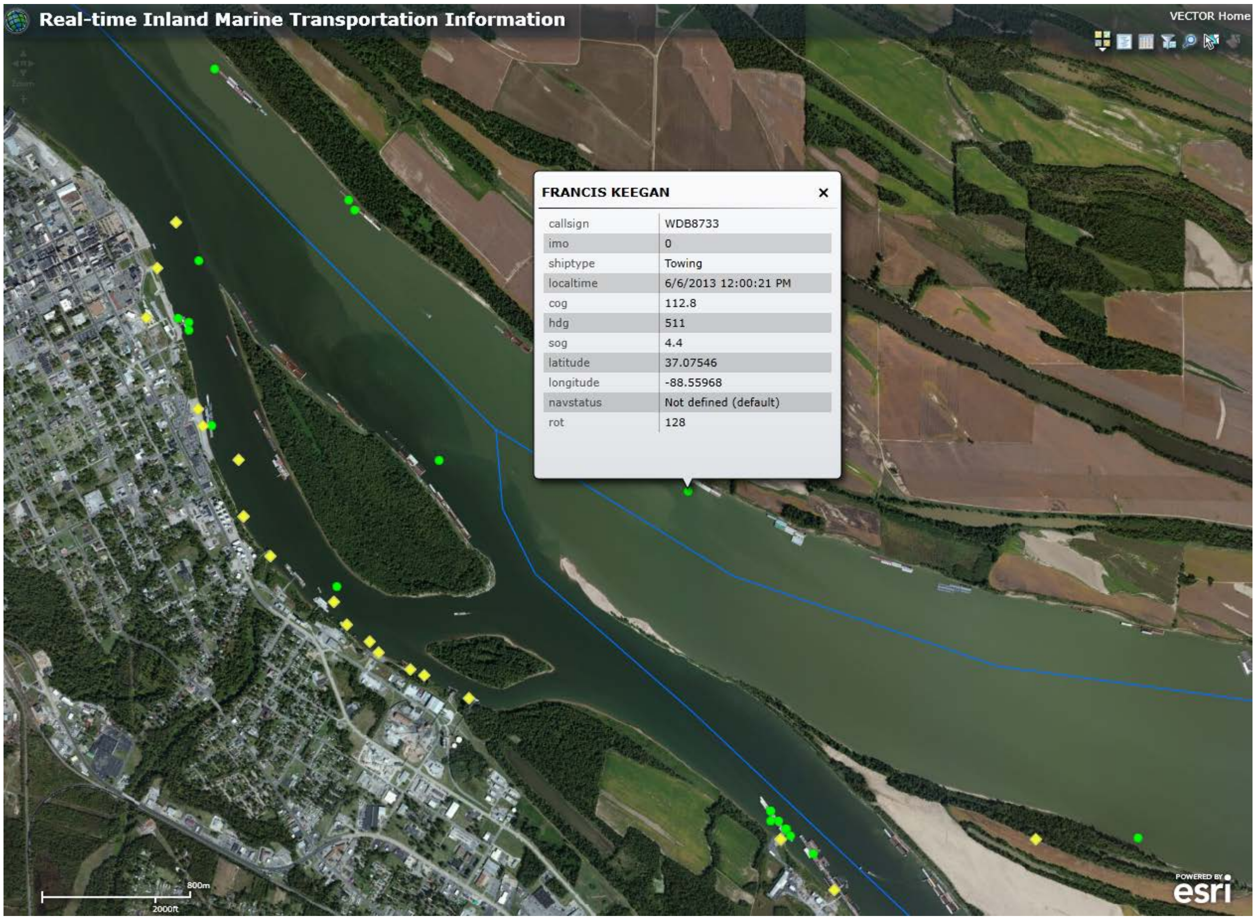


Figure 6. Screenshot of the Paducah, KY area.

A screenshot of the interface in Paducah is shown in Figure 6. The vessels are green circles (the Francis Keegan has been clicked to show information about her status) and the yellow diamonds are from the USACE port series layer. The port series layer contains more than 9,000 water terminals within the United States (USACE Port Series, 2013).

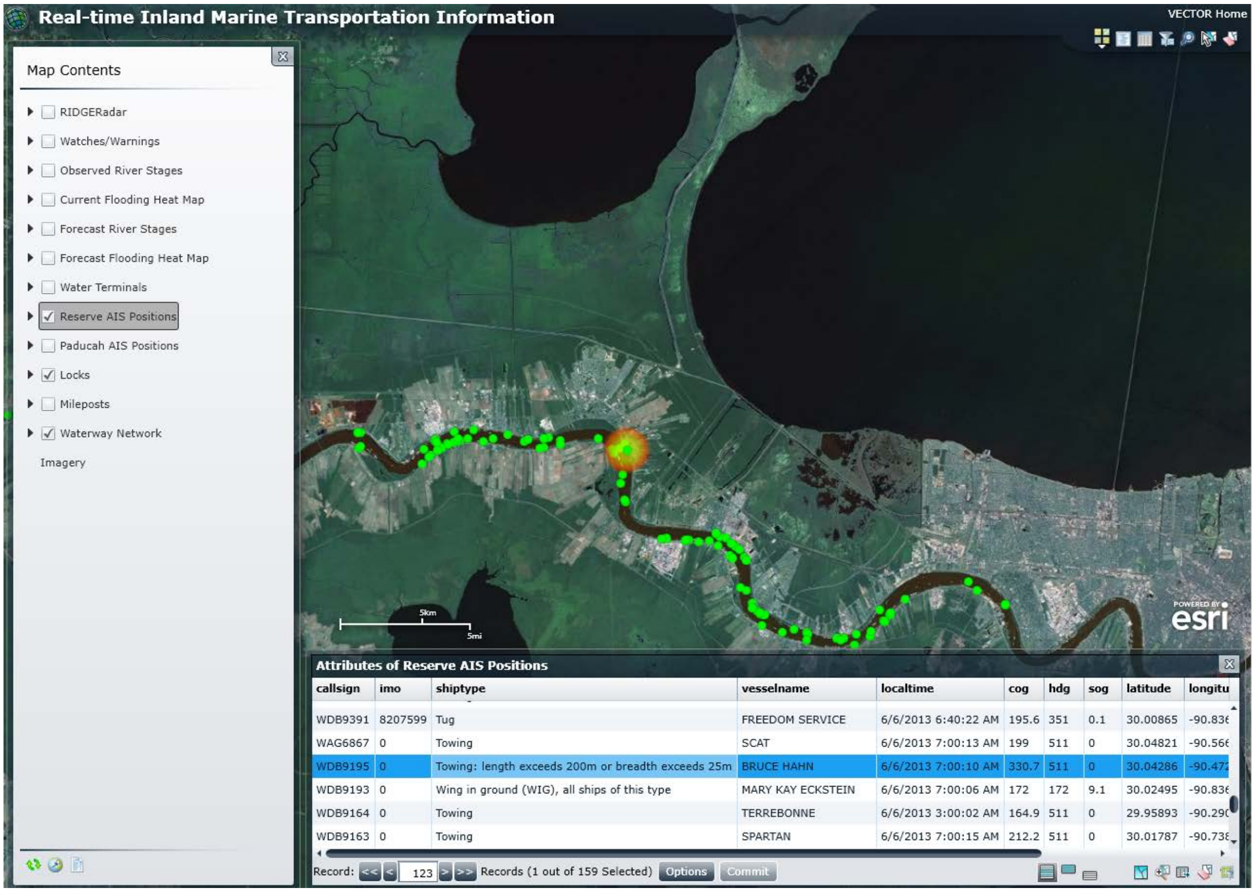


Figure 7. Interface Screenshot in Reserve, LA.

Figure 7 shows some of the features that are present within the interface while zoomed in to the Reserve, LA area. The layers may be toggled on and off on the left side of the frame, a user may also right click to refresh the layer, get information about the layer (metadata), or change the extent of the map to the layer's features. Users may open the attribute table to get a tabular view of the data. While looking at the tabular data, clicking on a row to highlight the record will flash the feature on the screen, as seen in Figure 7 when the Bruce Hahn is selected. The toolbar in the upper right corner of the map gives the user several options:

- Change the underlying base map from satellite imagery to streets, topographic, and shaded relief maps, among others.
- Toggle the map contents window (where layers can be turned on and off).
- Open and close the attribute window, where several feature selection options are available when working with a layer (e.g., zoom to selected, export selected records to CSV format, and show only records from the current map extent).
- Filter the layer and perform ad-hoc querying and selection based upon vessel attributes (e.g., show only towing vessels, show vessels whose status is "Moored," vessel types of "Tanker," etc.).
- Search for content to place in the map. Users may add content from ArcGIS Online and other web sources. Potential layers might include traffic layers, marinas, boat ramps, or other weather-related layers.

- Select features by pointing. Features may be selected by using the pointer to click on them individually, or drag a rectangle around them.
- Finally, users may clear their selection with the toolbar button on the far right of the screen.

It should also be mentioned here that the tools described here are just the ones included with the ESRI Silverlight Viewer and most commonly included in a map. There are additional tools that could be added, and of course, custom tools might be developed and added by users. The advanced programming interfaces (APIs) are well-documented and their free viewers (Silverlight and Flex) are quite straightforward to use. The Silverlight Viewer was used in this project rather than creating the application from scratch since it possessed all of the features necessary out of the box to support visualization and ad-hoc querying.

Conclusions and Recommendations for Future Research

This research has documented an effort to process large amounts of AIS data in near real-time and provide a secure data visualization and ad-hoc Internet GIS querying interface. The AIS data used in this and previous research efforts has proven to be of high quality and precision. It is capable of supporting several different types of analyses, and the related layers (lockages, weather and river gage information) have shown that the content may be combined in real-time and historical applications. For example, with additional AIS antennas at strategic locations throughout the inland waterway network and the number of barges for each tow coming from the corpslocks.com application, a near-complete picture of the vessel traffic picture may be realized. Examining historical data may help visualize casualties, incidents and near-miss incursions on the waterways.

Future research should examine automated detection of vessel events, including lockages, fleeting operations, and arrival/departure notifications for shippers and marine transportation companies. In addition to these detection events, simple geoprocessing functions such as point in polygon tests and proximity analyses could be applied with the weather GIS layers to notify mariners and shipping companies of impending severe weather and to help predict future positions where vessels may be in danger (e.g., slowing down ahead of a storm likely to produce straight-line winds may be advisable so a tow does not encounter the weather). Of course, mobile apps where the data is available on tablet and smartphone platforms should also be explored. The viewer has been demonstrated to be a valuable data visualization tool and it may be customized even further to provide extract, transform and load (ETL) data functions.

References

Dobbins, J., Langsdon, L. "Generation of Inland Waterway Trip Information using Automatic Identification System (AIS) Data." Transportation Research Record, National Research Council, Washington, D.C., August 2012, *in press*.

United States Coast Guard Navigation Data Center website: http://www.navcen.uscg.gov/?pageName=AISMessages. Last Accessed June 2013.

United States MISLE Database Download, website: https://homeport.uscg.mil/mycg/portal/ep/contentView.do?channelId=-18374&contentId=149309&programId=91343&programPage=%2Fep%2Fprogram%2Feditorial.jsp&pageTypeId=13489&contentType=EDITORIAL Last Accessed June 2013.

Microsoft Developer Network (MSDN) data type: STPointFromText (geography data type): http://msdn.microsoft.com/en-us/library/bb933979.aspx. Last accessed June 2013.

NOAA weather website with ArcGIS Server Map Services. http://gis.srh.noaa.gov/ArcGIS/rest/services. Last accessed June 2013.

NOAA weather service, RIDGE Radar Map Service: http://gis.srh.noaa.gov/ArcGIS/rest/services/RIDGERadar/MapServer. Last accessed June 2013.

NOAA weather service, Watch Warn Map Service: http://gis.srh.noaa.gov/ArcGIS/rest/services/watchWarn/MapServer. Last accessed June 2013.

NOAA Advanced Hydrologic Prediction Service (AHPS) Map Service: http://gis.srh.noaa.gov/ArcGIS/rest/services/AHPS_Gauges/MapServer and http://water.weather.gov/ahps/index.php Last accessed June 2013.

USACE Port Series Layer. Geographic information system (GIS) layer and data source. http://www.navigationdatacenter.us/ports/ports.asp. Last accessed June 2013.

Mining Automatic Identification Systems (AIS) Data for Improved Vessel Trip Analysis Capabilities (RI-4)

Great Lakes Component

Final Report

Samir K. Dhar
Ph.D. Candidate and Research Asssistant
University of Toledo
May, 2016

Chapter 1

Introduction

Automatic Identification System (AIS) technology is implemented and used by maritime authorities to monitor and track vessels on waterways. The majority of vessels involved in commercial activity are mandated to have an AIS instrument installed. AIS fitted vessels automatically transmit AIS messages at a regular time intervals. These AIS messages are received by receiving stations and the data are then redirected to maritime authorities. The maritime authorities then use this AIS information to manage vessel traffic within its jurisdiction and beyond. With thousands of vessel involved in commercial activities around US coastal and inland waterways, the volume of AIS messages increases at a rapid pace. As the volume (number of records) of AIS data increase every minute, managing and extracting useful information from the data becomes difficult and time consuming.

Tracking and monitoring increasing numbers of maritime vessels (commercial or recreation) on waterways poses an important issue of safety that is of a major concern by the maritime community to maintain safe navigation. Tracking of vessels from an operational standpoint consists of an exchange of information between a vessel and port authorities on shore, or from one vessel to another vessel in real time to better manage traffic. The information exchanged helps vessel traffic management personnel in identifying a vessel, its location, its speed, course, and destination According to the US Coast Guard (USCG), the concept of monitoring and managing vessel traffic was initially accomplished using a radio call followed by a shore-based radar station. Vessel Traffic Services (VTS) is the authorized body established by port authorities to track and manage traffic within its jurisdiction. VTS is a monitoring body directed by SOLAS (Safety of Life At Sea) under Chapter V – Safety of Navigation and IMO (International Maritime Organization) resolution "IMO Resolution A 857 (20)", adopted on 27 November 1997 (U. S. Coast Guard Navigation Center, 2014) (International Maritime Organization, 2014). VTS operated by the USCG provides real-time monitoring, tracking, and navigational assistance to vessel(s) that are sailing within its region to prevent vessel collisions, groundings, and ramming.

In the past, VTS personnel made use of visual (Closed Circuit Television) and VHF radiotelephone to identify, locate, and communicate with vessel(s). The earlier system had heavy human involvement and was prone to human errors. Then came the radar based system that could locate vessel(s) within a specified range and plot it on the radar screen. To reduce or prevent such unnatural calamities, the time to detect and respond needs to be closer to real time. The radar technology has a plot time of around 4-6 minutes that includes detecting a vessel, compute its location, speed, closest point of approach, and time to closest point of approach. Present day the VTS authorities, along with the vessel operators, use Automated Identification System (AIS) to communicate important location information to the concerned authorities and to nearby vessels. An AIS equipped vessel broadcasts information every 2 seconds when travelling at high speeds (14 Knots and higher) and less frequently when travelling at lower speed or when at berth. The AIS acquired information is used to monitor shipping activity, and then is later recorded and stored in a database for future use or analysis. Over a period of time, the size of the AIS data increases exponentially with the increase of shipping activity within that region. Along with terrestrial AIS, initiatives such as LRIT and E-Navigation are put into place to capture AIS information via satellite from a global perspective. Since satellite operate over higher altitudes, the field of view is larger and, therefore, can cover more area and gather AIS information from more vessel.

1.1 Problem Statement

The continued growth and improvements in the AIS technology is leading to better management of marine traffic in the waterway system. IMO makes it mandatory for commercial vessels of 300 gross registered tonnage (grt) to have AIS installed so that its movements could be monitored within the waterway system and ports. Accurate position information is the key to managing vessels and helping them to navigate through tight channels, ports, and through locks. Especially in ports where maritime activity is dense, the need to efficiently manage vessel is very important and AIS technology provides the VTS officials near real time information about vessels within its jurisdiction. However, with more and more vessels becoming involved in maritime trade, accumulation of AIS information is increasing at a rapid pace.

AIS information captured from shore stations or via satellite gets stored in data stations for current and future use. The huge amount of AIS information collected can unquestionably overload the database within days, leading to delay in query and processing time. Therefore, the problem is how to reduce the size of the AIS database without compromising important and pertinent information about vessel activity within the waterway system. Applications using historic AIS data can definitely benefit from reduced numbers of AIS records since querying or processing fewer records results in a decrease in processing time and faster availability of results. Also, with the e-navigation initiative that relates to Marine domain Awareness, the need to efficiently manage this overwhelming amount of AIS data becomes more important.

1.2 Objectives

The study here addresses this issue of high volume of AIS data that is growing with time within the Great Lakes region. The Great Lakes Navigation System (GLNS) is a close quarter navigation system with a mix of domestic and international shipping traffic along with heavy recreational mix that directly impacts the regional and national economy. The primary aim of the study is to address the high volume of AIS data by developing a technique to reduce the data without losing any pertinent information. Furthermore, AIS data reduced to a manageable size permits further analysis. In this study an application of the reduced AIS data is used for detecting vessel calls at specific terminals and tracking vessel paths.

The study's focus is to understand AIS technology and its contribution to mariner's safety, preventing collisions of vessels, and protecting the environment. Next, the sheer volume of AIS data and associated problems needs to be examined. The key objectives of this study are:

1) Develop a method to automatically acquire AIS data for the Great Lakes region.

2) Create a database to store the acquired AIS data.

3) Mine AIS data to identify important events, anomalies, and errors

4) Develop methods to rectify errors within AIS data.

5) Develop an algorithm to reduce the number of AIS records.

6) Test the algorithm within the Great Lakes region and examine the results.

7) Develop an algorithm to determine vessel calls at a particular dock by using distilled AIS data records.

8) Tracking vessel path to understand a vessel's behavior.

Trade and commerce facilitates the exchange of commodities between two trading entities or social groups. This social phenomenon of trade using the waterway is increasing. The shipping industry has not only grown in numbers of participants, but also in the number of vessels and the sizes of the vessels has seen a significant increase. With an increase in trade, particularly imports, there has been an increase in vessel calls at ports. AIS technology is providing additional help to the VTS authorities in order to better manage their waterway system. With every vessel broadcasting their information every 2 seconds when cruising at a speed of 14 knots or higher, VTS officials are provided with more accurate and near real time information about vessel traffic. However, every vessel broadcasting information so frequently can overwhelm the data storage and management capabilities. As a result, reducing the number of AIS data records becomes an important factor in minimizing processing time and delivering quicker results via software application.

Chapter 2

GLNS & Vessel Traffic Service

2.1 Background and Context

In the United States (US), the Great lakes waterway transportation system is an important and integral part of the US's overall transportation system. The Great Lakes cover an area of 750 miles with a series of interconnected (five) lakes named Lake Superior, Lake Michigan, Lake Huron, Lake Erie, and Lake Ontario (Figure 2-1) (US EPA, 2012). The freight industry utilized these waterway very frequently to transport goods/commodities from port to port. The Great Lakes, and St. Lawrence Seaway provides the only waterway connections between the Midwest States and Eastern Coast in the US (Duluth Seaway Port Authority, 2014). Navigation developments done by US and Canada accomplished, by the US and Canada made it possible for vessels to traverse the rough Niagara Peninsula, St. Lawrence River, and the St. Mary's River with ease (Martin Associates, 2011).

The GLNS is well connected contiguously with series of locks and canals, thus providing safe and efficient navigation. The inland connection starts with the Soo Locks that provide easy passage through the St. Mary's River thereby, connecting Lake Superior to Lake Huron and the lower lakes. The St. Lawrence Seaways connect Lake

Erie to the Atlantic Ocean through Lake Ontario and St. Lawrence River. The St. Lawrence Seaway system is connected by 5 short canals comprising of two main sections of lock systems. The first section is the Welland Canal that connects Lake Erie and Lake

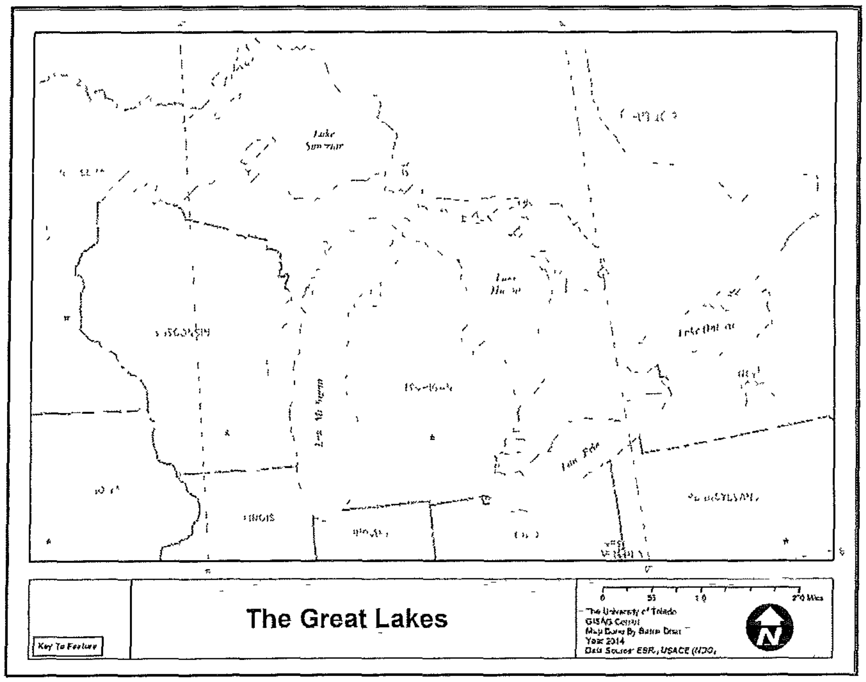


Figure 2-1 The Great Lakes

Ontario with a series of 8 locks managed and operated by Canada. The second section is the Montreal-Lake Ontario section that is comprises of 7 locks, out of which 5 are managed by Canada and 2 by the US.

The Great Lakes, along with its connecting rivers and the St. Lawrence Seaway, makes up the Great Lakes Navigation System (GLNS) that provides a contiguous connection from the Gulf of St. Lawrence to the western end of Lake Superior, which is a distance of 2,342 statute miles (USDOT, 2013). Over the years, the GLNS has been an important waterway system, not only to facilitate domestic and international trade within the Mid-west region, but also to all mariners that use it for recreation or other uses. There are 32 major US ports in the Great Lakes with around a thousand (994) individual

Navigation Points of Interest (NPI) (Dock/Wharfs/Piers/Fleeting Area) that support the maritime industry (Figure 2-2 & 2-3). According to the U.S Army Corp. of Engineers

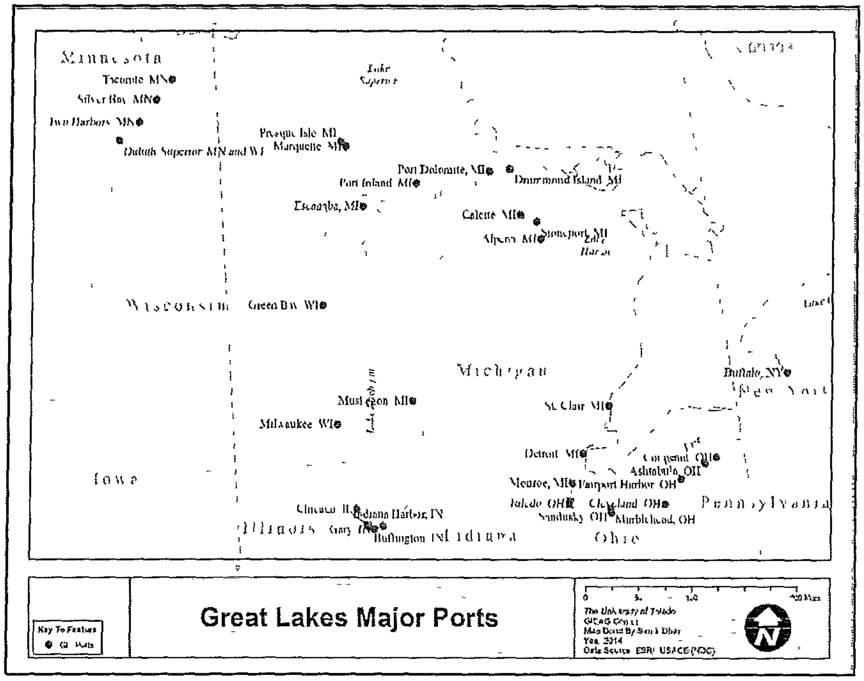


Figure 2-2 Great Lakes Major Ports

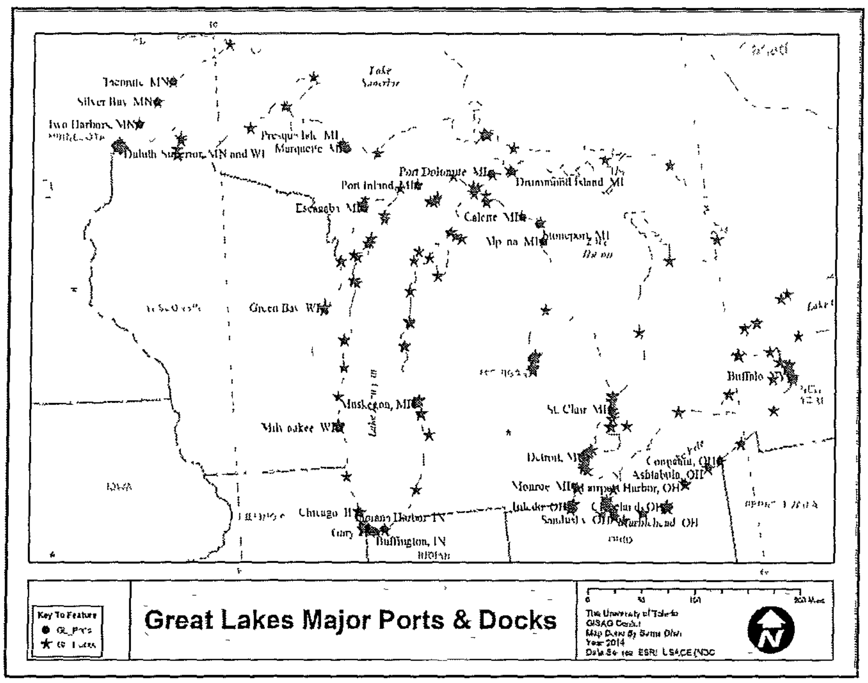


Figure 2-3 Great Lakes Major Ports and Docks

(USACE) the GLNS is very instrumental, not only to the economy within the Mid-west region, but also is very pivotal for the nation's economy. The Great Lakes region has several industries that support the region's economy and use the resources that get hauled over the GLNS. Manufacturing establishments, steel mills, cement plants, construction firms, power plants, agriculture related industries, and other small industries rely heavily on this low-cost transportation system to either move raw materials or to transport finished goods and products.

2.1.1 Great Lakes Economy

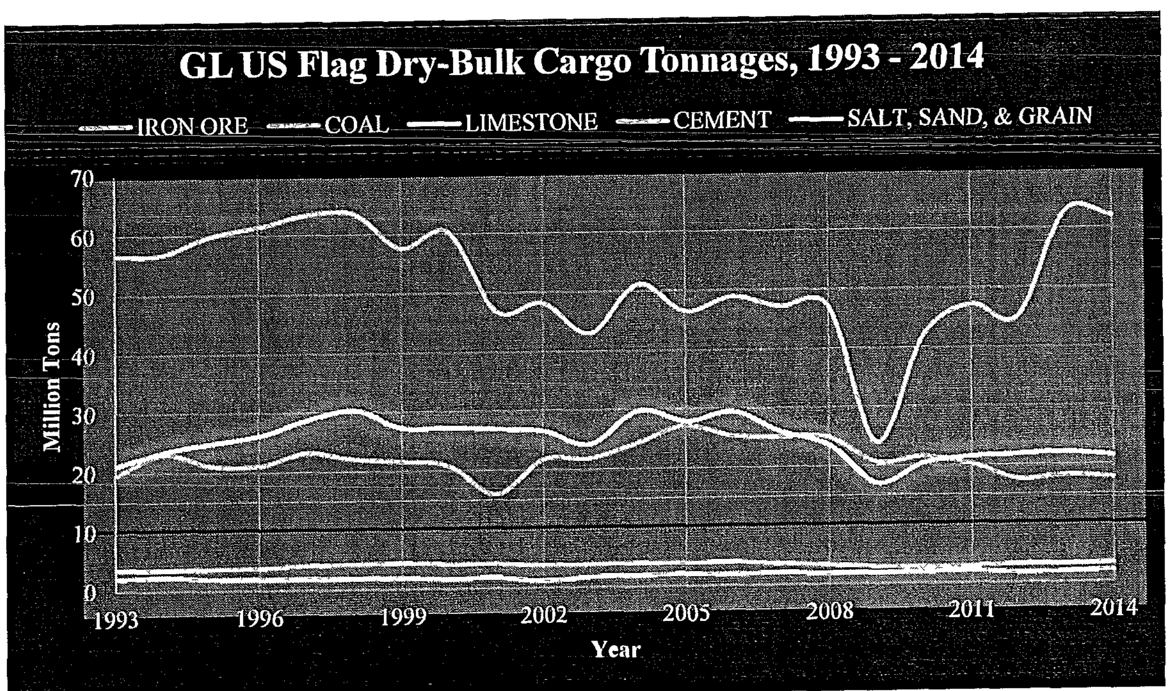
In the US the GLNS is heavily used to move bulk commodities such as iron ore, lime-stone, minerals, coal, sand, salt, agriculture products, and etc. within the domestic front and it also serves the international market. According to the American Association of Port Authorities (AAPA), the GLNS supports 227,000 jobs in the U.S. and Canada and generates $33.5 billion in business revenue (AAPA, 2013). The AAPA also reports that in the U.S. alone, the system supports over 128,000 jobs and generates $18.1 billion in business revenue. According to a study performed by Martin & Associates, 322.1 million metric tons of cargo were handled in 2010 within the GLNS (Martin Associates, 2011).

Year	Tonnage
2003	156,484
2004	178,434
2005	169,411
2006	173,013
2007	160,959
2008	152,381
2009	108,689
2010	129,455
2011	134,899
2012	126,825

Table 2.1 Freight Traffic, 2003-2012 (thousands of short tons) (Source USACE)

The USACE in their annual "Transportation Facts and Information" booklet, released in 2013, reports that 22 of all the ports in the GL region are among the top 100 ports by tonnage in the nation (USACE, 2013). Table 2.1 shows the total tonnage of all commodities moved (foreign and Domestic) on the Great Lakes from 2003 through 2012.

On an average, the GLNS moves 150 million short tons of commodities annually. Table 2.2 shows the annual tonnages of major commodities moved on the GL. Even though the total tonnages started declining from 2007 due to the financial crisis the maritime industry seems to be making a modest recovery by 2010 as the tonnage increased 20% from the year 2009. The Majority of goods transported via the Great Lakes, in regards to tonnage are the dry-bulk commodities with iron ore, coal, and limestone being the top three commodities. Graph 2.1 shows the trend for the major dry-bulk commodities transported (iron ore, coal, limestone, cement, and salt, sand, & grains) over a period of 21 years ranging from 1993-2014.



Graph 2 1 Great Lakes U S -flag dry-bulk cargo Tonnages, 1993–2014 Data (Source LCA)

Most of the iron ore (taconite) movements originate from the Iron Range in Minnesota and the Upper Peninsula of Michigan (Stewart, 2006). Iron ore in the form of taconite gets shipped on vessels to steel mills located around or near the Great Lakes. Coal is primarily used in coal-fired power generating plants located around the GL, along with steel mills, cement plants, and other industries. The majority of the coal mined in the Powder River Basin and the Appalachian gets moved by either rail or barge to the nearest port(s) along the GL, and then they are hauled in vessels to their final destination.

YEAR	IRON ORE	COAL	LIMESTONE	CEMENT	SALT, SAND, & GRAIN	TOTAL
1993	56,602,334	19,540,306	21,169,370	3,607,593	2,088,322	103,007,925
1994	56,827,247	23,164,768	23,481,283	3,532,783	2,384,820	109,390,901
1995	59,846,200	21,143,967	24,913,305	3,689,192	1,983,515	111,576,179
1996	61,404,696	21,056,459	26,137,520	3,734,530	1,918,393	114,251,598
1997	63,370,630	23,244,252	28,755,341	4,159,146	1,944,893	121,474,262
1998	63,523,224	22,057,219	30,358,476	4,251,903	1,898,540	122,089,362
1999	57,683,677	21,633,198	27,310,498	4,417,055	1,905,946	112,950,374
2000	60,332,678	20,760,474	27,288,089	4,144,774	1,616,944	114,142,959
2001	46,924,703	15,965,758	26,988,622	4,136,897	1,852,205	95,868,185
2002	48,195,327	21,743,831	26,554,243	3,817,911	1,147,511	101,458,823
2003	43,016,285	21,879,426	24,239,110	3,851,487	1,758,127	94,744,435
2004	51,201,511	24,416,349	29,861,141	3,965,401	1,889,249	111,333,651
2005	46,572,119	27,207,350	27,935,513	3,892,822	2,052,645	107,660,449
2006	48,972,112	25,360,113	29,489,410	3,997,703	1,912,416	109,731,754
2007	47,206,383	25,170,629	25,966,057	3,602,488	2,095,694	104,041,251
2008	47,223,494	24,971,623	23,632,070	3,294,071	1,831,557	100,952,815
2009	24,031,087	20,674,888	17,067,232	2,865,323	1,828,213	66,466,743
2010	42,028,418	21,539,866	20,410,266	2,782,259	1,923,704	88,684,513
2011	47,224,743	20,239,327	21,434,839	2,817,846	2,067,506	93,784,261
2012	45,189,027	17,579,948	21,794,394	3,183,388	1,727,879	89,476,648
2013	43,852,041	18,237,640	22,111,494	3,129,748	1,823,769	89,156,705
2014	45,624,207	17,772,489	21,459,429	3,248,033	2,035,985	90,142,157

Table 2.1 US- Flag Carriage of Bulk Cargo on the GL, 1993-2014 (in short tons) (Source Lake Carrier Association)

Limestone is primarily used by cement, construction industry, steel mills, power plants, and taconite plants These industries located around the Great Lakes receive limestone from quarries located in Michigan and Ohio, mainly by vessel. There are other commodities that uses the GLNS such as salt, grains, sand, gravel, chemicals, petrolcum, and petroleum products. These industries located around the GL region rely and thrive on the low-cost transportation network provided by the GLNS. According to the USACE, the GLNS system saves U.S. businesses $3.6 billion of transportation cost annually when compared to the next mode of transportation (rail or truck). Besides the saving it provides the maritime industry it also brings in $4.6 billion (2010) in federal, state, and local taxes through direct and indirect activity on the GLNS (Martin Associates, 2011). The USACE that manages the US portion of the GLNS aims in providing safe and efficient infrastructure to the marine community by constantly improving infrastructure and implementing latest technology to monitor and navigate vessels through locks and channels. The following sections will provide an in-depth literature review about the Vessel Traffic Service (VTS) and the technology used by VTS official, in monitoring and tracking vessels.

2.2 Vessel Traffic Service

The maritime community with all its tradition and complexity is one of the oldest modes of transportation in the world. The marine community is comprised of commercial vessels and recreation vessels that use the waterways for the purpose of moving freight, transporting people, fishing, research, and pleasure cruises. As the number of these users increase over a period of time, the issue of safety becomes more pertinent. Tracking of commercial or recreation vessels on any waterways is the first step for maintaining safe navigation and for better management of marine traffic. Identifying, monitoring, and tracking of vessels incorporates an exchange of information between ships and shores in real time in order to better manage vessel traffic.

According to the USCG, the concept of monitoring and managing vessel movements is initially accomplished using shore-based radar stations. This identifying, tracking, and monitoring of vessels within a region by an authorized entity makes vessel navigation safe and efficient.

2.2.1 Vessel Traffic Service History

VTS is a combination of operational procedure directed by IMO and SOLAS, electronic equipment, regulation, and personnel ((Lin & Huang, 2006). The operation of VTS initiates when any vessel enters the VTS region, the officer on board inputs the vessel's identity, geographic location, and voyage related information into the VTS system. The VTS system gets the necessary information about the vessel and starts to monitor her movements until that vessel leaves the VTS area. The first shore-based radar system to provide VTS was installed and became fully functional in the port of Liverpool in 1949. In the US, the first Vessel Traffic Service (VTS) center with radar system capabilities along with VHF-FM radiotelephone, was installed in the port of Long Beach, California. In the USA, the first Federal VTS managed by the Coast Guard originated in the San Francisco Bay (NAVCEN, 2011). This initiative stemmed from a R&D effort in 1968 called the Harbor Advisory Radar. Participation in the system was voluntary, so few vessels participated and contributed to the services.

Initially, VTS essentially appeared to the users as an information (vessel traffic) provider for decision makers. The decision makers then used the information to assess the vessel traffic situation within a particular port. It was only after the big collision of two tankers (Arizona Standard and Oregon Standard) in January 18, 1971, under the Golden Gate Bridge that maritime related safety issues related to VTS were taken seriously. Subsequently, the Port and Waterways Safety Act of 1972 bestowed power to the US Coast Guard to install, maintain, and operate VTSs within the US navigation waterways (coastal, inland, and great-lakes) (NAVCEN, 2011).

Initially, the Coast Guard begin to install VTS in congested ports such as Puget Sound, Louisville, Houston-Galveston, Prince William Sound, Berwick Bay, and St.

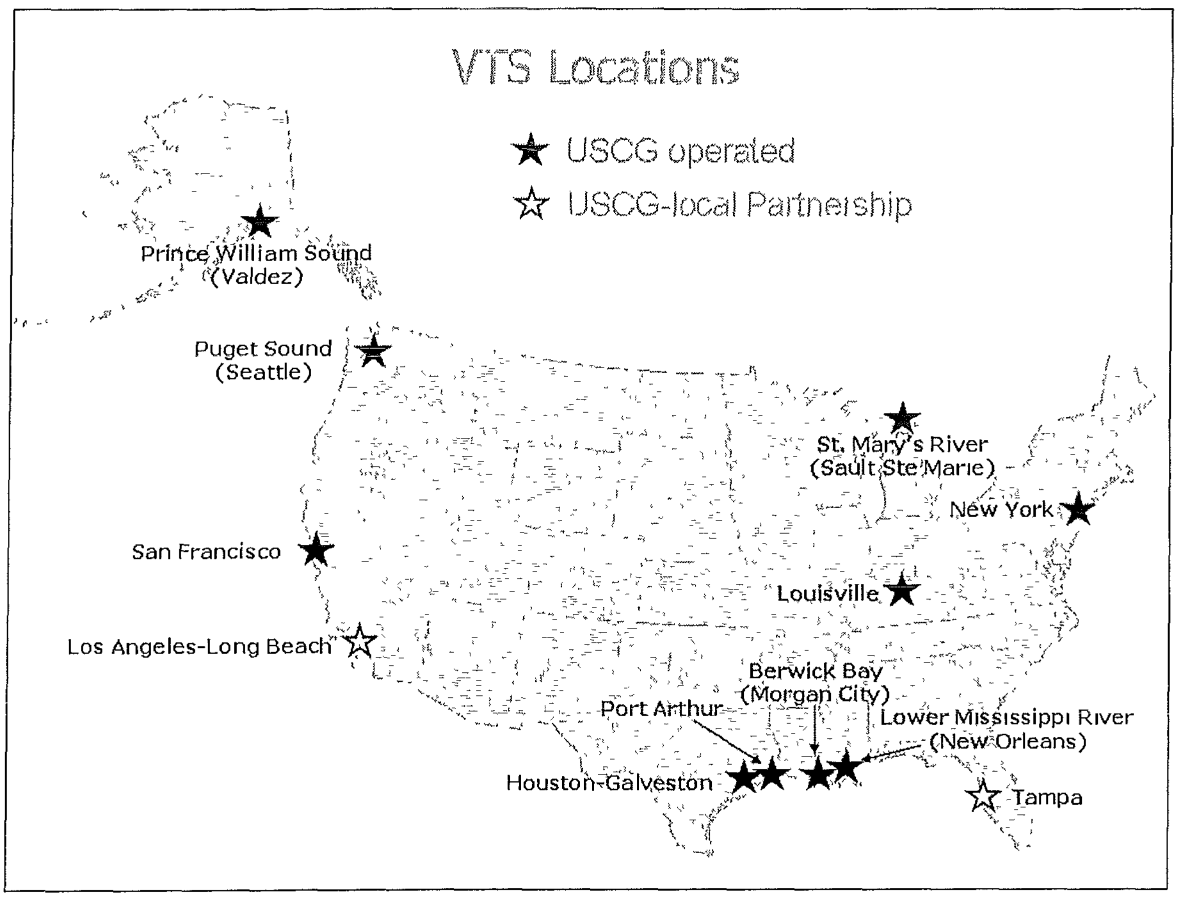


Figure 2-4 VTS Location in US (Source. USCG - NAVCEN)

Mary's river at Sault Ste Marie (Figure 2-4) besides others, such as New York and New Orleans that provided VTS services on a voluntary basis (NAVCEN, 2011). Slowly, the effort was cut for a few years due to monetary issues. It was revived again after the EXXON VALDEZ disaster. Exxon Valdez is a 301m long oil tanker that ran aground in Prince William Sound spilling thousands of barrels of crude oil in Alaska causing maritime safety and environmental concerns. It was following this incident that the Oil Pollution Act of 1990 was enacted by congress, which mandated the Coast Guard to make participation in VTS of all maritime vessels mandatory (NAVCEN, 2011).

2.2.2 Automatic Radar Plotting Aids

Automatic Radar Plotting Aids (ARPA) is a shore based radar monitoring system that was implemented during 1980's. This radar based ARPA system aided in accurately locating vessels, measuring the speeds of vessels, identifying vessel's course and heading, calculating closest point of approach (CPA), and time to CPA (Lin & Huang, 2006). The functionality of ARPA is based on the radar emitting short pulses of radio waves that propagates everywhere irrespective of direction and when it hits an object (vessel or shore), part of the pulses are reflected back. This partial reflected pulse captured by an antenna is then used to compute the objects location, speed, CPA, and TCPA. The antenna is usually placed at a higher elevated location so it can cover more area. The height of the antenna is very essential in computing the distance of the vessel from the shore due to curvature of the earth (Lin & Huang, 2006). Smaller vessels like a fishing boat, are not detected unless they are close to the shore. Also, blind spots such as mountains can hinder spotting a vessel. In the early stages when using ARPA all the information such as speed, course, CPA, and TCPA were manually computed. However, as technology improved it was coupled with computers where the computation and plotting were done automatically, leading to a reduction in the rate of plotting time from 4 – 6 minutes to 2 minutes (Lin & Huang, 2006). The output from the ARPA is coupled with electronic chart to provide a better relative location of the vessel. Identity of the target vessel is acquired through VHF calls. Even though ARPA provides better surveillance to authorities, there are constant efforts made to improve situation awareness in marine waterways by adopting new technology to track vessel(s).

2.3 Automatic Identification System (AIS)

Marine navigation safety and security is slowly becoming a growing concern with the increase in maritime vessel traffic and vessel size. More accurate location information of a vessel is required to aid mariners' navigation safety. VTS authorities implemented Automated Radar Plotting Aid (ARPA) which is a radar based system to assist in acquiring vessel's navigation information to decide whether vessels are in their proper course or inline of collision with another vessel. This ARPA system takes around 4-6 minutes to show any course alteration of a vessel. Latest computer technology provides faster processing and helped ARPA system to reduce its computation time. The VTS authorities still needed information at quicker rate to efficiently manage its waterways. This led to the development of the Automatic Identification System (AIS). AIS is an autonomous radio-based technology developed under the request and guidance of the International Maritime Organization (IMO) to foster safe maritime navigation within waterways by tracking vessel traffic in real time.

Chapter 3

AIS Technology

The overall AIS technology is comprised of an on-board transponder that broadcasts information over VHF maritime channels (87B – 161.975 MHz or 88B – 162.025 MHz). This allows marine vessels to communicate with each other as well as with coastal base stations, Search and Rescue (SAR), and Aids to Navigation (AtoN). Information communicated includes a vessel's identity, location, speed, draft, and more navigational data is transmitted at regular interval via a VHF transmitter built within the transponder. These broadcasts or signals are captured or received by transponders installed on either a land side system or on a vessel. The captured information is then parsed (decoded) and can be either displayed on a screen that is overlaid with navigation charts or can be stored in a database for future use. An AIS equipped vessel broadcasts information every 2 seconds when travelling at a speed greater than 14 knots, and less frequently when travelling at low speed and even less when the vessel is idle or at berth (Table 3.1) .

Ship's Dynamic Conditions	Reporting Interval
Ship at anchor or moored and not moving faster than 3 knots	3 min
Ship at anchor or moored and moving faster than 3 knots	10 s
Ship 0-14 knots	10 s(1)
Ship 0-14 knots and changing course	3 1/3 s
Ship 14-23 knots	6 s(1)
Ship 14-23 knots and changing course	2 s
Ship >23 knots	2 s
Ship >23 knots and changing course	2 s

Table 2 1 CLASS A AIS Equipment Reporting Intervals (Source ITU)

3.1 AIS Regulations

All and every vessel over 300 gross tons is mandated under port security regulation to have AIS equipment installed/fitted on its board. The USCG has provided detailed guidelines on which types of vessels are needed to have AIS fitted and vessels that can be exempt from it. According to the regulation TITLE 46 – Shipping, Chapter 701 Port Security, Sub-Chapter 1 section 70114, the rule mandates vessel operating in US waterways must have AIS fitted (USCG - DHS, 2008).

> ***§70114. Automatic identification systems***
> *(a) System Requirements -(1) Subject to paragraph (2), the following vessels, while operating on the navigable waters of the United States, shall be equipped with and operate an automatic identification system under regulations prescribed by the Secretary·*
> *(A) A self-propelled commercial vessel of at least 65 feet overall in length.*
> *(B) A vessel carrying more than a number of passengers for hire determined by the Secretary.*
> *(C) A towing vessel of more than 26 feet overall in length and 600 horsepower.*
> *(D) Any other vessel for which the Secretary decides that an automatic identification system is necessary for the safe navigation of the vessel*
>
> *(2) The Secretary may-*
> *(A) exempt a vessel from paragraph (1) if the Secretary finds that an automatic identification system is not necessary for the safe navigation of the vessel on the waters on which the vessel operates, and*
> *(B) waive the application of paragraph (1) with respect to operation of vessels on navigable waters of the United States specified by the Secretary if the Secretary finds that automatic identification systems are not needed for safe navigation on those waters.*

AIS has been well accepted globally and mariners have adopted technical and broadcasting standards developed collectively by the International Maritime Organization (IMO), the International Telecommunication Union (ITU), and the International Electronic Commission (IEC). The International Convention for Safety of Life at Sea (SOLAS) drafts regulations (19.2.1 Chapter V) that details specification for shipborne navigational systems and equipment and makes it mandatory for AIS carriage for vessels operating in international waters.

3.2 AIS Equipment Setup

A general AIS system is comprised of a VHF transmitter, two VHF TDMA receivers, one VHF DSC receiver, and standard marine electronics links to a display unit and other sensors. According to the ITU Radio-communication standards, an AIS unit should allow automatic exchange of navigation data to shore-based station and to nearby vessels (ITU, 2014). The system should use SOTDMA technology to broadcast primary AIS messages along with safety related communications and application specific messages. There are primarily two types of AIS devices (Class A and Class B) available for commercial vessels, and other AIS devices are used in base station, AtoN, and SAR transmitter (NAVCEN, 2014).

Class A: It is designed for typical commercial vessels such as cargo vessels and passenger vessels. It conforms to all performance standards and carriage required by IMO. It has the capability to broadcast different types of messages at a higher power that can cover greater distance. The device has the capability to receive text messages from base stations for safety or security purposes (NAVCEN, 2014).

Class B: This is typically designed for non-commercial related vessels either engaged in recreation or other purposes. Since they do not meet all of the performance standards adopted by IMO, it broadcasts at a lesser frequency than the Class A device and at a lower power (Table 3.2). It has to capability to receive safety related text messages, but it cannot transmit them. Two types of Class B devices are available· one using Carrier Sense Time Division Multiple Access (CS-TDMA) technology and the other using Self-Organizing Time Division Multiple Access (SO-TDMA) technology (NAVCEN, 2014).

Platform's condition	Nominal reporting interval	Increased reporting interval
Class B "SO" shipborne mobile equipment not moving faster than 2 knots	3 min	3 min
Class B "SO" shipborne mobile equipment moving 2–14 knots	30 s	30 s
Class B "SO" shipborne mobile equipment moving 14–23 knots	15 s	30 s[(3)]
Class B "SO" shipborne mobile equipment moving >23 knots	5 s	15 s[(3)]
Class B "CS" shipborne mobile equipment not moving faster than 2 knots	3 min	–
Class B "CS" shipborne mobile equipment moving faster than 2 knots	30 s	–
Search and rescue aircraft (airborne mobile equipment)	10 s[(2)]	–
Aides to Navigation	3 min	–
AIS base station	10 s[(1)]	–

Table 3 2. Reporting Intervals for Class B Equipment (Source. ITU)

AIS Base Stations: This device is a shore-based station that provides information to vessels regarding time, meteorological or hydrological information, navigation information, or position of other close by vessels (NAVCEN, 2014).

Aid to Navigation (AtoN): It can be either a shore-based or a mobile station that has the capability to transmit status of AtoN or any application related messages (NAVCEN, 2014).

Search and Rescue (SAR) Transmitter: It is mainly used in maritime survivor locating devices attached to life boats or rafts. This has the capability to transmit messages relating to position (latitude and longitude) (NAVCEN, 2014).

Typically an AIS system fitted in a vessel is used either as a lone system or in conjunction with a radar based system. Most, larger vessels use AIS in conjunction with the radar based system. Figure 3-1 shows a general mount diagram of a typical AIS Class A transponder with above deck and below deck designs. The above deck is comprised of two antennas a) VHF Antenna for transmitting and receiving AIS messages, and b) GPS antenna to get positional information. The below deck settings is comprised of the Pilot equipment is connected to a display unit and other sensors via a junction box. The sensors attached to the AIS device are few input devices such as a gyro-compass to acquire heading information, a speed and distance measuring device, and an electronic GPS system.

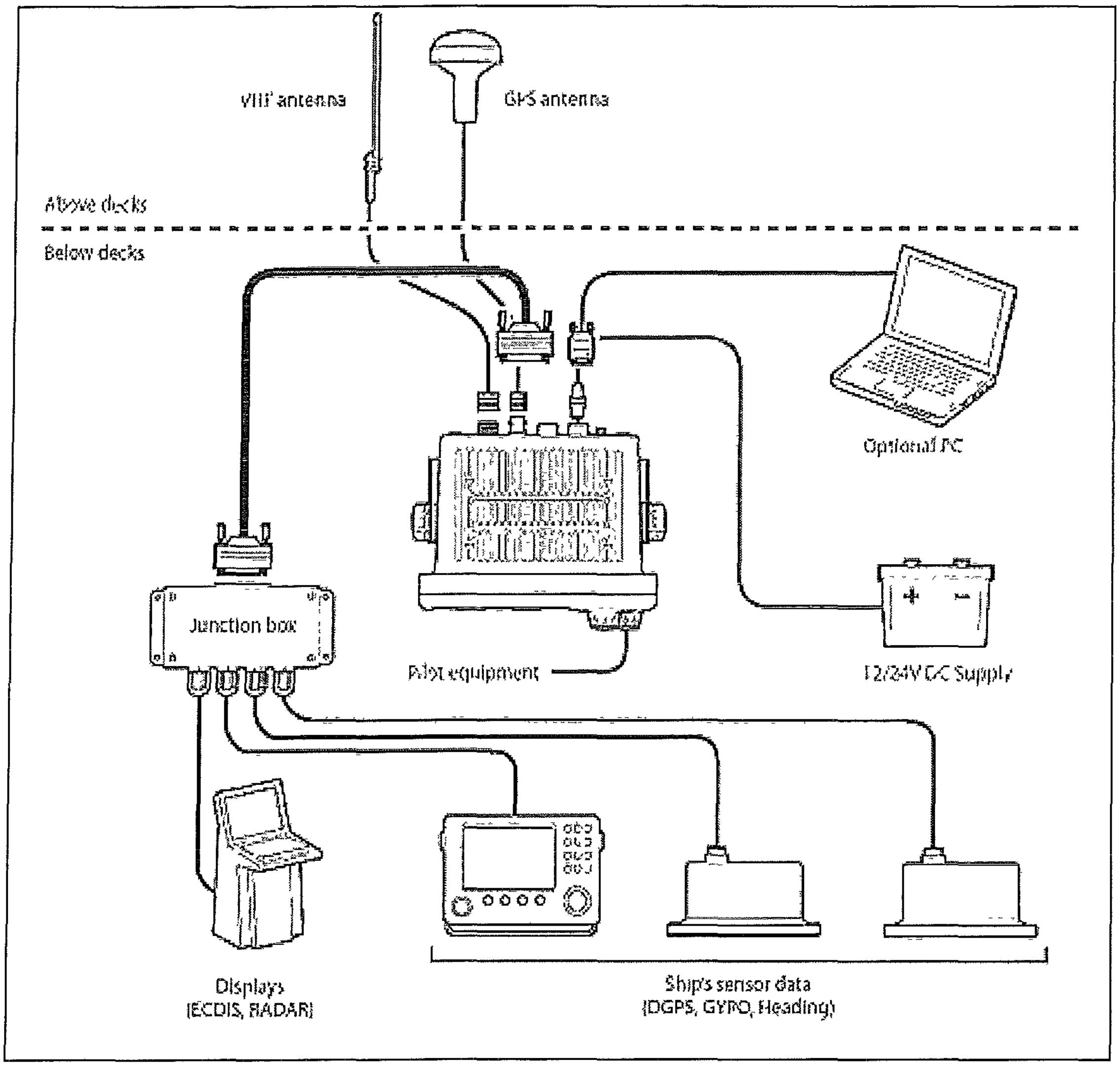


Figure 3-1 Typical AIS Class A transponder Mount Diagram (Source www aisnautica com)

3.3 Technical Model of AIS

The USCG in its Resolution MSC.347 (91) adopted on 30 November 2012 recommends, especially for the Class A AIS device, to meet the technical requirements specified by ITU in its recommendation Rec. ITU-R M.1371. The ITU specification on the technical structure of AIS device in its recommendation is in accordance with the Open System Interconnection model (OSI).

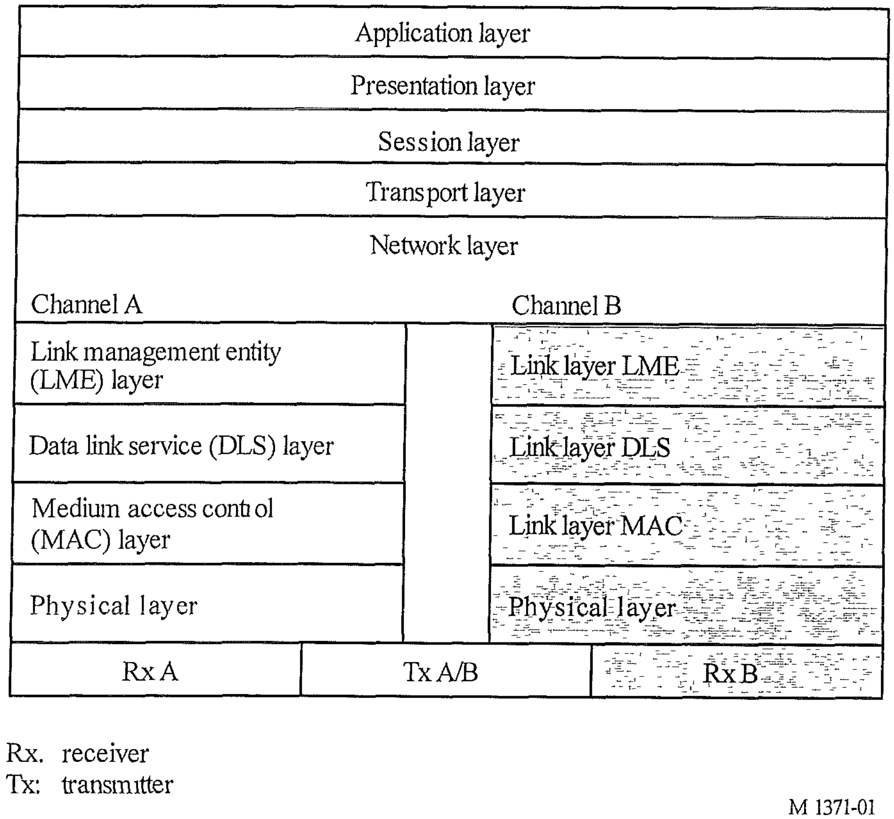


Figure 3-2 OSI Model of AIS (Source ITU)

Figure 3-2 shows the OSI model of AIS as specified by ITU and adopted by IMO and IEC. The standards specified by ITU in its recommendation only covers the Physical Layer, Link Layer, Network Layer, and Transport Layer.

Physical Layer: This layer facilitates the transfer of a bit-stream from the originator to the Link Layer. The AIS is designed to operate in the VHF maritime mobile band with two settings of high (25 kHz) and low (12.5 kHz) with high being the default setting. Table 3.3 presents specification for different parameters within both the settings (High and Low) which are independent of each other.

Symbol	Parameter Name	Units	Low setting	High setting
PH.RFR	Regional frequencies (range of frequencies within RR Appendix 18)(1)	MHz	156.025	162.025
PH CHS	Channel spacing (encoded according to RR Appendix 18 with footnotes)(1)	kHz	25	25
PH.AIS1	AIS 1 (default channel 1) (2087)(1) (see § 2.3.3)	MHz	161 975	161.975
PH AIS2	AIS 2 (default channel 2) (2088)(1) (see § 2 3.3)	MHz	162 025	162.025
PH.BR	Bit rate	bit/s	9 600	9 600
PH.TS	Training sequence	Bits	24	24
PH.TXBT	Transmit BT product		~0.4	~0.4
PH.RXBT	Receive BT product		~0 5	~0.5
PH MI	Modulation index		~0.5	~0.5
PH.TXP	Transmit output power	W	1	12.5(2) / 5(3)

(1) See Recommendation ITU-R M.1084, Annex 4.

(2) Except for Class B "SO"

(3) For Class B "SO".

Table 3 3 Parameter Settings for High and Low setting (Source ITU)

Link Layer: This layer presents specification for packaging of the bit-stream data. Figure 6 shows that the link layer is broken down into three sub-layers viz. Medium Access Control (MAC), Data Link Service (DLS), and Link Management Entity (LME). The **MAC** sub-layer uses the TDMA technology to provide access to the VHF data link. It uses the Universal Time Control to synchronize with base station and also to access the current unused slot in a frame. A frame is one minute that is divided into 2250 slots, with each slot identified by its index (0 – 2249), and its length is equal to 26.67 ms that is capable of handling 256-bit of information. The **DLS** sub-layer activates and releases the data links after receiving or transmitting a transmission. The data transfer use a bit-control protocol which is slightly based on the High-Level Data Link Control (HDLC) that defines the transmission packet structure. Figure 3-3 shows the structure

Training sequence	Start flag	Data	FCS	End flag	Buffer

Figure 3-3 Transmission Packet Structure (Source ITU)

of a transmission packet with a total length of 256 bits. The packet starts with a Training sequence which is a bit pattern of alternating 0's and 1's and is of 24 bits long. The Start and the End flag is of 8 bits long and it denotes the start and end of the data stream. The Data part of the packet structure is 168 bits long; therefore, in cases with a long transmission message, the data is broken down to multiple transmission packets. The FCS (Frame Check Sequence) calculates the checksum of the data portion using the cyclic redundancy check (CRC) 16-bit polynomial to detect error in transmission. The Buffer part is 24 bits long and is comprised of a combination of bit stuffing (4 bits), distance delay (14 bits), and synchronization jitter (6 bits). The LME sub-layer manages the operations of the Physical layer and the two sub-layers i.e. MAC and DLS of the Link layer. It provides access to the data links by using the TDMA access scheme to organize an unused slot. With AIS devices being autonomous and continuous in its operations, the SOTDMA access scheme is used since it quickly solves slot allocation conflicts.

Network Layer: This layer is used to establish and maintain channel connection during data transmissions. It also manages the priority level of different message types and handles them accordingly with high priority messages being handled first. It determines the reporting rate (Rr) of messages and also the even distribution of transmission packets between two channels (87B and 88B) by alternating between the channels. The network layer also mitigates issues such as data link congestion by releasing messages furthest away from it.

Transport Layer: This layer converts the bit-stream data into transmission packets of the specified size before transmission and also organizes the received packets in sequence. Any communication with the higher layers (e.g. corresponding acknowledgement of message) has to be accomplished via presentation interface.

3.4 AIS Message

AIS unit broadcasts automatically broadcast vessels dynamic information and other information as programmed to shore-based receivers and to nearby vessels in a self-organized fashion (Figure 3-4) (ITU, 2014). The AIS unit is usually programmed to broadcast different message types as required by the VTS authorities or depending on vessel types. There are 27 very specific message types as defined by ITU (ITU – R M

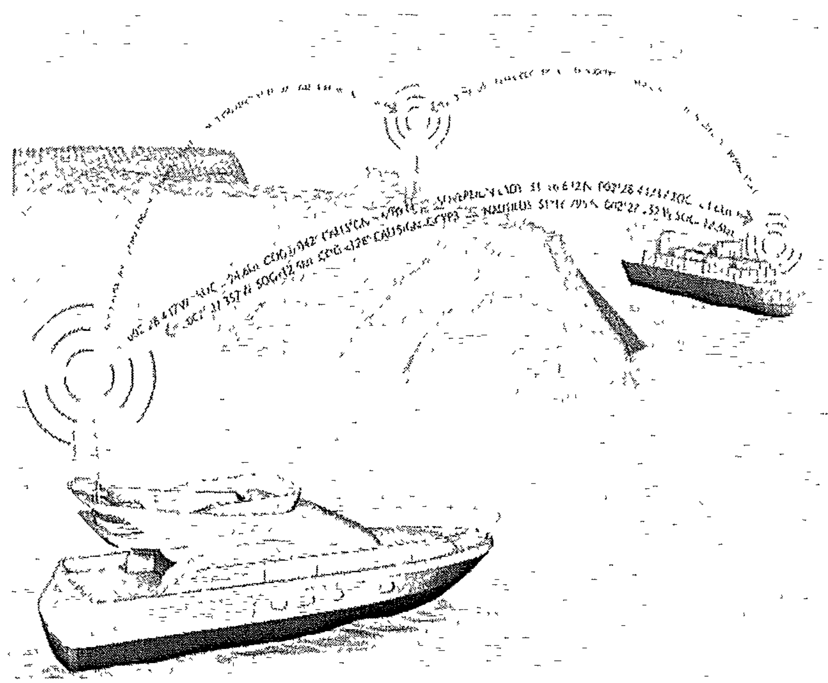


Figure 3-4 AIS Message Broadcasted by Vessel(s) (Source IMO)

1371) and adopted by USCG. There is also a provision for additional message types (Message type 28 – 63) that could be incorporated in the future if the need arises. Table 6 is a detail list of message types that the USCG has adopted as defined by IMO along with ITU.

A commercial vessel fitted with a CLASS A AIS unit will broadcast message types 1, 2, 3, and 5 every 2 to 10 sec depending on the vessel's speed, and every 3 minutes when the vessel is anchored or at berth. AIS messages broadcast by commercial vessels are primarily comprised of three part viz. static, dynamic, and voyage related information. The Static information provides the Vessel's unique ID, name, and physical characteristics. The dynamic part (position report) of the AIS message contains information regarding the speed of the vessel, location of the vessel, draft at that location, heading, and course of the vessel. The voyage related information is comprised of the vessel's destination and the expected time of arrival (ETA). Message types 1, 2, and 3 relate to the dynamic parts of the AIS information and the common parameters among these three message types as shown in Table 6. The common parameters in the message related to the vessels current navigation status, speed of the vessel, its positional information, heading, and current time in UTC format. Values for certain parameters such as navigation status and rate of turn are given in Table 3.4 & 3.5 respectively.

Type	Message Name	Message Description
1	Position Report Class A	Scheduled position report, Class A shipborne mobile equipment
2	Position Report Class A (Assigned schedule)	Assigned scheduled position report, Class A shipborne mobile equipment
3	Position Report Class A (Response to interrogation)	Special position report, response to interrogation, Class A shipborne mobile equipment
4	Base Station Report	Position, UTC, date and current slot number of base station
5	Static and Voyage Related Data	Scheduled static and voyage related vessel data report, Class A shipborne mobile equipment
6	Binary Addressed Message	Binary data for addressed communication
7	Binary Acknowledge	Acknowledgement of received addressed binary data
8	Binary Broadcast Message	Binary data for broadcast communication
9	Standard SAR Aircraft Position Report	Position report for airborne stations involved in SAR operations only
10	UTC and Date Inquiry	Request UTC and date
11	UTC and Date Response	Current UTC and date if available
12	Addressed Safety Related Message	Safety related data for addressed communication
13	Safety Related Acknowledgement	Acknowledgement of received addressed safety related message
14	Safety Related Broadcast Message	Safety related data for broadcast communication
15	Interrogation	Request for a specific message type can result in multiple responses from one or several stations
16	Assignment Mode Command	Assignment of a specific report behavior by competent authority using a Base station
17	DGNSS Binary Broadcast Message	DGNSS corrections provided by a base station
18	Standard Class B CS Position Report	Standard position report for Class B shipborne mobile equipment to be used instead of Messages 1, 2, 3
19	Extended Class B Equipment Position Report	No longer required. Extended position report for Class B shipborne mobile equipment, contains additional static information
20	Data Link Management	Reserve slots for Base station(s)
21	Aid-to-Navigation Report	Position and status report for aids-to-navigation
22	Channel Management	Management of channels and transceiver modes by a Base station
23	Group Assignment Command	Assignment of a specific report behavior by competent authority using a Base station to a specific group of mobiles
24	Static Data Report	Additional data assigned to an MMSI, Part A Name, Part B Static Data
25	Single Slot Binary Message,	Short unscheduled binary data transmission Broadcast or addressed
26	Multiple Slot Binary Message With Communications State	Scheduled binary data transmission Broadcast or addressed
27	Position Report For Long-Range Applications	Class A and Class B "SO" shipborne mobile equipment outside base station coverage
28 - 63	Undefined Reserved for future use	N/A

Table 3 4 AIS Message Types (Source USCG- NAVCEN)

Parameters	Description
Message Type	Identifier for this message 1, 2 or 3
Repeat Indicator	Used by the repeater to indicate how many times a message has been repeated See Section 4 6 1, Annex 2, 0-3, 0 = default, 3 = do not repeat any more
MMSI	MMSI number
Navigation Status	See Navigation Status Table 8
Rate of Turn (ROT)	See ROT Table 9
Speed Over Ground (SOG)	SOG in 1/10 knot steps (0-102 2 knots), 1023 = not available, 1022 = 102 2 knots or higher
Position Accuracy	The position accuracy (PA) flag 1 = high (<= 10 m), 0 = low (> 10 m), 0 = default
Longitude	Longitude is given in in 1/10000 min, divide by 600000 0 to obtain degrees
Latitude	Latitude is given in in 1/10000 min, divide by 600000 0 to obtain degrees
Course Over Ground (COG)	COG in 1/10 = (0-3599) 3600 (E10h) = not available = default 3 601-4 095 should not be used
True Heading (HDG)	Degrees (0-359) (511 indicates not available = default)
Time Stamp	Second of UTC timestamp
Maneuver Indicator	0 = Default, 1= No special Maneuver, 2= engaged in special maneuver
Spare	Not used Should be set to zero Reserved for future use
RAIM flag	0 = RAIM not in use(default), 1 = RAIM in use
Radio status	Diagnostic information for the radio system (Rec ITU-R M 1371-5 Table 49)

Table 3 5· Common Parameters for Message Type 1,2,3 (Source USCG NAVCEN)

Message type 5 reports the static information of the vessel along with the current voyage of the vessel. This message type is broadcasted every 6 minutes and is used only by CLASS A shipborne devices when reporting static and voyage related data (USCG NAVCEN, 2014). Table 3.6 details the parameters that make up the message type 5 that primarily includes the vessel's unique identification number (IMO or CG number), vessel name, vessel type, destination, and expected time of arrival.

Parameter	Description
Message ID	Identifier for this Message 5
Repeat Indicator	Used by the repeater to indicate how many times a message has been repeated See Section 4 6 1, Annex 2, 0-3, 0 = default, 3 = do not repeat any more
MMSI	MMSI number
AIS Version	0 = station compliant with Recommendation ITU-R M 1371-1 1 = station compliant with Recommendation ITU-R M 1371-3 (or later) 2 = station compliant with Recommendation ITU-R M 1371-5 (or later) 3 = station compliant with future editions 0 = station compliant with Recommendation ITU-R M 1371-1 1 = station compliant with Recommendation ITU-R M 1371-3 (or later) 2 = station compliant with Recommendation ITU-R M 1371-5 (or later) 3 = station compliant with future editions
IMO Number	IMO ship ID number
Call Sign	Craft associated with a parent vessel, should use "A" followed by the last 6 digits of the MMSI of the parent vessel Examples of these craft include towed vessels, rescue boats, tenders, lifeboats and life-crafts
Vessel Name	Official name of the vessel
Ship Type	0 = not available or no ship = default 50-99 = as defined in Table 8 100-199 = reserved, for regional use 200-255 = reserved, for future use Not applicable to SAR aircraft
Dimension to Bow	Dimension to Bow in meters
Dimension to Stern	Dimension to Stern in meters
Dimension to Port	Dimension to Port in meters
Dimension to Starboard	Dimension to Starboard in meters
Position Fix Type	Reference point for reported position
ETA month (UTC)	1-12, 0=N/A (default)
ETA day (UTC)	1-31, 0=N/A (default)
ETA hour (UTC)	0-23, 24=N/A (default)
ETA minute (UTC)	0-59, 60=N/A (default)
Draught	Draught at that location in meters/10
Destination	Destination Port
DTE	Data Termninal Equipment ready, (0 = avilable, 1 = not avialable = default)
Spare	Spare. Not used Should be set to zero Reserved for future use

Table 3 6 Parameters for Message Type 5 (Source USCG NAVCEN)

3.5 AIS Working Concept

A schematic diagram Class A ship-borne AIS station as specified by the IEC 61993-2 document primarily consist of a transponder (one VHF transmitter and two VHF receivers) that transmits and receives AIS messages over two dedicated marine VHF channels using TDMA technology (Figure 3-5). Decoding and encoding of the message received is performed before it gets parsed for it to be displayed. A link to the display unit that could be either radar, Electronic Chart Display and Information System (ECDIS) or dedicated devices. The unit is connected to sensors via a junction box in order to acquire information such as speed, heading, and course. Position information is acquired by a GPS antenna along with time reference. DSC (Digital Selective Calling) VHF receivers are tuned to receive special commands via a specified channel (70) and are a core part of the global Maritime Distress Safety System (GMDSS). Messages broadcast by a vessel can be received by base stations or any vessel(s) within a particular range. The AIS transponders use the TDMA technology operating on two VHF radio

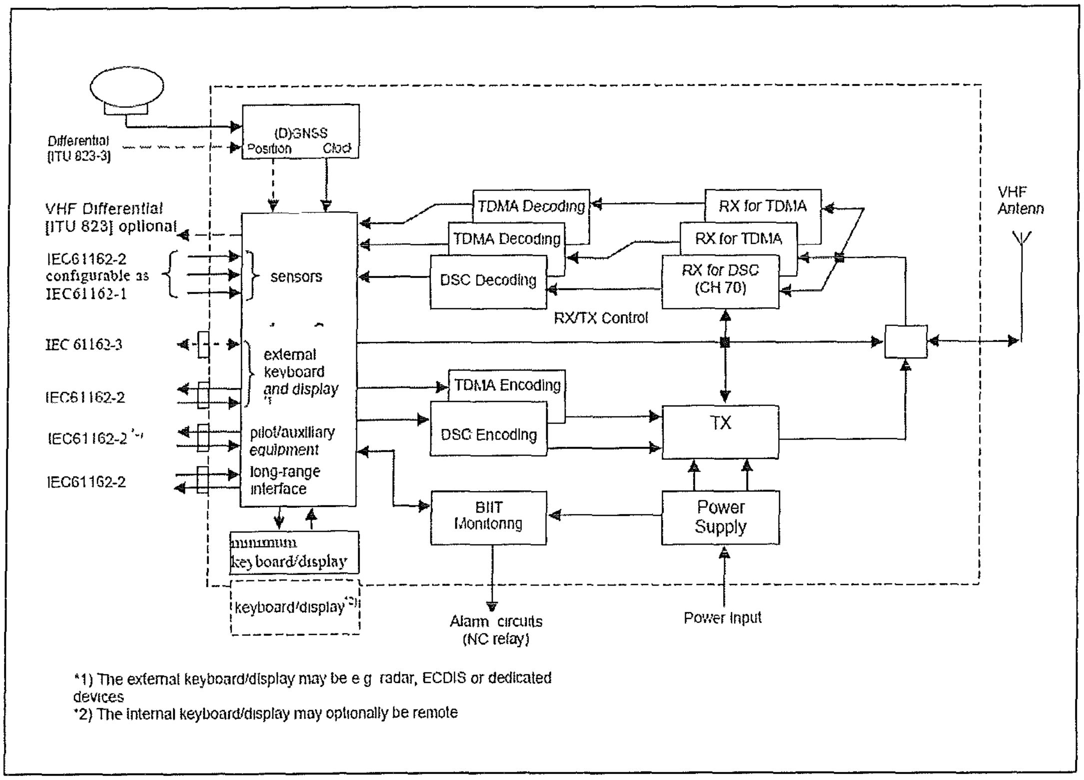


Figure 3-5 Schematic Diagram of Class "A" AIS Station (Source IEC)

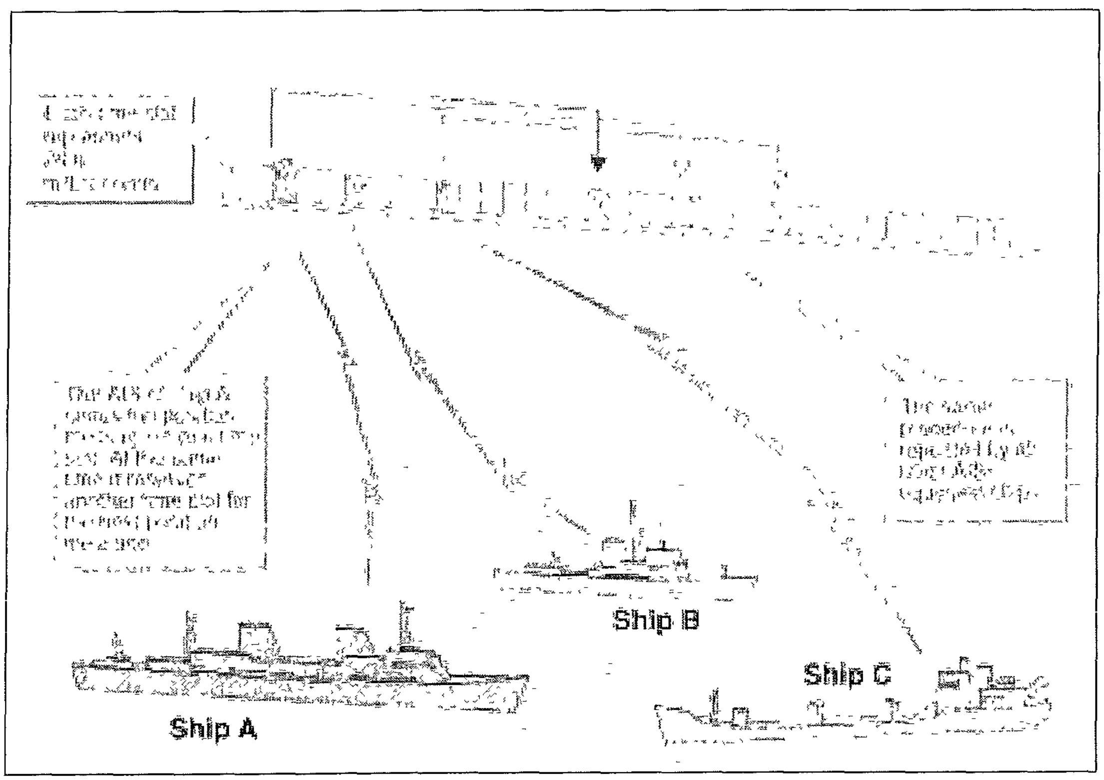


Figure 3-6 TDMA Technology (Source NAVCEN USCG)

frequencies with each band being shared in time between multiple users by dividing the band into 2250 time slots per minute with each time slot being 26.6 millisecond (ITU, 2014) (Figure 3-6). There are different types of transmission access technologies for different AIS units, but all transmission uses 9.6 kb GMSK (Gaussian Minimum Shift Keying) FM modulation over 25 or 12.5 kHz channels using High Level Data Link Control (HDLC) packet protocols (USCG NAVCEN, 2013). Most of the access scheme use a common time reference to accurately determine the start time of each TDMA slot series. The following is a list of TDMA technologies available and authorized by ITU that makes it possible for autonomous and continuous transmission of AIS messages:

- Self-Organized Time Division Multiple Access (SOTDMA): It is specifically designed for Class "A" AIS units that allows reserved transmission time slot for every unit within a common time zone (USCG NAVCEN, 2013) Every transmission uses a time slot indicator for its next transmission, hereby making a transmission map for each AIS unit. When moving from one area to another with different time slot allocation, the unit reconfigure their transmission map dynamically (ITU, 2014).
- Random Access Time Division Multiple Access (RATDMA): This technology is mainly used for Aids to Navigation (AtoN) by base station(s) to text message necessary information of a non-repeatable nature. Key elements of RATDMA share a common time reference, and randomly selects a current unused time slot to transmit. It is primarily used by Class A units for network entry; this is when an AIS unit is first turned on and has no allocated time slot using SOTDMA, therefore, an initial RATDMA transmission is used to acquire a time slot (ITU, 2014).

- Fixed Access Time Division Multiple Access (FATDMA): This type if acccss scheme is used by base station(s) configured to a particular time slot. This managed system has to be configured manually while installing the equipment so that the time slot is blocked and other AIS unit(s) cannot use this time slot.
- Carrier Sense Time Division Multiple Access (CSTDMA): Mainly defined for Class B AIS units that are required to listen for reservation messages (such as identification or safety related messages) and must comply with these messages. This technology acquires a time slot randomly using signal strength (background noise level) as criteria for its availability. If the signal strength measure is higher than the background level then the time slot is assumed to be in use by another AIS unit; if the signal strength is close or below the background level then it is assumed to be unused (ITU, 2014).
- Pre-Announced Time Division Multiple Access (PATDMA): This access scheme is used by devices (SART) that are capable of only transmitting messages. It uses the common time share to acquire a time slot that is randomly selected and announces its use of the time slot for the following 8 minutes, repeating the same process after the end of eight minutes.
- Incremental Time Division Multiple Access (ITDMA): This technology is used by base stations for non-periodic messages that could be either safety related, network entry when switching on the unit, or a change in reporting intervals of a periodic message by a vessel. It uses the common time reference and uses its internal slot map to select an unused slot in order to transmit a message.

The AIS message transmitted follows a specific format defined by the IMO. The message is an encoded binary text format that follows NMEA standards that the IMO adopted in conjunction with ITU. The binary text message gets further encoded in a 6 Bit ASCII before transmission. A general structure of an AIS message consists of a header and the data packet. The NMEA 0183 V 4.10 standards is comprised of two distinct codes or tags for receiving and transmitting AIS data. The received AIS data primarily from other vessels has a prefix code of *"! AIVDM"* and the AIS data being transmitted from a vessel is prefixed with "*! AIVDO*" (Arundale, 2014).

Sample AIS message: *"!AIVDM,1,1,,A,**133m@ogP00PD;88MD5MTDww@2D7k**,0*46"*

The ASCII format of AIS messages have been set and detailed in IEC-PAS 61162-100, "Maritime Navigation and Radio-communication Equipment and Systems" that adopts the NMEA 0183 V 4.10 standards. The above sample is of message type 1 that relates to position report of the vessel. Each component of the message is separated by a comma. The two 1's separated by the comma correspond to number of sentences and sentence number, respectively, in the message. The letter "A" signifies the channel used to transmit the message in this case Channel A (161.975 MHz). The next section of the message is comprised of 168 bits encoded as 6-bit ASCII AIS payload information that relates to the information regarding the vessel. The "0" signifies the end of the AIS message followed by the NMEA checksum field after the delimiter character "*". Most AIS equipment has inbuilt decoder to decode the received message before it is displayed or stored. Besides having ground based transmitters that aid vessels in navigating safely and avoiding collision, there is a nationwide network designed to increase Maritime Domain Awareness (MDA) using satellites.

3.6 Long Range Identification and Tracking

Tracking and monitoring of vessel close to the shore using AIS, as mandated by IMO is becoming more prevalent as many countries have adopted it to maintain safe navigation of vessels. Although AIS has its own limitations as per its range (20nm to 100nm) is concerned and it cannot effectively track vessels traversing beyond the coastal bound for non-US ports (Chen, 2014) (Cairns W. R., 2005). According to the US Coast Guards, there is around 5000 large sized vessels within 2000nm of the US at any time that not only represents economic loss, but also threatens US borders (Cairns W. R , 2005). Monitoring and tracking of these vessels becomes very important to safe guard the security of the US. To expand surveillance capabilities and to be able to monitor vessels at a distance, the USCG has adopted the Long-Range Identification and Tracking (LRIT) system as mandated by IMO for global identification of vessels. It is developed to collect and distribute information about vessel position received from IMO member States ships that are subjected to SOLAS (NAVCEN USCG, 2014).

A general LRIT system is comprised of shipborne LRIT information transmitting equipment, communication Service Providers, Application Service Providers, LRIT data centers, Vessel Monitoring Systems, LRIT Data Distribution Plans, and International LRIT data exchanges (IMO, 2014) (Figure 3-7). LRIT is a web-based system wherein a user can, in real time view or request vessel status, vessel information, vessel position, and regulate vessel reporting rates. The US National LRIT system obtains intelligence (human/machine) and it passes this information to the Command Center via Common Operation Picture (COP) where software(s) automatically analyzes the information and has the capability to alert the Command Center in case of a potential anomaly (Cairns W. R., 2005) (Figure 3-8). The USCG COP is integrated with other systems at the Federal, State and Local levels for sharing information with different agencies. To accomplish this effort the use of satellite and other existing technologies are being looked at to supplement the surveillance system.

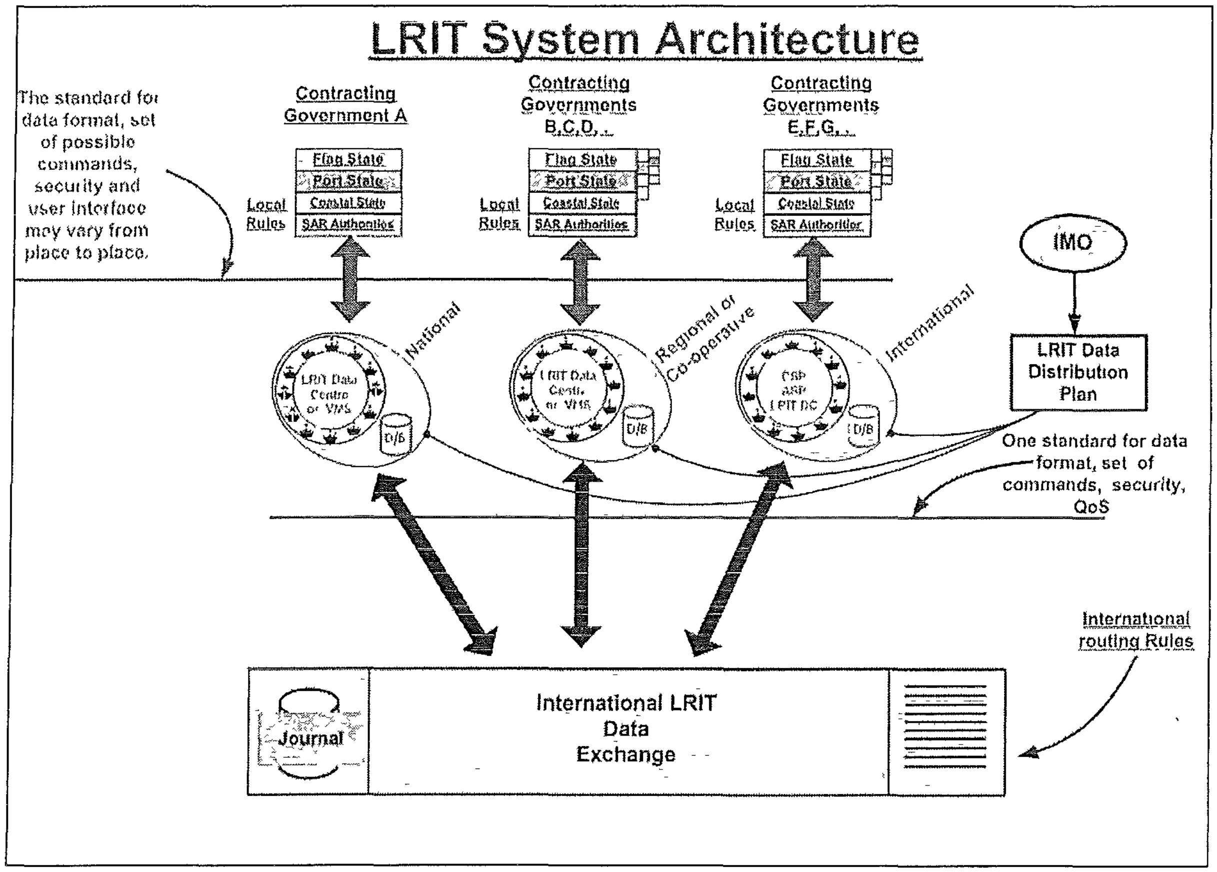


Figure 3-7 LRIT System Architecture (Source. IMO)

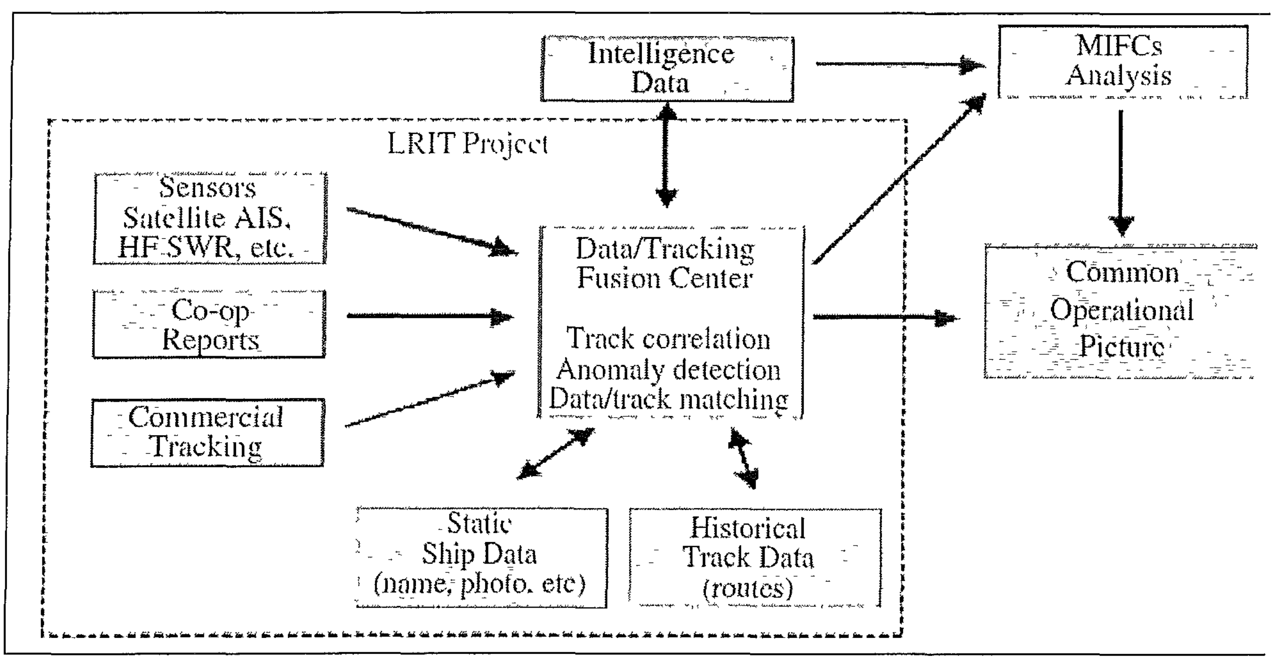


Figure 3-8 LRIT System National Architecture (Source. Cairns, 2005)

3.7 Satellite AIS

The discussion about the concept of using satellites for maritime surveillance started in 2003 with the Norewegian Defence Research Establishment, Forsvarets Forsklngistitutt (FII), presenting a paper at the fourth International Academy of Astronautics (IAA) Symposium on Small Satellites for Earth Observation (Hoye, Eriksen, Meland, & Narheim, 2008). Later in 2005, FFI presented their findings at the COMSAR annual meeting at London, whereby reporting possibility of the use of micro-satellite technology carrying passive sensors (Synthetic Aperture Radar Imagery and AIS message receiver) for ship detection (Carson-Jackson, 2012). Since then significant progress has been made in integrating AIS information capabilities with satellite's abilities for greater range to detect vessels.

The satellite-based AIS system is comprised of five components that include low earth orbiting satellites, ship-borne AIS equipment, Land Earth Stations (LES) to collect

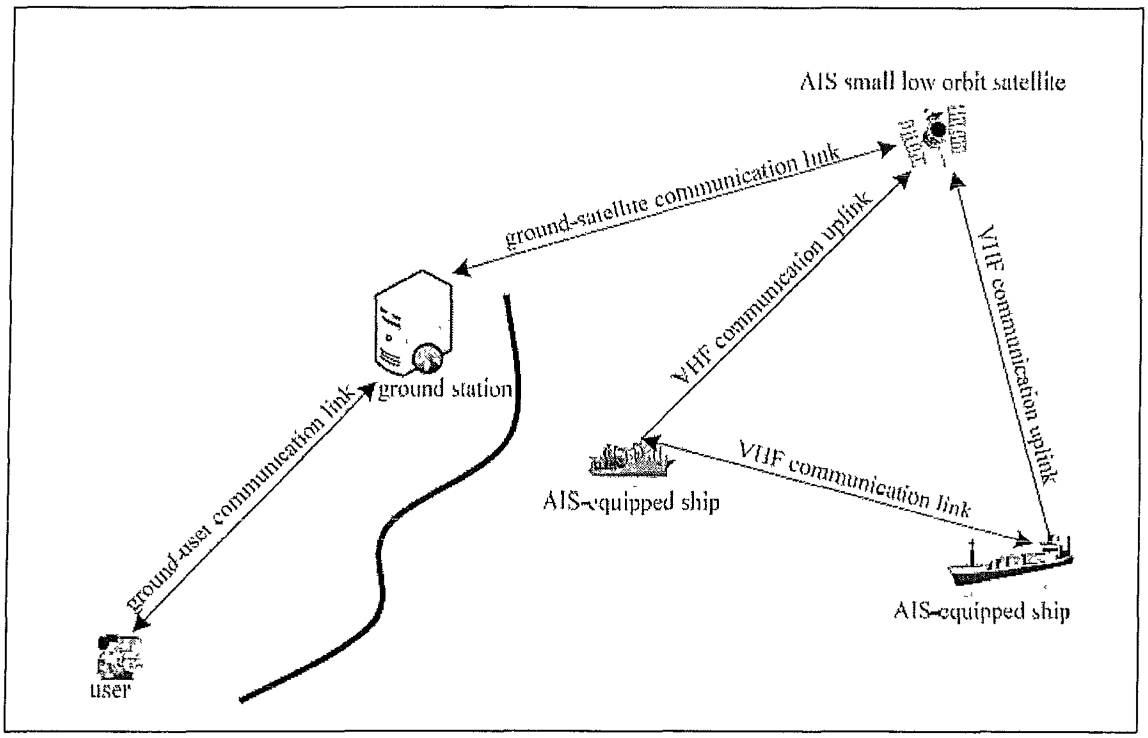


Figure 3-9 Satellite-based AIS Architecture (Source Chen, 2005)

data from satellites, users (e.g. VTS, USCG), and communication links for the user to LES (Chen, 2014) (Figure 3-9). The receiver on the satellite receives all the AIS messages broadcasted by AIS-equipped vessels. The satellite then transfers all the captured AIS messages to a LES when it comes close while orbiting. In its early stage of infancy the satellite system faced several issues that were solved or are in the process of becoming solved. Some of the issues or limitation that got highlighted by recent research are as follows:

- Orbit selection for satellite: Vertical reach of Class A AIS message is 1000km from the ground up (Chen, 2014). Recent research has shown that a single satellite orbiting on the Low Earth Orbit (LEO) can complete one revolution around the earth in 90 minutes which is quicker than if a satellite uses another orbit such as Geostationary Earth Orbit (GEO) or Medium Earth Orbit (MEO). Additionally, the distance from the surface of the earth is lesser (80 – 2000 km) when compared with GEO (35,786 km) and MEO (2000 – 35,786), thereby making it advantageous to capture AIS messages broadcasted by vessels (Carson-Jackson, 2012).
- Type of satellite: According to the FFI, small satellites would be ideal for maritime surveillance. A small satellite generally could weigh anywhere from 1kg to 100kg, with a lifespan between 5 to 20 years (Carson-Jackson, 2012). The issue here with small satellite is the limited power and data storage capabilities. With AIS-equipped vessel broadcasting messages at every 2 seconds, and with the satellite's capability to cover greater areas,

both of these factors easily lead to large amounts of data to be stored on the satellite before it reaches a LES to off load all the captured messages.

- Number of satellite: A satellite at LEO can provide information about a vessel with an update rate of 4 – 14 times a day (Hoye, Eriksen, Meland, & Narheim, 2008). A constellation of four satellites can provide a ship's information once every hour and twice if eight satellites are used. To globally monitor vessels, few satellites will definitely lead to time delay in updating vessels' information. Therefore, there is a need to have a substantial number of satellites to persistently monitor a region or area.
- Message Types: A "Class A" AIS unit broadcasts information every 2 seconds when cruising over 14 knot. Class A units have the capability to broadcast different message types as mandated by IMO. Different message types provide different information depending on the use, so the message size could vary from being too short to too long LEO satellites moves with a higher velocity to maintain their orbit that can easily miss the shorter AIS messages (Carson-Jackson, 2012). With different message types available, satellites are not configured for the newer message types.
- AIS Message Collision: Currently two dedicated VHF channels are used to broadcast AIS information using Time Division Multiple Access (TDMA) technology. Different message types are broadcasted via VHF channels that are captured by AIS stations at shore and in the vessels. Satellite capturing of these AIS messages can lead to overlap or message collision. The overlap occurs when AIS messages are sent by vessels

from adjacent time slots (SOTDMA Cell / Region), and these regions are well within the Field of View (FOV) of the satellite. The AIS messages received by the satellite tend to partially overlap due to similar time slot but different SOTDMA regions or different signal path lengths (Hoye, Eriksen, Meland, & Narheim, 2008). Figure 3-10 shows two vessels, ship

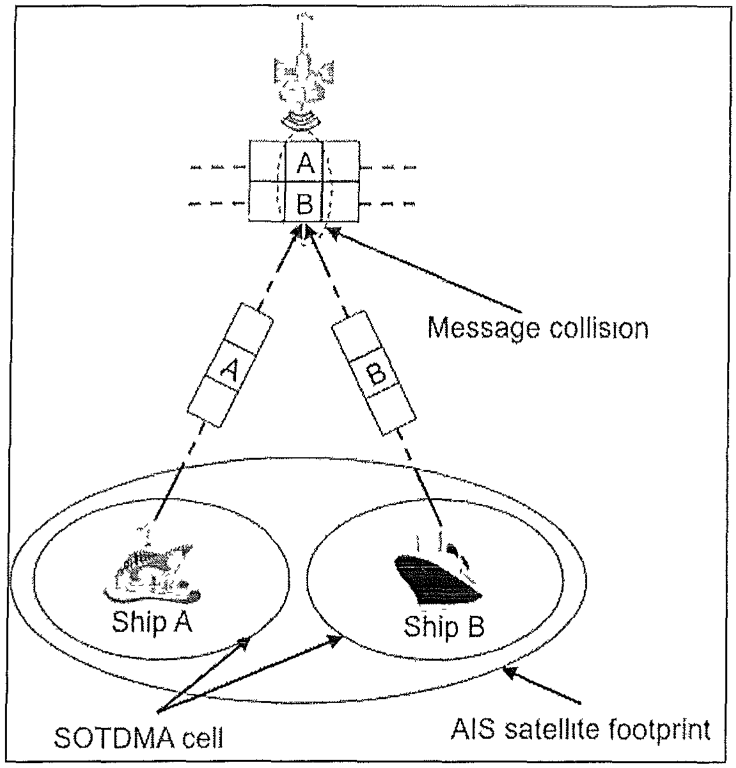


Figure 3-10 Message overlap due to similar time slot (Source Yuli, 2014)

A and ship B cruising within two different SOTDMA regions and both the regions are within the FOV of the satellite with an equal distance between the satellite and vessels. With both the vessels transmitting AIS messages using the same TDMA time slot but different SOTDMA region, the messages will be captured by the satellite at the same time slot, leading to message collision (Yuli, 2014).

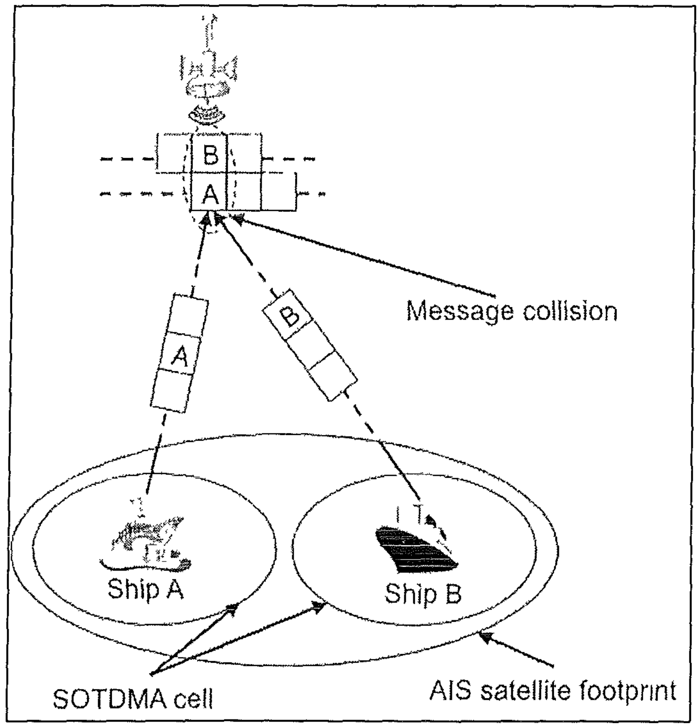


Figure 3-11 Message Collision Different Signal Path Length (Source Yuli, 2014)

Figure 3-11 exhibits AIS message collision, but the scenario is slightly different. In this case, the AIS message originates from different SOTDMA regions and with different time slots but are captured by the satellite at the same time causing data collision. In this case, the difference of transmission TDMA time slot is compensated by the difference of distance of each vessel from the satellite (Yuli, 2014).

- Other Technical and Environmental Issues: Research has shown that the characteristics of a ship's antenna can affect the transmission of AIS messages while communicating with a satellite (Cervera, Ginesi, & Eckstrin, 2011). Another issue with satellite AIS is the Doppler effect/shift where the frequency of the transmitted AIS wave changes due to the relative velocity of the vessel and the satellite (Burzigotti, Ginesi, &

Colavolpe, 2012). Environment can also play an important role in affecting transmitted AIS signals. When the linearly polarized AIS signal transmitted from a vessel passes through the ionosphere, the magnetic field in the medium and other factors changes the polarization angle of the signal, which is Faraday rotation (Cervera, Ginesi, & Eckstrin, 2011).

The above mentioned issues have a negative effect on the performance of the satellite AIS system. In recent years, advancement has been accomplished to either overcome the problem or identify different solutions. The report presented by the ITU recommends changes in some parameters to improve the performance of the satellite AIS system (ITU-R M.2169, 2010). The report addressed the technical limitations and proposed with some recommendations such as:

- A separate short AIS message (96 bits) type (Message type 27) with a reporting interval of 3 minutes can aid in solving data collision.
- Vessels within the range of terrestrial AIS base station should stop transmitting satellite AIS messages (i.e. Message Type 27). This will definitely decrease on AIS data size.
- Satellite AIS systems should only be limited to Class A. This will reduce the amount of AIS data as there is a significant growth in AIS users among the Class B category.
- Separate dedicated VHF channels (channels 75 and 76) are recommended for satellite AIS system.

- Revision of "Rec.ITU-R M.1371-3" to reflect the necessary changes so the AIS devices comply with these guidelines and be upgraded accordingly.

Almost all the recommendations put forth by the ITU were incorporated into the Rec. ITU-R M. 1371-3 in 2015. This correlated well with the timing of the initiative "e-Navigation", which is a global effort in collecting maritime vessel information for further analysis.

3.8 e-Navigation

In 2006, the IMO Maritime Safety Committee (MSC) included a new project of "e-Navigation" for its sub committees of Navigation (NAV) and Radiocommunications and Search and Rescue (COMSAR) (IMO, 2015). The goal was to develop a strategic vision to integrate new and existing electronic navigation in order to enhance navigation safety and reduce the burden on navigators. With all the navigation communication technology and services in their place, the effort was to integrate new and old technology to provide accurate, secure, and cost effective systems for global coverage of vessels (IMO, 2015). The sub-committees agreed that the concept of e-navigation should be based on users need and that all navigation equipment and software should comply with a set of performance standards. IMO's Maritime Safety Committee, in its report "MSC85-Report", provided guidelines to develop and implement the concept of e-Navigation (Amato, Fiorini, Gallone, & Golino, 2011).

Besides actively working on the vision and scope of e-Navigation, the groups were expeditiously working on other ancillary items, such as laying down the system architecture, identify components and limitation, benefits and obstacles, policy framework, and identifying roles and responsibilities of organizations and their member states. Since this initiative dealt mainly with navigation communication systems, the human element involved was addressed by the sub-committee of Human Element, Training, and Watch-keeping (HTW). The definition for e-Navigation, as coined by the sub-committee and agreed by IMO is provided below:

> *"e-Navigation is the harmonized collection, integration, exchange, presentation, and analysis of maritime information onboard and ashore by electronic means to enhance berth to berth navigation and related services, for safety and security at sea and protection of marine environment."*

From the US perspective, Committee of Marine Transportation System (CMTS), which is a federal inter-departmental committee headed by the Secretary of Transportation envisions e-Navigation as an integration of information and not just electronic equipment (Cairns B. , 2011). The vision of the US e-Navigation effort is as follows:

> *"To establish a framework that enables the transfer of data between and among ships and shore facilities, and that integrates and transforms that data into decision and action information."* (CMTS, 2012)

The vision adopted by IMO, or the one CMTS has coined, both meet the prerequisite of collection, integration, exchange, analysis, and standardization aspect of navigation communication for the purpose of enhancing maritime safety, security, and maritime awareness. The e-navigation system architecture (Figure 3-12) presents a conceptual and technical framework. The framework shows the entities within the ship and the shore environments along with the physical link(s) that interconnect them. The ship environment is comprised of the Integrated Navigation System (INS), the Integrated

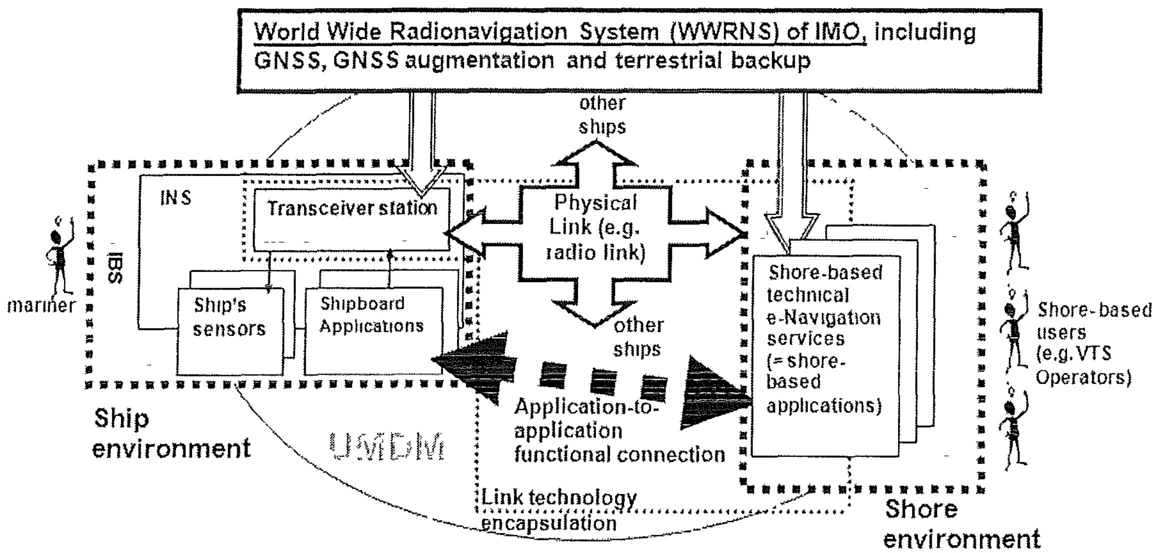


Figure 3-12 e-Navigation Architecture (Source IALA)

Bridge System (IBS), Transceiver stations, along with other sensors and on-board applications. The Shore side technical environment provides the interfaces to the shore-based application via the Physical Links in the form of services (Amato, Fiorini, Gallone, & Golino, 2011). The architecture also incorporates the World Wide Radio-Navigation System (WWRNS) as a system external to e-navigation assisting in providing accurate time and position (GNSS).

The e-Navigation initiative is a work in progress, and future addition and improvements are desired to attain better integration of systems, improved connectivity, seamless data exchange, human-focused interfaces, decision-focused information, and efficient inter-agency coordination and participation (CMTS, 2012). In a recent concluded fifth "e-Navigation Underway" conference held in January 2015 on board of a

DFDS ferry M/S PEARL SEAWAYS, received a number of new recommendations from its participants. Some of the new ideas were the concept of "Maritime Cloud", modernizing GMDSS, augmenting ECDIS, and Human Centered Designs, to name a few, that could make the system more robust and user friendly.

3.9 Application(s) of AIS

AIS, since its inception, is very aimed towards providing better and more accurate information about the vessel, its location, and other navigational dynamics. The benefits of AIS are not only designed to aid the maritime community or the government authorities by helping them to better track and monitor vessel(s), traffic control, and increase maritime situational awareness, but also to benefit the logistic industry by not only experiencing safe navigation through waterways, but by also cutting costs and increasing efficiency with timely, accurate information regarding berth availability at port. With accurate information about commercial vessel(s) location, the port operators can efficiently and effectively plan the use of its resource to benefit shippers by reducing wait time at the port, which translates to increasing efficiency at ports, thus improving the supply chain and savings to shippers.

Besides the improvement in the areas of situational awareness, navigational safety and improved maritime logistics, researchers and analysts have found profound use of AIS data. Below is a list of some of the applications of AIS data:

- Collision avoidance application to alert VTS personnel of potential collision of vessels within its surveillance region. The application uses fuzzy logic to predict the collision time and position (Kao, Lee, Chang, & Ko, 2007). Similar studies have been done to identify collision candidates in complex traffic patterns using Pendersen's model to calculate the causation probability of collision for a vessel (Silveira, Teixeira, & Soares, 2013)

- Interpolation of method is used to infer the physical path taken by the vessel by using two positional AIS report of that vessel (Peters & Hammond, 2011).
- Traffic analysis can be done using AIS data, to construct statistical characteristics of the port that provides information about distribution of transit location, termination and duration by vessel category, traffic type, and intensity of traffic (Calder & Schwehr, 2009).
- The use of AIS data to estimate emission in inland waterway by taking the vessel's physical characteristics and its speed over space and time (Perez, Chang, Billings, & Kosub, 2009).
- AIS data is used to quantify vessel utilization patterns in major navigation channels in south Louisiana, so in the future this information can be used to predict rates of vessel-induced channel widening and identify areas of concern (Kaiser, 2014).
- Vessel traffic flow velocity studies are conducted in certain waterways to determine the degree of vessel traffic congestion. AIS data is used to acquire the speed of vessels traversing through that region and analyze using a fuzzy inference model to determine the main traffic congestion degree under varying vessel traffic flow velocity (Hu, Yong, Shi, & Chen, 2010).

Besides some of the applications mentioned above, other potential uses of AIS data are listed below:

- Identify new location of new maritime activities
    - The location of the unknown activities can be discovered.
    - Abnormal activities like lightering can be discovered and monitored.
    - New Anchorage location can be identified as a Navigation Point of Interest (NPI).
    - New Dock's location can be identified and inventoried.

- Shipping companies are making efforts for fuel saving under the situation of high energy price Decision makers can take actions based on real-time AIS data, including the slow steaming, weather, and performance monitoring to apply the energy saving solution.
- Applying the AIS Data Archive to Answer Navigation Engineering Questions.
    - Identify channel routes usually used by vessels.
    - Measure the actual area of the channel for dredging purposes.
    - Identify the changes in draft at locations over time.
    - Vessel traffic separation by type of vessels.
    - Traffic flow by vessel type, speed, and time of day.
    - Improve maritime traffic management standards.
- Commodity flowing within a region can be accurately estimated using AIS data. Accurate shipping volumes (value, tons, and ton-miles) and shipping traffic patterns can be estimated for a region in order to:
    - Develop long range transportation plans.
    - Study site impact for future developments.

    - Evaluate benefits of intermodal transportation.
- Traditionally, data is collected based on survey data as well as mathematical models to estimate the O/D flow data; with the application of AIS, accurate O/D flow data can be collected for analysis automatically.
- Application of Intermodalism:
    - Simulate and estimate transportation costs by mode.
    - Estimate transit time of commodity movement more accurately.
    - Estimate fuel costs and consumption by mode and commodity.
    - Better management and efficient use of port assets.
    - Make port operations (AIS – ETA) more efficient by reducing waiting period for truck or rail.

AIS is a new technology, and over time, its use seems to benefit the maritime community immensely. Literature shows the development of AIS technology and how it is being widely used globally by commercial vessels to report its location and identity to other nearby vessels and authorities on shore. VTS authorities use AIS along with the old Radar technology to monitor vessels within its region. The entire objective of AIS is to promote safe navigation in waterways and avoid situation such as Exxon Valdez oil spill in 1989. With the inception of AIS, government agencies and other organizations have envisioned its use from a global perspective to monitor and track vessels engaged in maritime commerce. Literature also shows the initiatives and efforts made by these agencies to use existing technologies such as satellite to effectively monitor vessel movements in open water. Even though the use of satellite is not perfected, improvements and developments have been achieved over a period of time. Cooperation among participating countries and agencies has given rise to ideas and concepts such as LRIT and e-Navigation.

LRIT and e-Navigation provides an opportunity to integrate disparate data, specify data standards, and create services that the marine community can use. Literature has showed that the implementation of AIS is a slow process and the marine community is not trained, or they do not care. Initially, there were instances where vessels transmitted garbage data because either their device was not calibrated properly, or the user was not trained to handle the device. Also, with new applications are being developed, such as collision avoidance, to help vessel operators avoid collision and provide better situational awareness. With over 40,000 vessels engaged in maritime commerce and with widespread use of AIS on vessels, authorities are receiving wealth of information about vessel movements and activities.

AIS data shows considerable promise in tracking commodity flows through the system as well as documenting traffic volumes at important locations to identify areas that needs infrastructure development. AIS data, when collected over a period of time can be very overwhelming and a daunting task to manage it too. Literature has shown that researchers involved in working with AIS data do face similar issues of erroneous data and huge volumes of data. What is lacking is studies devoted to addressing the volume of AIS data and how to reduce it so it can be better managed. Also, the quality of the AIS data needs to be examined prior to reducing the data. As studies have shown most of the error within the AIS data is due to faulty calibration of the AIS unit, inputting erroneous information, or faulty installation with on-board sensors. This study is an effort to address the sheer volume of AIS data: its acquisition, storage, management, data-mining for errors, techniques to overcome errors, identify data coverage gaps, methodology to identify stop-events, methodology to reduce the volume of AIS data, and use the reduced data to detect stop-events, vessel path tracking, and detecting vessel call at specific terminals.

Chapter 4

AIS Data Analysis & Mining

This chapter describes the methods employed in this study to overcome the prevailing issues with the AIS technology. AIS being a relatively new technology, initial adaptation within the maritime community had its own issues. Mariners were not well trained to handle AIS equipment therefore issues such as faulty calibration of AIS units, inputting erroneous information, and use of ambiguous information is observed in the AIS data. Besides errors in AIS messages, issues of messages not being captured by any receiving stations pose a definite concern to safety. Another biggest concern is the sheer high volume of AIS data. Since, thousands of vessels is transmitting AIS messages repeatedly within seconds is overwhelming. Storage of these messages and performing analysis can be very challenging and time consuming. Also, understanding vessel's behavior in certainly segments of the waterway is also an issue and tracking the path of a vessel can be useful.

The area for this study is the Great Lakes region with particular interest in the commercial vessel activity therein. In terms of methods, this chapter describes an AIS data source: 1) data acquisition and storage, 2) data management, 3) conversion of data into geographical file format, 4) GIS analysis to identify data coverage gaps, 5) data-mining for errors, 6) detection and remediation of errors. The methods employed here will provide a thorough understanding of AIS data and the usability of the data after it has been reduced to a manageable amount.

4.1 Study Area

The effort here is to collect AIS broadcasts from commercial vessels within the Great Lakes region (Figure 4-1). There are government authorities, different agencies, and public or private organizations that collect AIS data for their own use. In this study the AIS data for the Great Lakes region is acquired from "Great Lakes and Seaway Shipping Online". This is a non-profit organization that is comprised of boat enthusiasts that keep track of vessels operating within the Great Lakes. They have AIS antennas installed on the shores, covering major parts of the Great Lakes that include Lake Superior, Lake Michigan, Lake St. Clair, Lake Erie, Lake Ottawa, St. Lawrence Seaway, and inland rivers (Figure 4-2).

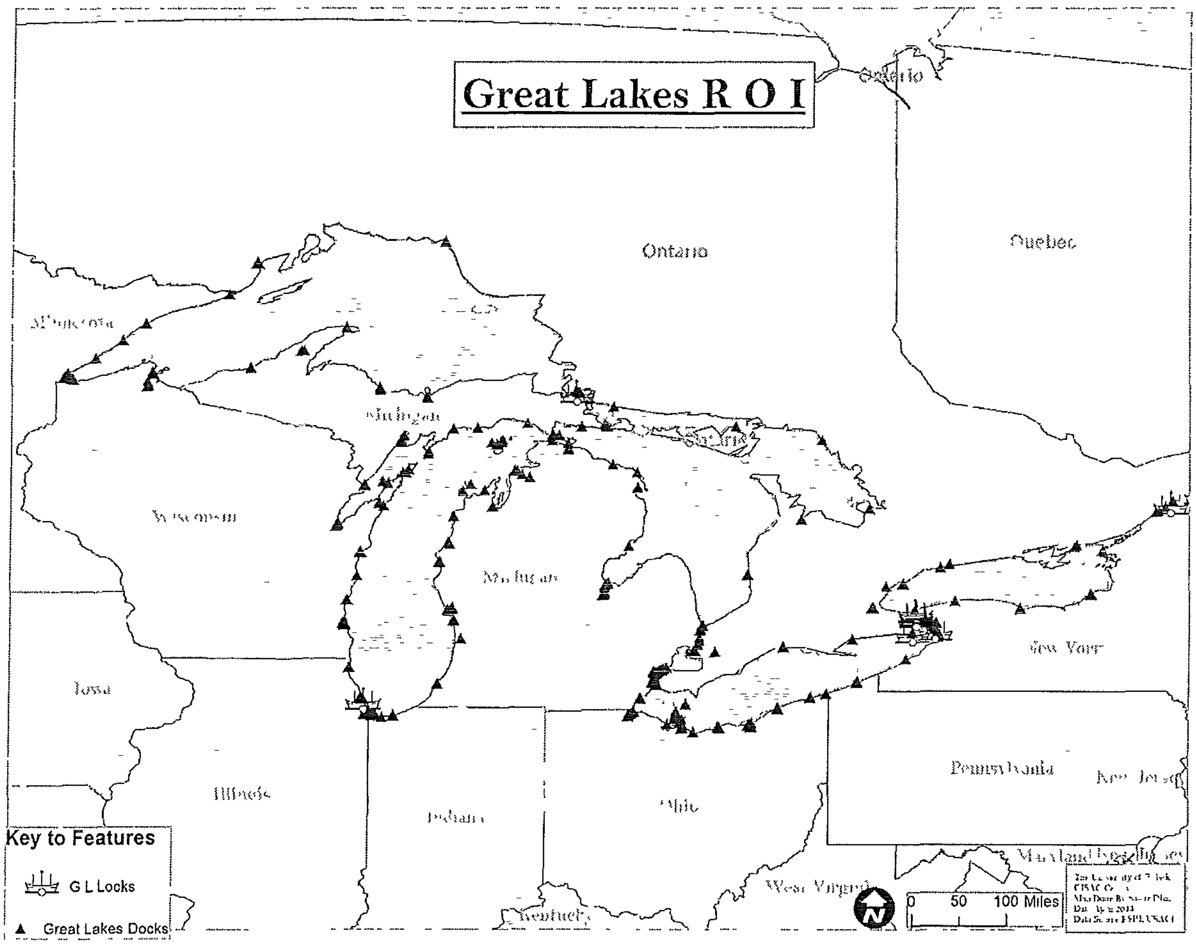


Figure 4-1 Great Lakes Region

Efforts are being made to install more antennas within the region to get comprehensive coverage by raising funds or by collaborating with other entities.

4.2 AIS Data Acquisition

The process of acquiring AIS data from "Great Lakes and Seaway Shipping Online" is an autonomous and automatic process The method employed here is to create a "Window's Service" that is then installed in a local server located in the GISAG Center at the University of Toledo. The "Great Lakes and Seaway Shipping Online" has a Web Server where they upload the AIS data for a five minute time period. This AIS data

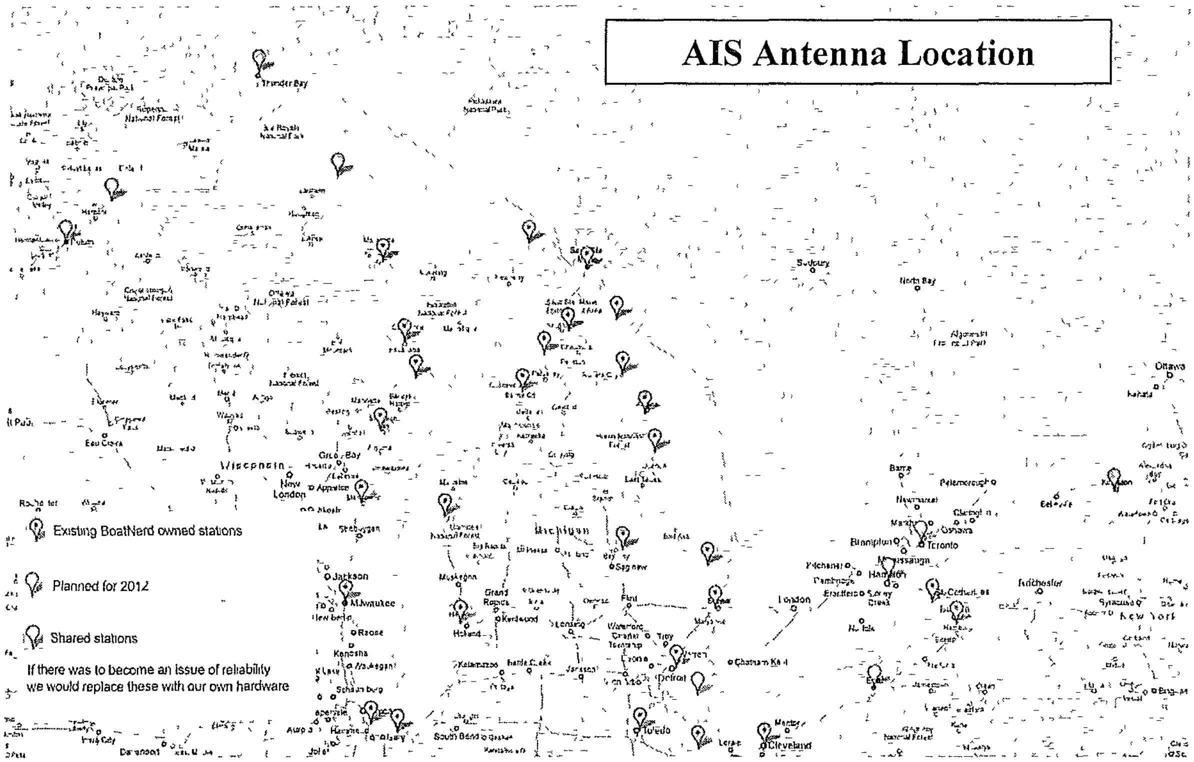


Figure 4-2 AIS Antenna Location (Courtesy BoatNerd)

stream for five minutes is automatically decoded and parsed before it is saved in a Text File format. The "Window's Service" at the GISAG Center accesses the Web Server at the "Great Lakes and Seaway Shipping Online" facility and grabs the latest Text file that contains the AIS data stream. The script then obtains the file and interprets information line by line, storing the data in a SQL Server that contains a database completely dedicated for storage.

The Window's Service fires every five minutes and the entire process of accessing the Web Server and fetching the latest Text file is done automatically. Each file received contains roughly around 5000 or more records. The AIS information received contains a composite of three message types viz. static, dynamic, and voyage related information. Table 4.1 gives a list of all the data items are in the Text file that is received. The AIS information received has data broadcasted by each and every vessel type that is required to have an AIS device installed on board. AIS information from commercial, recreation, and other types of vessels is acquired. With information being acquired from every vessel type within the proximity of the receiving antennas, the size of the data-table tends to increase phenomenally.

Data Documentation AIS	
ITEM	DESCRIPTION
Id	Unique Identification of record
Mmsi	Maritime Mobile Service Identity Set on installation - note that this might need amending if the ship changes ownership
Status	0 = under way using engine,
	1 = at anchor,
	2 = not under command,
	3 = restricted manoeuvrability,
	4 = constrained by her draught,
	5 = moored,
	6 = aground,
	7 = engaged in fishing,
	8 = under way using engine,
	9 = reserved for future amendment of Navigational Status for HSC,
	10 = reserved for future amendment of Navigational Status for WIG,
	11 - 14 = reserved for future use, 15 = not defined = default
Type	Type of ship/cargo
Lat	Latitude - to 1/10000 minute
Lon	Longitude - to 1/10000 minute
speed	Speed over ground - 0 1-knot (0 19 km/h) resolution from 0 to 102 knots (189 km/h)
course	Course over ground - relative to true north to 0 1 degree
heading	True Heading - 0 to 359 degrees from eg gyro compass
draft	Draught of ship - 0 1 meter to 25 5 meters
width	Width of ship - to nearest meter
length	Length of ship - to nearest meter
name	Name - 20 characters to represent the name of the vessel
call	Radio call sign - international radio call sign, up to seven characters, assigned to the vessel by its country of registry
dest	Destination - max 20 characters
eta	ETA (estimated time of arrival) at destination - UTC month/date hour minute
reg	Registration Number
imo	IMO ship identification number - a 7 digit number that is unchanged upon transfer of the ship's registration to another country
timestamp	The universal time to nearest second that this information was generated*

Table 4 1 AIS Data Documentation

Year	Total	Commercial	% Commercial	Stop-Events
2010	23,680,761	14,724,854	62 18	5,864,724
2011	40,948,632	23,506,497	57 40	10,286,244
2012	64,896,614	36,386,660	56 07	16,332,033
	129,526,007			**32,523,438**

Table 4 2 Raw AIS Data Count

4.3 AIS Data and Data Analysis

Table 4.2 shows the annual totals of AIS record counts in the database. There is a definite increase in data records every year after 2010, and that is attributed to installation of more antennas around the Great Lakes. Commercial vessels make up more than 50 percent of the total AIS data records, where the remaining records are either recreational or vessels used by US Coast Guards. Stop-events are events that a vessel exhibits when it is not in motion since the vessel is either at berth or is waiting nearby a port waiting for an available berth. The criteria to identify a stop-event is evaluating the value of the AIS record's "Speed" data variable.

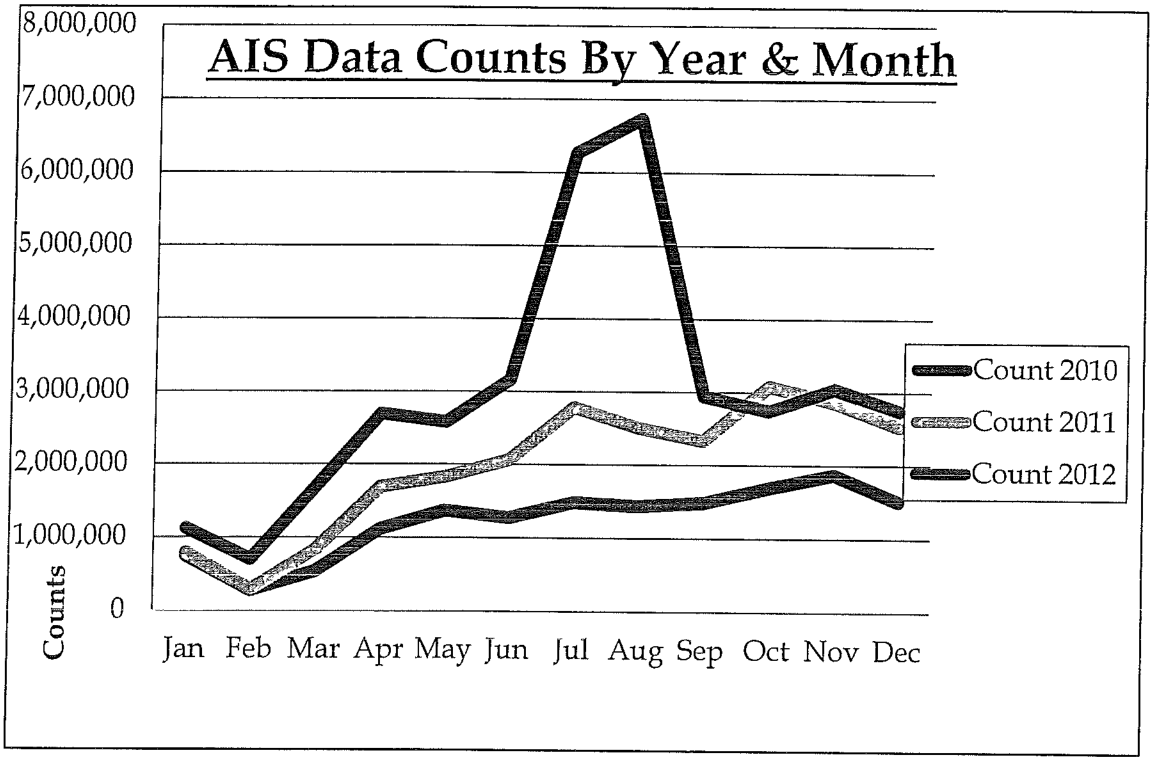
After consultation with personnel from USACE and USCG a value of the "Speed" data variable equal to 0.1 knots or less indicates the stop-event for that vessel. Also, the data variable "Status" should be able to identify the stop-events, and this is a value that the vessel operator has to manually enter. Most of the time, the operators are too busy with other activities and are unable to update the vessel's status. Efforts are being made to link data variables, such as when the speed of the vessel is 0.1 knots or less the value of the variable "Status" should automatically change to "0 = At Anchor," or if the vessel is cruising at a speed of 14 knots and over the status value should be updated to "8 = Under way using engine".

In this study the AIS data is stored in a tabular format that can be queried and examined using Structured Query Language (SQL). Chunks of data can be queried using SQL and imported to an Excel sheet for further analysis. Since the AIS data has location variable (Latitude and Longitude) data it can also be imported into a Geo-Database for it to be transformed into a geographical file format (Shapefile). Geographical files of AIS data can be visually examined using ESRI ArcMap. Initial exploration of the AIS data showed high amounts of activity during the months of July and August (Table 4.3 and Graph 4.1).

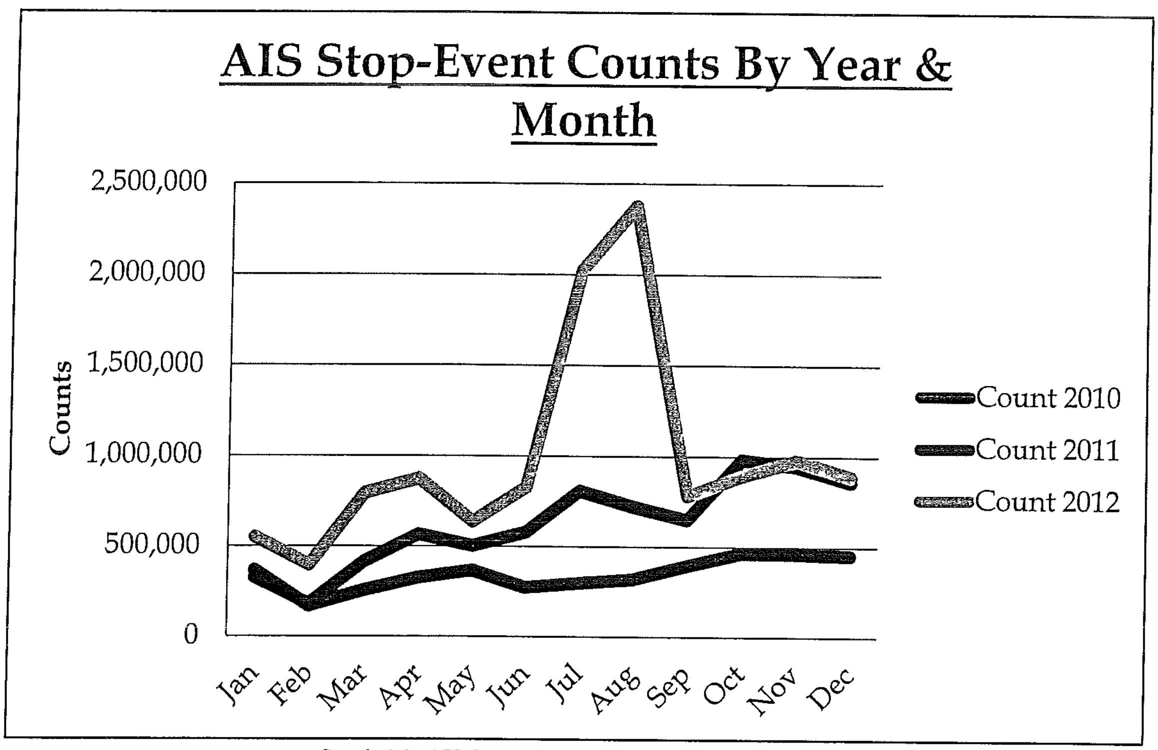
Month	Count 2010	Count 2011	Count 2012
Jan	747,744	790,732	1,118,114
Feb	278,320	288,604	700,241
Mar	535,210	832,243	1,684,007
Apr	1,113,324	1,687,929	2,689,628
May	1,367,239	1,826,914	2,577,397
Jun	1,271,831	2,061,754	3,155,059
Jul	1,480,056	2,782,450	6,265,702
Aug	1,430,716	2,500,948	6,733,851
Sep	1,480,326	2,326,782	2,937,515
Oct	1,684,851	3,078,440	2,737,760
Nov	1,869,872	2,854,881	3,043,715
Dec	1,504,454	2,525,310	2,743,671

Table 4 3 AIS Data Counts by Month and Year

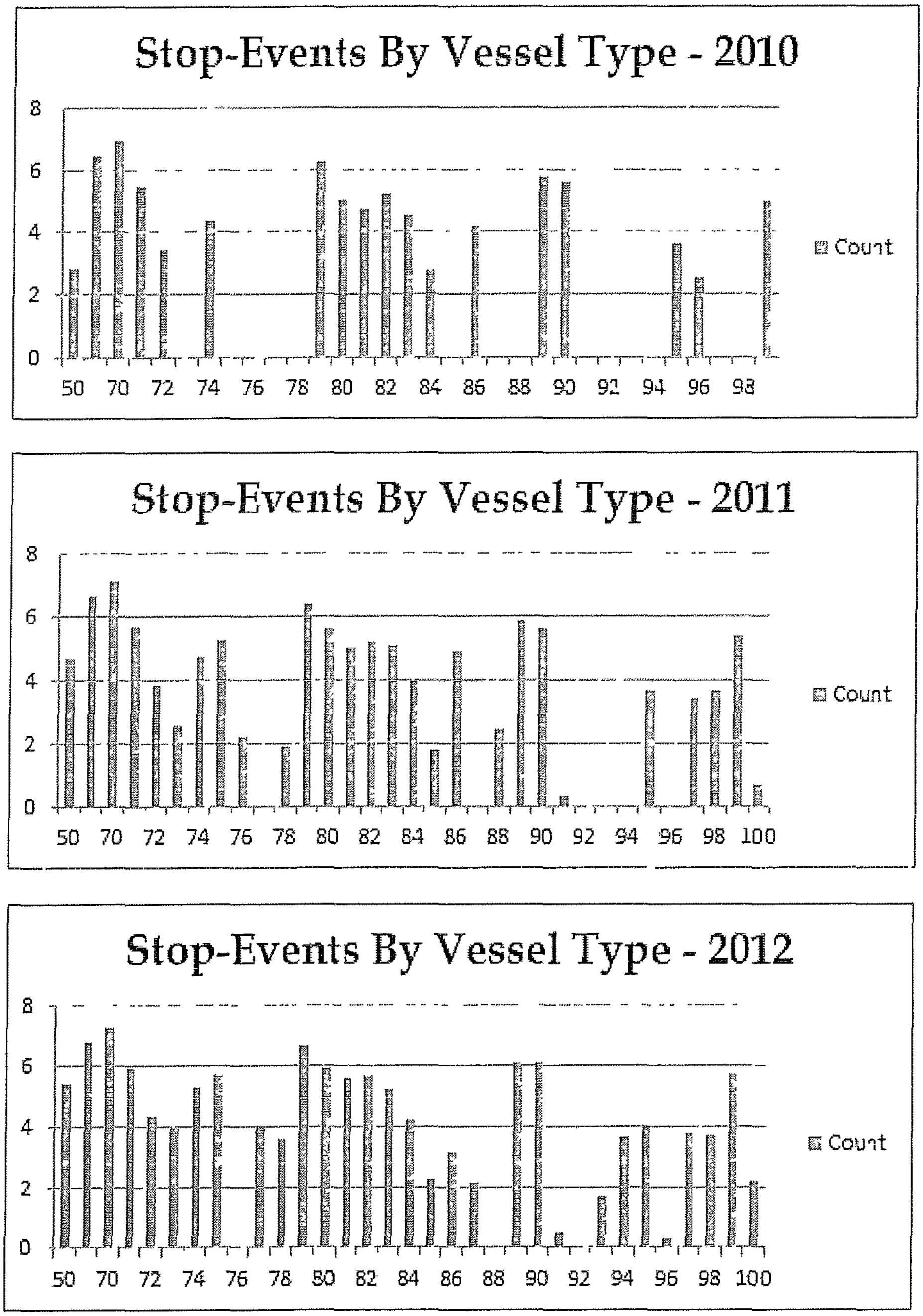
Stop-Events were identified for every month for the years 2010 -2012 (Graph 4.2). The stop-events were also categorized by the type of vessel. Graph 4.3 shows the majority of vessel types that were active within the GL region include Pilot vessels, Tug boats, Bulk cargo vessels, and Tankers for the year 2011 and 2012. In 2010, the stop-events by vessel type reveals less information due to fewer antennas being installed and, also, some of the AIS devices were not calibrated to broadcast accurate information.



Graph 4.1. AIS Data Counts by Year & Month



Graph 4 2 AIS Stop-Events Counts by Year & Month



Graph 4 3· Stop-Events by Vessel Type (2010-2012)

4.4 Geo-Visualization of AIS Data

The AIS data has spatial (Latitude and Longitude) and temporal (Time) information associated with each and every record. Here small chunks of data are queried from the SQL Server and imported directly into a geo-database using SQL. ArcMap 10.x is used to convert the selected tabular AIS data into a geographical file format. Initial visual inspection of the AIS data indicates a very busy pattern, emphasizing abundance of maritime activity in the region over a period of time (Figure 4-3). The observed pattern is created by AIS-Vessel relaying information periodically, every 2 or more seconds depending on speed or course. Also our initial visual examination of the map, which is AIS data for the month of June 2011, reveals areas where AIS coverage was inadequate. Also, some AIS data points are on land, indicating

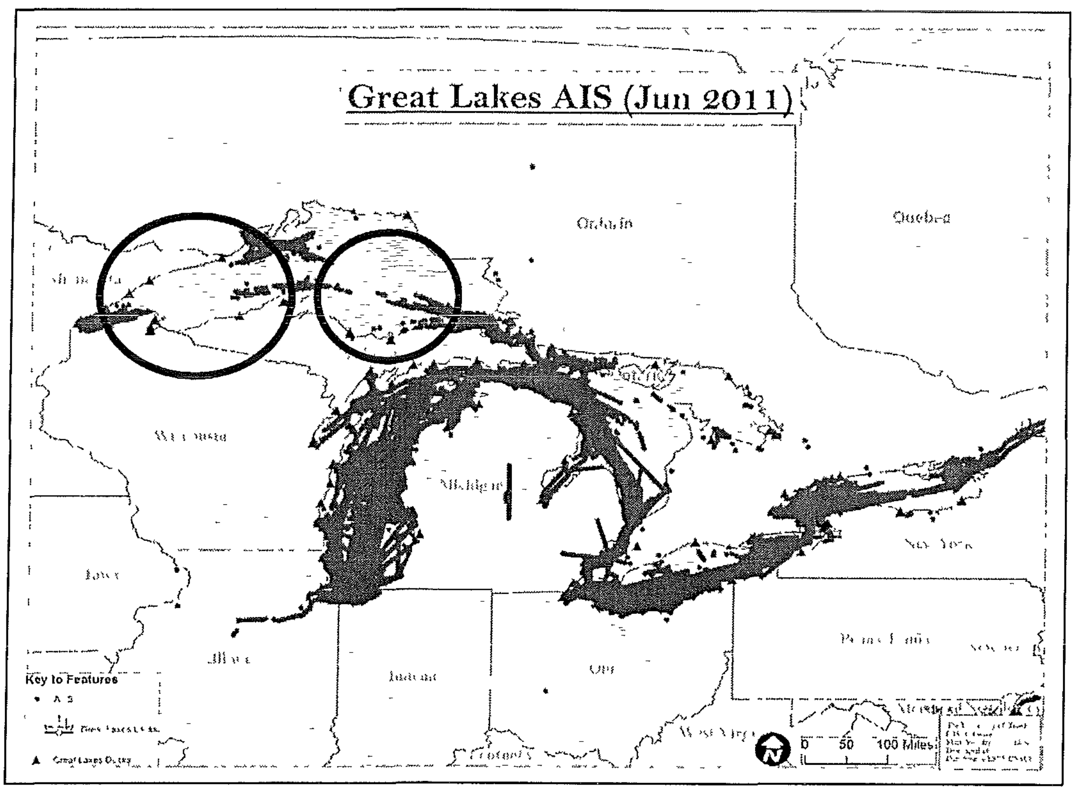


Figure 4-3 AIS Data (Shapefile format)

that the AIS device is not calibrated correctly and the data broadcasted has errors associated with it.

4.5 AIS Data-Mining for Errors

AIS data is inspected for erroneous information. Initial findings reveals several issues within AIS data. Majority of errors relate to vessel operators relaying erroneous information about its identity, or sometimes incomplete information. All the static information relating to the Vessel's identity and characteristics needs to be manually fed into the AIS unit during calibration. Several records exhibited spelling errors, incomplete name of vessel, or incorrect official vessel identification number. Faulty calibration of AIS units along with fat fingering faulty information seems like a plausible cause for such erroneous data.

Vessel Name	Call Sign	IMO #	V Type
ALBATROSS I V	HP6761	7338690	70
ALBATROSS I V	WDF9469	7338690	70
ALGOCAPE	VGJC	6703214	79
ALGOCAPE	VGJC	6226030	70
ALGOSOO	VGJD	7343619	79
ALGOSOO	VGJD	8119273	70
ALGOSOO	VGJD	7634288	79
ALGOWOOD	VCTD	7910216	70
ALGOWOOD		7910216	70
ANGLIAN LADY	VOLP	5141483	52
ANGLIAN LADY	VN-X0	5141483	52
ATLANTIC RAVEN	8PAF6	9187942	52
ATLANTIC RAVEN	CFN5827	835842	99
ALGONOVA	CFN5191	9378589	83
ALGONOVA	CFN5191	9378589	84
ALGONOVA	CFN5191	9378589	81
ALGONOVA	CFN5191	9378589	89
ALGONOVA	CFN4191	9378589	89
ALGONOVA	CFN5191	9378589	82
BLOCK	WXY6216	574870	71

MMSI	IMO
212091000	9436460
212093000	0
212093000	9409742
212156000	9356907
212156000	0
212170000	0
212170000	9271511
212177000	0
212177000	9346433
212215000	9459981
212215000	0
212337000	9393149
212337000	0
212340000	0
212340000	9413913
212372000	9393151
212380000	0
212380000	9314820
212425000	9133769
212425000	0

Table 4 4: AIS Data Errors

id	mmsi	mtime	type	lat	lon	course	heading	draft	timestamp
70415967	3160048	1309648339	130	43 1389	-79 1958	0	0	0	7/2/11 11 12 PM
70415982	3160048	1309648339	130	43 1389	-79.1958	0	0	0	7/2/11 11 12 PM
70737914	3160048	1309830591	130	43.1389	-79 1958	0	0	0	7/5/11 1 49 AM
70737929	3160048	1309830591	130	43 1389	-79.1958	0	0	0	7/5/11 1 49 AM
70170021	3160051	1309498361	130	42 0695	-82 6662	0	0	0	7/1/11 5 32 AM
70170029	3160051	1309498361	130	42 0695	-82 6662	0	0	0	7/1/11 5 32 AM

Table 4 5 Duplicate Records in AIS Data

Occurrences of similar MMSI associated with different vessel identification numbers were also discovered (Table 4.4). This suggests the possibility of the AIS unit being used for more than one vessel. Instances of duplicate or similar records is identified so that those records can be rectified. Duplication of similar AIS feeds usually happen if

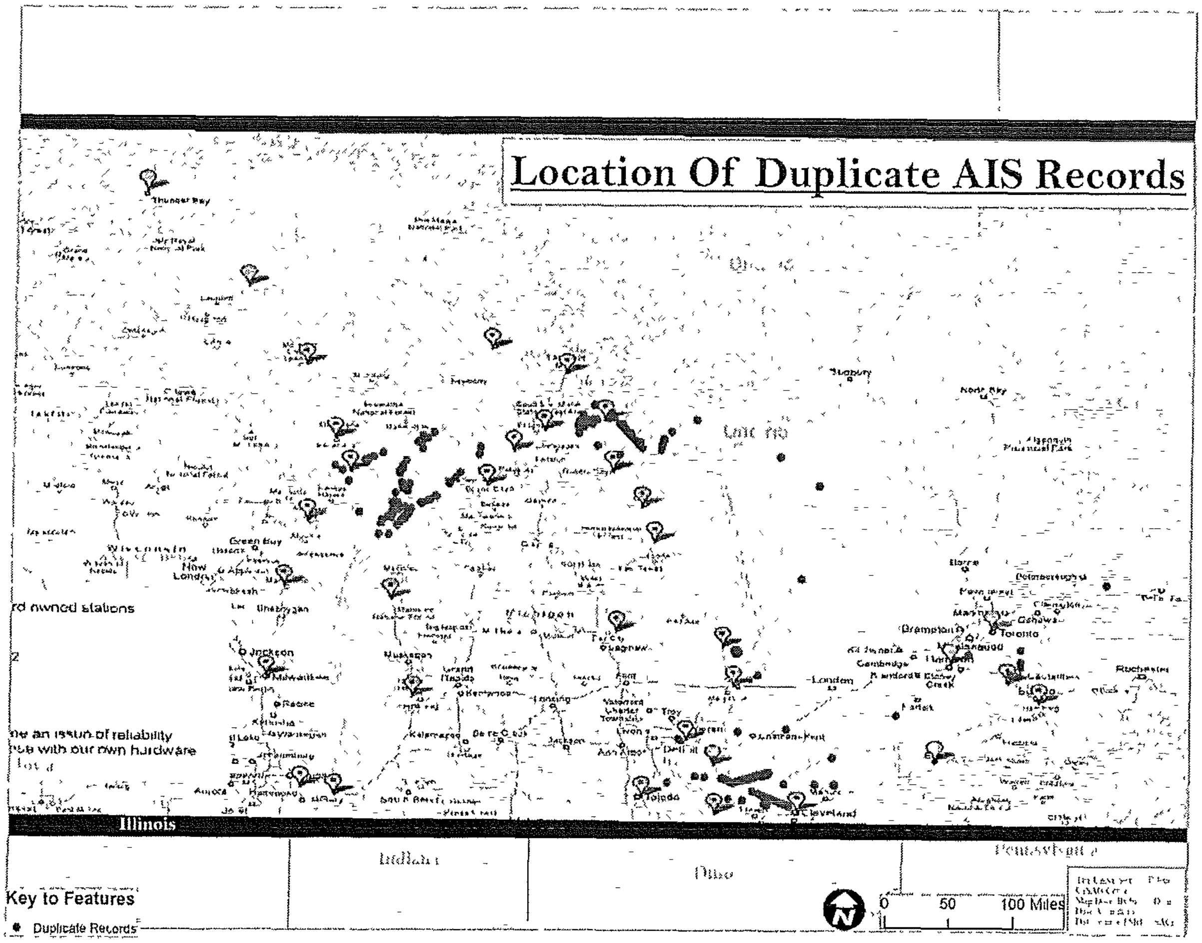


Figure 4-4 Mapping Duplicate AIS Records

the vessel(s) are within range of one or more receiver(s) or antenna(s) (Table 4.5). These duplicate records is selected and mapped in ArcGIS to identify location(s) of these AIS feeds (Figure 4-4).

Another issue of Gaps in AIS coverage is identified visually by mapping the AIS data. Gaps in data suggest that the coverage in certain area is inadequate (Figure 4-3)

Several records were identified that contain gibberish information (Table 4.6). This is typical of either the AIS device not being calibrated at all, or, the information entered manually was erroneous.

Vessel Name	**Call Sign**
	WZ3378
- %YG H;NN PATHFNDR	.S!FM?1
!IHE BCUP	H IH A*
#Z N E BLOCK	9 F^48
(^*^)AMERICA EXPRESS	9VJJ8
) WX X P A BU:	L) 2 0
*#YUM O	B ^OLD0
,RPN 0XP	WMTFE
.3LB3 N8 ._V!WLC#	W 7 PB
/ # DF HOP<AND	P.A)
////U8 AP) CHAMPLAI	A// F
: X TG	navaid
_ _ _OD A SYL_ZX;:	WLTVY%G
_> U(WX:ZUM HA2H 0	navaid
_I0KT / FD ASHER	BCL_GL
0 4 <#L1;04/ ,*9D X	navaid
1 3?% FA%QSIC	94DXR5:
2 AA PA !A /	XD P
2*ZHR DE_	W6) / 7
4HDDT D 5SLARKE	/ BDPFA

Table 4 6 Garbage Data

Many of these highlighted errors can be categorized as faulty calibration of an AIS device, data input errors of vessel's information, and some purely memory slips or negligence of timely data input (Harati-Mokhtari, Wall, Brooks, & Wang, 2007). Ancillary data sources such as the Vessel database from the USCG or IMO is used to solve issues regarding to vessel's identity, such as the IMO number or other identification number and the proper name of the vessel. Examples of missing values such as speed, course, heading, and draft, as seen in Table 17 is due to faulty calibration of the AIS device and can be fixed with proper connection of the AIS device to the vessel's sensors on board. Most of the errors occur due to inadequate training of navigators and operators, leading to discouragement with using the system (Harati-Mokhtari, Wall, Brooks, & Wang, 2007) With the volume of acquired AIS data the rate of errors not corrected can become very high and lead to inaccurate analysis. Error detection in the early stages is of important and officials should promptly take proper corrective actions in notifying vessel operators. Also, high volume of AIS data makes processing of AIS data, or even performing any kind of analysis, very time consuming or the software is not capable to handle huge data volumes (i.e. Big Data). In the following section a methodology is explained where high volume AIS data is distilled using multiple variable criteria. Methodology for identifying stop-events is developed to aid in analyzing events such as vessel calls or any other activities besides vessel call.

Chapter 5

AIS Data Reduction

5.1 AIS Data Reduction Technique

As seen in Table , AIS record count can be 64 million and over, and as more vessels install AIS devices the count of AIS records will only increase. Consequently, the effort here is to reduce the volume (number of records) of the raw AIS data (Big Data) without compromising pertinent information. The initial step is to first identify few variables (attributes) within the data that can be used to reduce the raw AIS data. A vessel broadcasts AIS message within certain time intervals that are based on the speed of the vessel and its course. The dynamic part of the AIS message changes with time and the remaining part of the messages (static or voyage) remain unchanged for a vessel on a particular voyage. Thorough observation reveals that within the dynamic part of the message the attributes that constantly keep changing over time are the vessel's location and its timestamp. The remaining part of the dynamic message changes over a period of time with respect to destination, depth of channel, and type of waterway (e.g. river, lakes, and ocean). Emphasis is on those variables that change within every other broadcast of AIS messages such as the location (latitude and Longitude) attribute along with speed and heading information. The location attribute, along with the vessel's heading information, shows if a vessel is traveling in a straight line or if it is changing its direction/course. This helps in documenting the path a vessel takes to traverse from point of origin to the point of destination. Location information, along with speed of the vessel, also highlights events such as stop-event, where the vessel is docked at a terminal over a period of time with its speed being 0.1 knots or lower.

For the data reduction analysis, a sample data of 85,093 records comprising AIS information from few vessels for a small period of time is used (Figure 5.1). The algorithm is coined within the MS SQL Server 2005 framework. The algorithm is coded using the

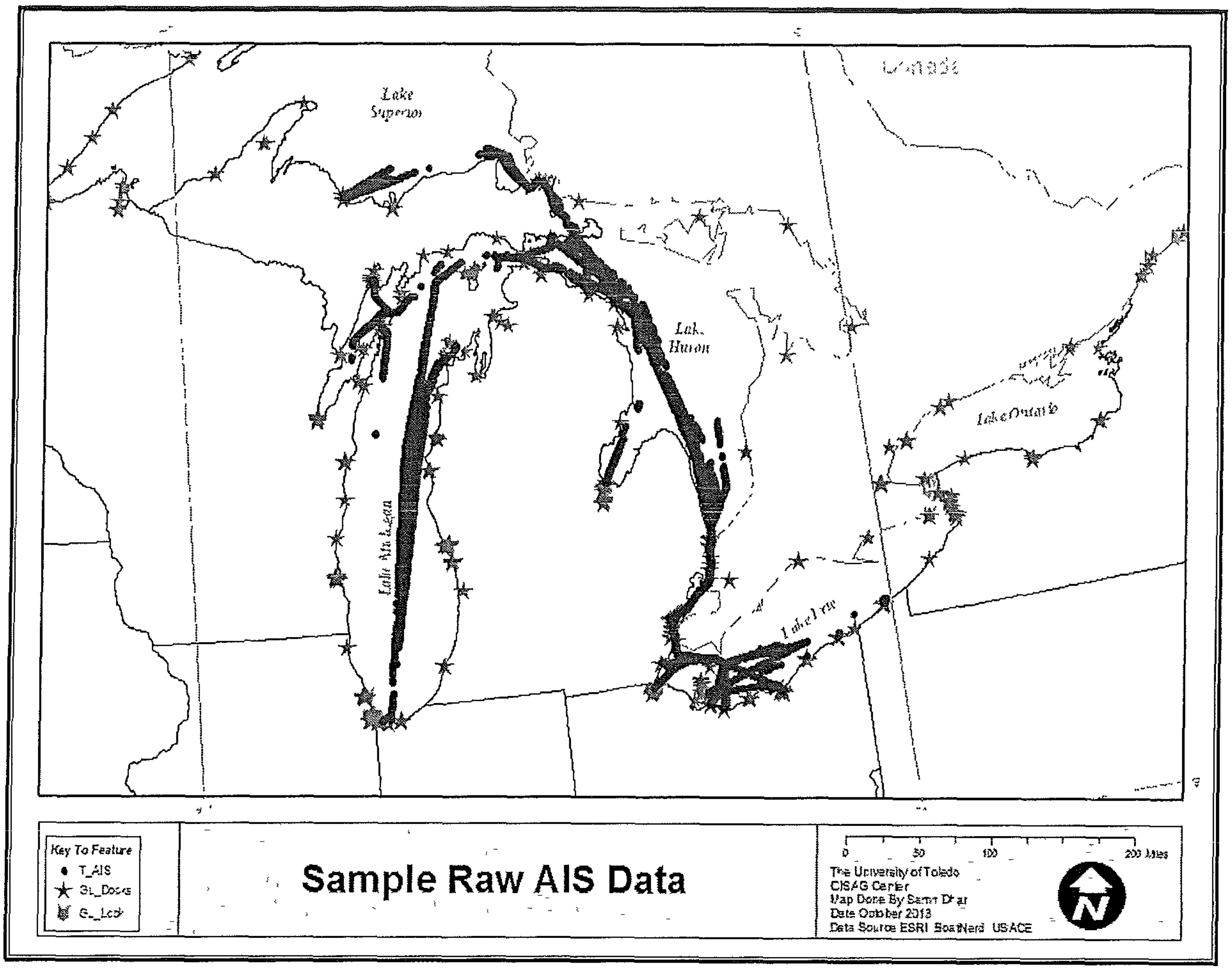


Figure 5.1 Sample Raw AIS Data

procedural language Transactional Structured Query Language (T-SQL) in a form of a procedure that is executed at the server's end. Figure 5-2 is the pseudo code of the algorithm that is developed to reduce the volume of the raw AIS data to a manageable form. Each and every record is processed individually and compared with a set of variables to assess if the current record under process should be kept or discarded. The very first record that is processed gets automatically assigned to be the initial values for the comparative variable purposes.

In this Big Data reduction algorithm, none of the records is deleted from the database but is flagged accordingly of it being relevant or irrelevant. The algorithm also checks for duplicate records and flags them for future removal. The tolerance value of the variables including speed and heading is varied systematically, and a series of tests is conducted to ascertain the optimal combination that results with a maximum number of reduction of records without compromising important information.

5.2 AIS Distill Pseudo Code

The procedure initially selects every individual unique vessel's MMSI value from the sample data and stores it in an array. These MMSI value is processed one by one to select records. The procedure then makes a call to an outside sub-procedure that selects all records unique to the input MMSI value. The selected records is stored in a temporary table in a chronological order with the latest record time-wise being towards the end. The procedure loops through each and every record sequentially from the temporary table until all the records are processed. The first record is used to initialize the value of comparative variable and is flagged "Y" denoting it as a relevant record. The next record is processed where the value of the variables (Lat, Long, Speed, Heading, Time-Stamp) is compared with the comparative variables.

The values of the variables (Lat, Long, and Time-Stamp) is compared with the values of the comparative variables. If the values of variables of the record in process is similar to the values of the comparative variables, then that record is flagged "N" which denoted irrelevant record or possibility of duplicate process. Now if the value of the variables between the record in process and the comparative variables differs then the difference in values is tested further using the variable (Speed and Heading) of the record in process is evaluated against it comparative variable If the difference in absolute value is below the preset tolerance level, the record in process is flagged "Y" and then the comparative value is re-initialized with the value of the record in process. This process loops through every record in the temporary table until all the records are processed. Then all the AIS records relating the next MMSI value in the array is selected in the temporary table and the entire process of flagging each record is repeated again until all the records are processed.

```
Pseudo Code --Procedure AIS_Distill Begins
Create a Cursor to Store Unique MMSI Selected from the AIS Table
Declare variables to hold attribute(s) values
Open cursor_MMSI
Repeat
        Call Sub-Procedure to select all records unique to MMSI being processed
        GO through each AIS records from a temporary table
          Repeat
                Read the next record's attributes into variables
                IF the current record is the first record then
                        Initialize the Initial Values of comparing Variables
                        Update AIS table; flag this record as Selected or Marked as "Y"
                EndIF
                        IF current record's lat. long, and time-stamp is equal to comparative variables
                        Update AIS table, mark the current record as duplicate
                EndIF
                        IF current record's lat and long not equal to comparative variables
                        IF current record's speed and heading not equal to comparative variables
                                IF absolute difference between the current record's (speed, heading) and
                                        The comparative variables is >= the specified level then
                                        Update AIS table; flag the current record as 'Y'
                                        Re-Initialize the Values of comparative Variables with current
record
                                EndIF
                                EndIF
                EndIF
                Until all records pertaining to the unique MMSI are processed
                Empty temp tables
                Fetch Next MMSI
Until all unique MMSI are processed
Close Cursor
Release Resource
End Procedure AIS_Distill
```

Figure 5-2. Pseudo Code AIS Distill

5.3 Results

The procedure mentioned above is used to conduct series of test with every test run having a different tolerance level for the variables involved. At every test run one of the variable's (Speed and Heading) tolerance value is held constant while the value of the other is varied. This provides the optimum value for the tolerance level for the variable used in the reduction process. It is observed that in during tweaking the value of the tolerance level the percentage of reduction of records differed. With a higher tolerance level the percentage of reduction of record count increased.

Initially the tolerance level of the "Heading" variable is held constant while varying the "Speed" variable. This allows to ascertain the optimum value to be used for the "Speed" variable Results after several series of test indicates that the value of "15" is the optimal for the "Speed" tolerance level (Table 5.1). Any higher value for the speed provides no further significant increase in reduction of records count.

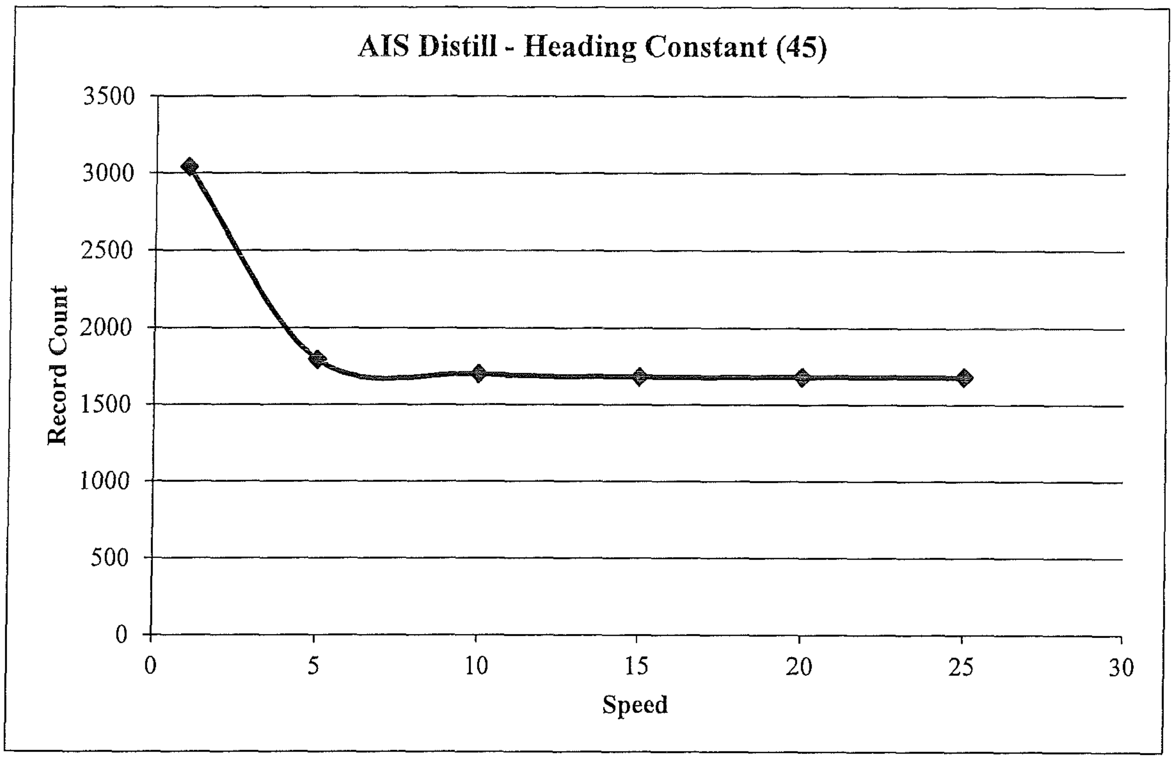
Similar test is carried out to find the optimal value for the tolerance level of the 'Heading' variable. Table 5.2 shows the results of the reduction percentage with the variable "Speed" being held constant. The optimum value of 45degree for the variable heading is chosen with the thought that even though the great lakes might seem similar to the open waters of the ocean but it still has rivers such as Maumee River, Detroit River, etc. that serves as shipping channel to shippers. With rivers having narrower channel that serpentine its way to a lake through some subtle to sharp bends, a higher value of *heading* would not help in properly defining the path of a vessel through a narrow water body.

Speed	Heading	Rec Cnt	% Reduction
1	45	3039	96.43
5	45	1794	97.89
10	45	1700	98.00
15	45	1681	98.02
20	45	1681	98.02
25	45	1681	98.02
Speed	**Heading**	**Rec Cnt**	**% Reduction**
1	40	3387	96 02
5	40	1889	97.78
10	40	1815	97.87
15	40	1800	97.88
20	40	1800	97.88
25	40	1800	97.88
Speed	**Heading**	**Rec Cnt**	**% Reduction**
1	35	3642	95.72
5	35	2016	97.63
10	35	1936	97.72
15	35	1918	97.75
20	35	1918	97.75
25	35	1918	97.75
Speed	**Heading**	**Rec Cnt**	**% Reduction**
1	30	3797	95.54
5	30	2209	97 40
10	30	2119	97.51
15	30	2102	97 53
20	30	2102	97.53
25	30	2102	97.53

Table 5 1 Reduction Percentage (Heading held constant)

Speed	Heading	Rec Cnt	% Reduction
6	5	6730	92.09
6	10	4595	94.60
6	15	3566	95.81
6	20	2936	96.55
6	25	2501	97.06
6	30	2170	97.45
6	35	1987	97.66
6	40	1851	97.82
6	45	1763	97.93

Table 5 2 Reduction Percentage (Speed held constant)



Graph 5 1 Record count after reduction process

Further on with the use of higher tolerance values for "heading" and "speed" variables the percentage of reduction shows less significant increase change to the point where the change in the percentage of reduction gets insignificant (Graph 5.1). Every record in the sample data that reflected its *speed* equal to or lesser than 0.1knots is flagged acceptable. With the exception of those records that had similar location value within a certain time frame. As this would denote a stop-event for a vessel either within a port or out in the open waters. The value of 15 for speed along with a value of 45 for heading helped in reducing the sample AIS data by 98 percent making it much more manageable for further analysis (Figure 5-3). This reduced or distilled AIS data can effectively be used for several applications. The next section details an algorithm developed to calculate individual vessel calls in a terminal, analyze vessel path, and method to associate average speed with the waterway network.

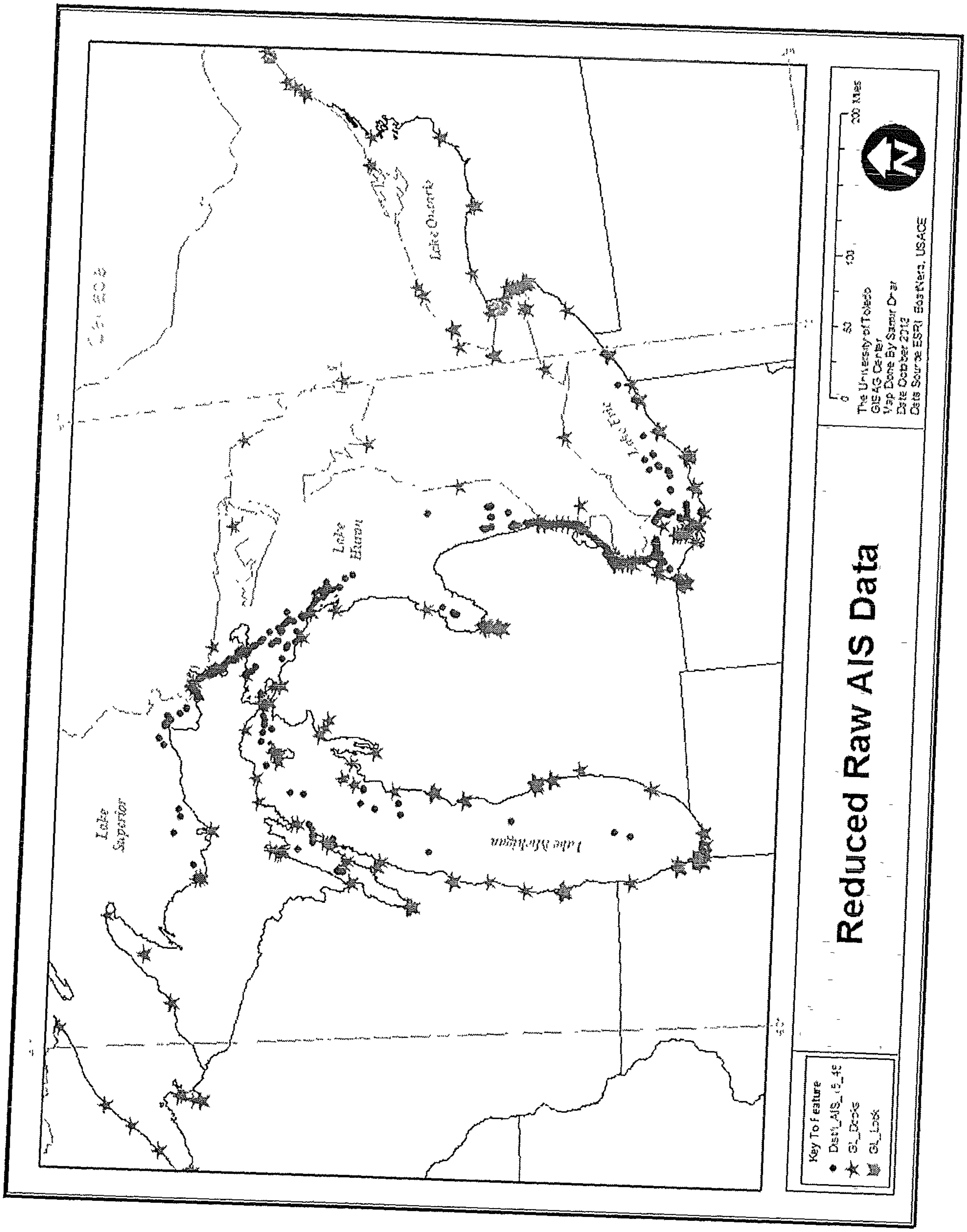


Figure 5-3 Reduced Raw AIS Data

Chapter 6

Application of Distilled AIS Data

6.1 AIS Application – Calculate Vessel Call(s)

The idea here is to use the distilled AIS data to calculate vessel calls made at a particular dock terminal. A vessel call is defined as when a vessel visits a dock to either load or unload commodities. In this example, the Dock at Stoneport, Michigan is chosen,

Figure 6-1 Stop-Event around Stoneport, MI

for which the vessel call is calculated for a set time period (Figure 6-1). The developed algorithm takes all the Stop-Events around the terminal. The reduced form of the AIS data captures all the Stop-Events without multiple records even though the vessel is halted for a long period of time.

The pseudo code shown in Figure 6-2 describes the process in which the vessel call is calculated. The procedure uses the reduced amount of AIS data to extract all AIS messages

```
Pseudo Code --Procedure AIS_DOCKVESSELCALLS(dock's Lat, dock's Long) Begins
Input to the procedure is the location of the Dock terminal
Declare variables to hold attribute(s) values, and other variables
Select all the records flagged 'Y' into a temp table from the Raw AIS Table.
Take the input location of the dock terminal and make a sweep around the dock to Flag those records within
the search limit
Repeat - Outer Loop
        Call Sub-Procedure to select all records within a set distance from the dock terminal
        Flag the records that match the criteria
Until all records are processed
Create and Open a Cursor to Store Unique MMSI Selected from the AIS Table flagged 'C'
Repeat - Outer Loop
  Select all the records flagged 'C' into a temp table for the next unique MMSI
        Repeat - Inner Loop
        Fetch the next record from the temp table
                IF This Record is the first record
                        Initialize the comparative variables
                        Increment Vessel call Variable
                ElseIf Location information is the same and date difference is more than 3
                        Initialize the comparative variables
                        Increment Vessel call Variable
                ElseIf Location information is not the same and date difference is more than 3
                        Initialize the comparative variables
                        Increment Vessel call Variable
                End If
                Until all records pertaining to the unique MMSI are processed
        Empty temp tables
        Fetch Next MMSI
 Until all unique MMSI are processed
Close Cursor
Release Resource
End Procedure AIS_DOCKVESSELCALLS
```

Figure 6-2 Pseudo Code Vessel Calls

within a specified distance around the dock terminal. Individual MMSI and its related records are examined for unique voyage, taking the location and time into consideration. The algorithm then checks for Stop-Events within that dataset and identifies unique vessel calls based on time and location.

6.1.1 Pseudo Code – Vessel Calls

The algorithm starts by selecting all the records from the raw AIS table that is flagged "Y" in a temporary table. All the records selected contains some records that resembles stop-event for vessels around a particular dock terminal. The next step is to identify such records and flag them for further analysis. This process is accomplished by selecting only those records that are within a set distance from the docks and is stored in another temporary table. Unique vessel's MMSI number is queried and stored in an array. The procedure then intakes single MMSI number and selects only those records that matched the MMSI number. A cursor is created that loops through each and every record pertaining to that particular MMSI number.

The procedure then uses the location variables (Lat & Long) and the time-stamp variable to detect vessel calls for that particular terminal. The value of locational and time-stamp variable of the very first record is assigned to the comparative variable. The value of the locational and time-stamp variable of the very next record is compared with the comparative variable. If there is no difference between the value of the comparative locational variables and the record in process, and if the difference between the time-stamp is greater than the defined days. Then the vessel call variable is incremented by a value of 1. This process runs through all the records for a particular MMSI number and then does the same operation for all the unique MMSI number in the array

6.2 AIS Application – Tracking Vessel Path

Tracking vessel's path provides useful information about its movements within the waterways and also helps in understanding the route it takes in relation to the waterway network Sometime unusual activity can be uncovered further investigation. In this effort tracking a vessel's path is done using historic AIS signals transmitted from a particular vessel. The raw AIS feed is initially distilled so it becomes very manageable and easy to use. The distilled AIS records is then used to create a vessel path of one vessel using ArcGIS software.

The locational variables i.e. Lat and Long along with time-stamp variable is used to create a vessel path. Using these variables the distilled AIS data for a particular vessel is arranged in a chronological order in an Excel sheet. The data is then imported into ArcGIS software for transforming it to a geographical file format (Shapefile) The shapefile is created using the "Conversion Tool" within ArcGIS Toolbox and then saved it as a shapefile. The "Conversion Tool" uses the Lat and Long variable in the AIS data to draw each feature that corresponds each record in the table.

Figure 6-3 shows the point shapefile of the AIS data of a particular vessel (366904930). Every point shape denotes the location from where the vessel transmitted the AIS signal. The next step is to create a line feature using the point shapefile. The "Data Management Tool" in ArcGIS Toolbox is used to create a line feature with the AIS point shapefile Time-stamp variable within the AIS point shapefile is used to chronologically order the creation of the line feature. This ensures that the track of the vessel moves in a chronological order.

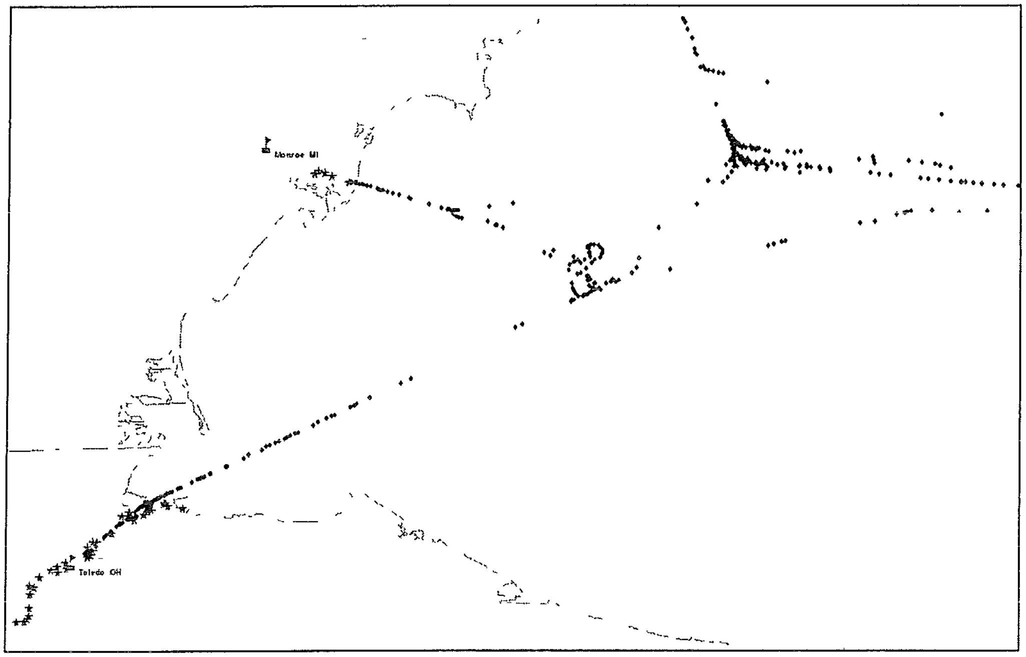


Figure 6-3 AIS Data Mapped (366904930)

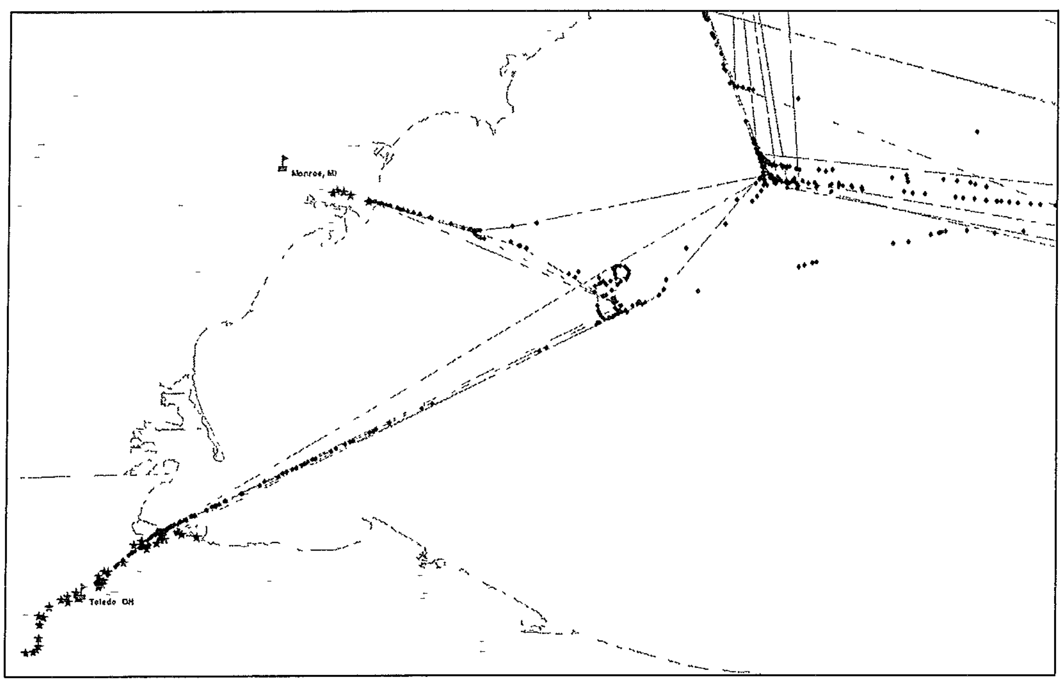


Figure 6-4 Vessel Path (366904930)

The Figure 6-4 shows the path of the vessel (366904930) and its movement for a certain time period. The next step is to overlay the waterway network acquired from the USACE to match the vessel's path with it. Visual inspection shows that the vessel path and the waterway network most of the time in open water does not coincide rather it runs parallel or is taking a different route (Figure 6-5). In the inland river the path follows the waterway network since rivers have restricted navigational channels. Vessel path can be used to rectify, re-align, or modify the existing waterway network to depict the actual route vessel(s) usually take while moving around the Great Lakes.

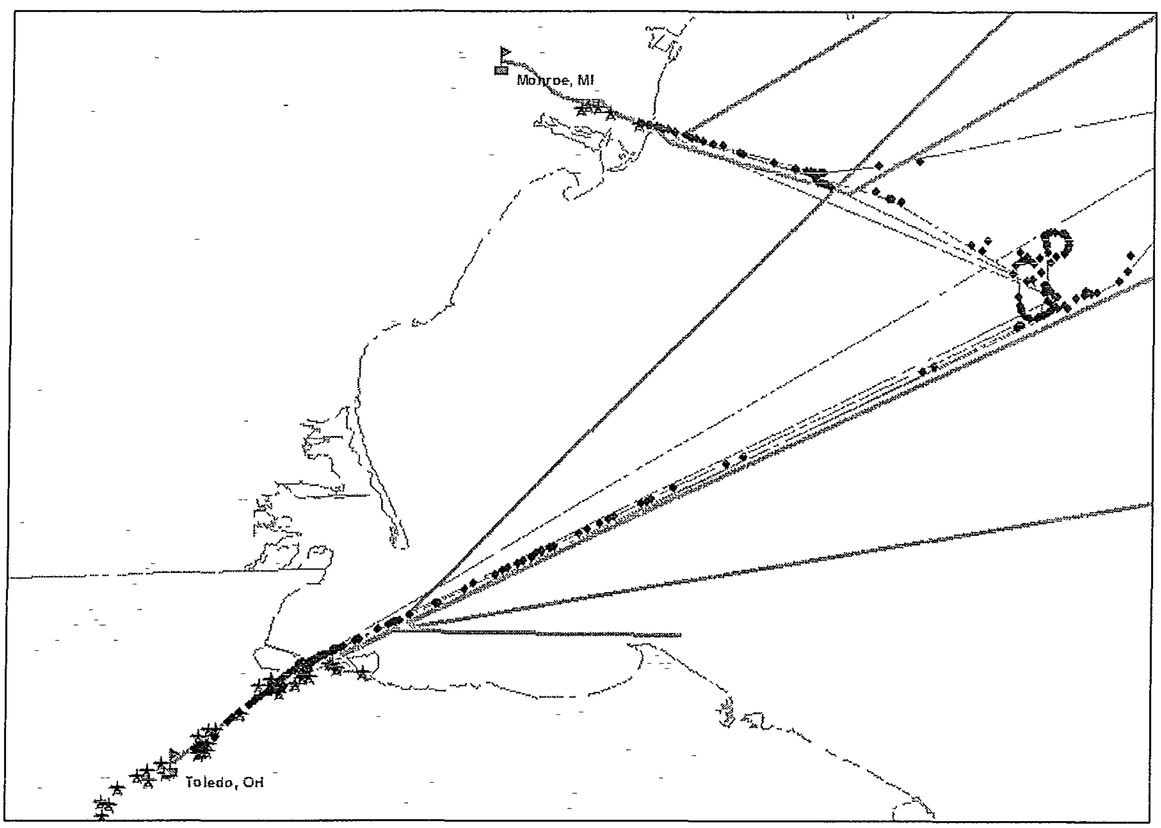


Figure 6-5 Vessel Path and Waterway Network

Vessel path can also show movements that looks little suspicious to the naked eye. In Figure 6-6 there is a movement where the vessel after coming out of Monroe, MI makes a figure 8 maneuver and then heads for Port of Toledo, OH. Such movements can be indicative of possible illegal activity and further scrutiny is deemed necessary by proper authorities. Vessel path can help in identifying high traffic areas where high level of monitoring could help in safety mariners.

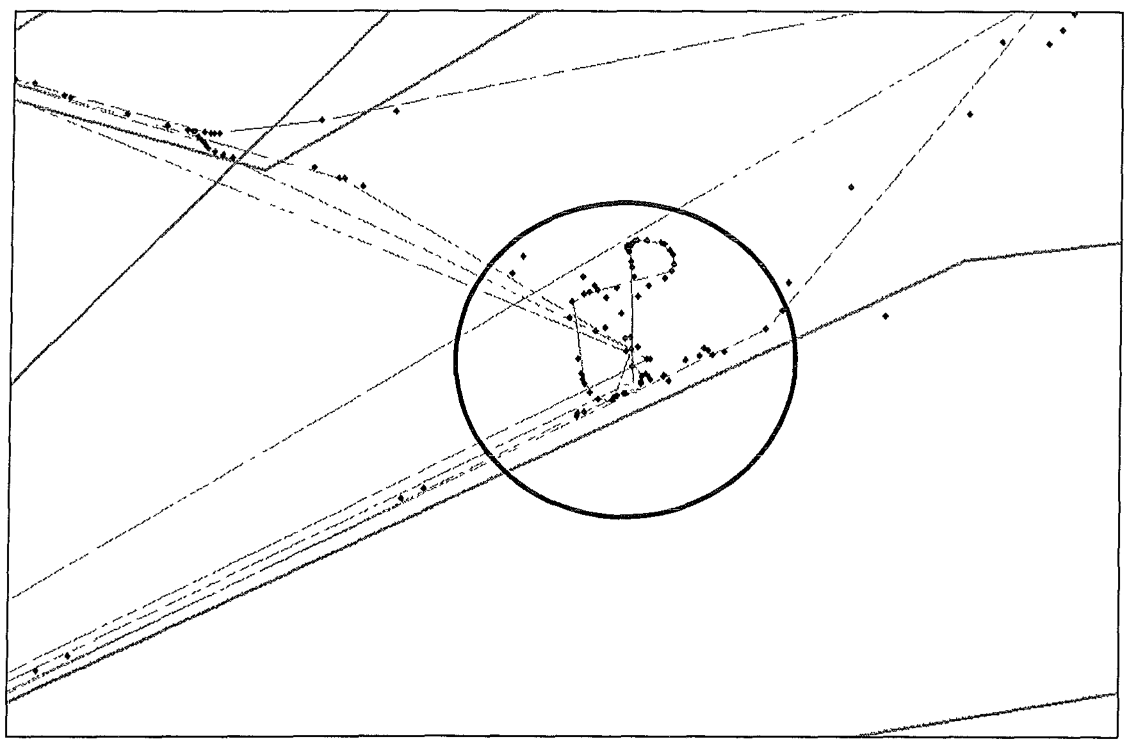


Figure 6-6 Unusual Vessel Movements (366904930)

6.3 AIS Application – Estimating Speed on Waterway Network

Transportation analyst, academics, and private entities use transportation network to estimate travel time between two locations. Travel time information is useful in routing freight over a network or series of connecting networks very efficiently. Prevailing transportation network that offers speed as variable in their data is the Highway Network. Besides the Highway network, the Rail or the Waterway network does not offer this feature or information in its data. Therefore the effort in this application of AIS is to provide a methodology to estimate average speed on the Waterway network.

For this study Distilled AIS data is used along with Waterway network. The AIS data has a speed variable that shows the speed of the vessel at that point of time and location This speed variable is used to estimate average speed over the Waterway network. ArcGIS software is used in estimating Speed on a link of the Waterway network. Below is a detail break down of the methodology to estimate average speed.

Step 1. Buffer tool inside ArcGIS is used to create a buffer around the waterway link. In this application a buffer area of 1mile radius is used, so that every AIS signal from different vessel can be captured within it (Figure 6-7).

Step 2. This buffer area is used to select all the AIS signals that is completely within the area. Selection by location tool within ArcGIS is used to select only those AIS records that is completely with in the buffer area (Figure 6-8).

Step 3. The highlighted records (Figure 6-9) within the buffer area is selected and then the average speed for the link under consideration is calculated

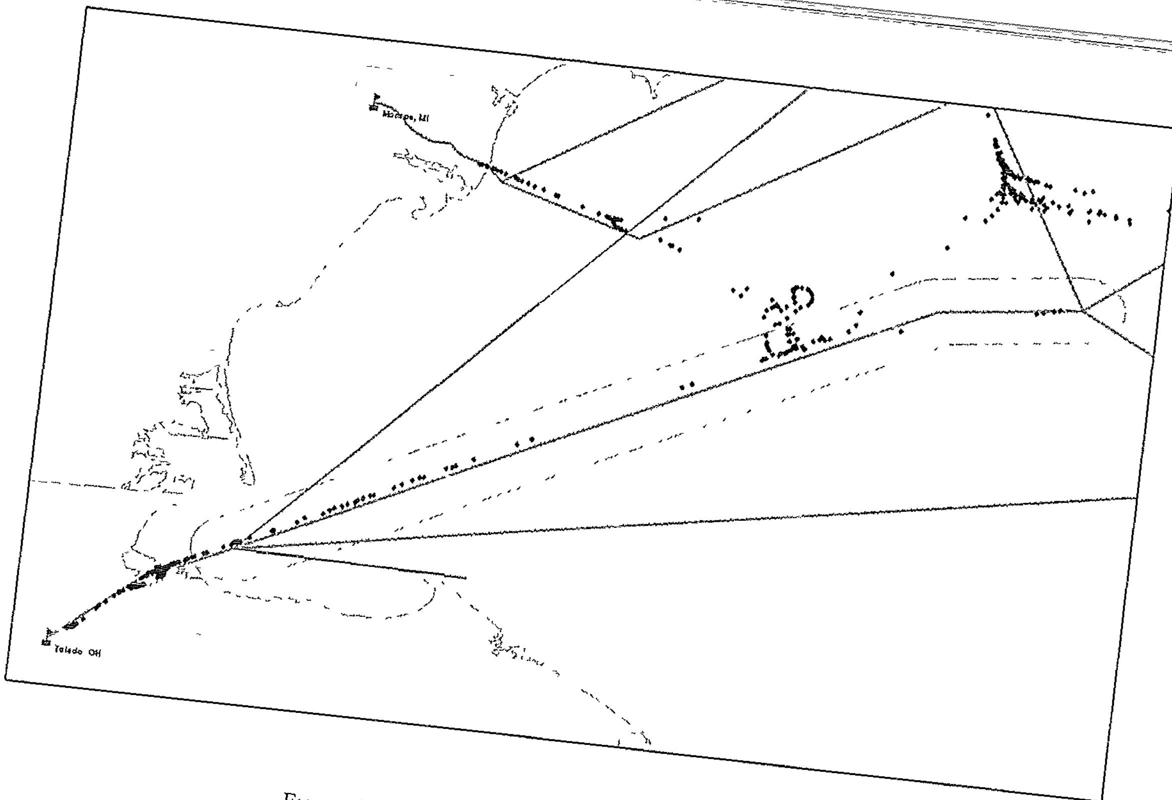


Figure 6-7 Buffer Area (1mile) around Waterway Link

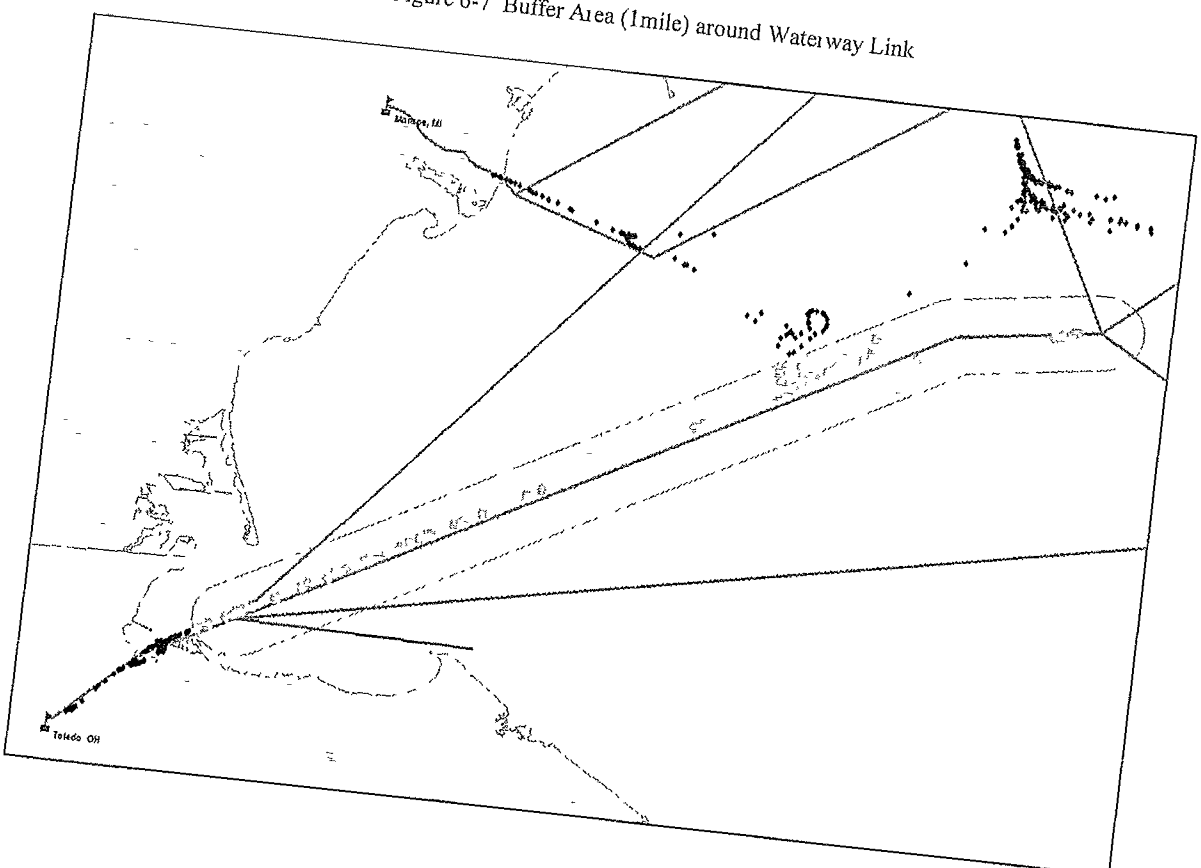
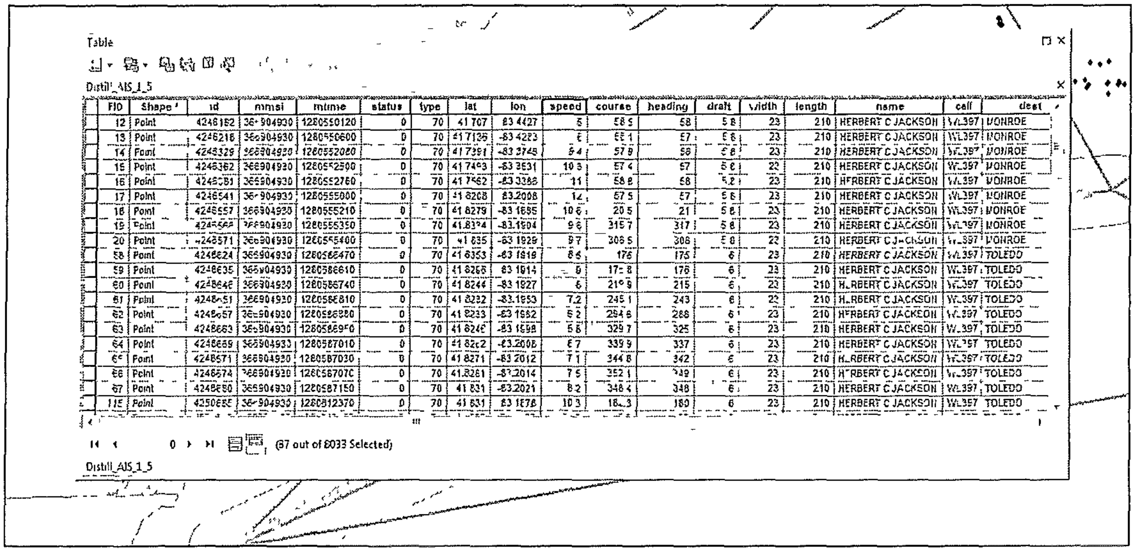


Figure 6-8 Selected Records within Buffer Area



FID	Shape *	id	mmsi	mtime	status	type	lat	lon	speed	course	heading	draft	width	length	name	call	dest
12	Point	4248182	366904930	1280550120	0	70	41.707	83.4427	5	58.5	58	5.8	23	210	HERBERT C JACKSON	WL397	MONROE
13	Point	4248216	366904930	1280550600	0	70	41.7126	-83.4223	6	55.1	57	5.8	23	210	HERBERT C JACKSON	WL397	MONROE
14	Point	4248329	366904930	1280552080	0	70	41.7391	-83.3748	9.4	57.9	58	5.8	23	210	HERBERT C JACKSON	WL397	MONROE
15	Point	4248362	366904930	1280552500	0	70	41.7493	-83.3531	10.3	57.4	57	5.8	22	210	HERBERT C JACKSON	WL397	MONROE
16	Point	4248381	366904930	1280552760	0	70	41.7562	-83.3386	11	58.8	58	5.2	23	210	HERBERT C JACKSON	WL397	MONROE
17	Point	4248541	366904930	1280555000	0	70	41.8208	83.2008	12	57.5	57	5.8	23	210	HERBERT C JACKSON	WL397	MONROE
18	Point	4248557	366904930	1280555210	0	70	41.8279	-83.1885	10.6	20.5	21	5.8	23	210	HERBERT C JACKSON	WL397	MONROE
19	Point	4248568	366904930	1280555350	0	70	41.8324	-83.1904	9.6	315.7	317	5.8	23	210	HERBERT C JACKSON	WL397	MONROE
20	Point	4248571	366904930	1280555400	0	70	41.835	-83.1929	9.7	308.5	308	5.8	22	210	HERBERT C JACKSON	WL397	MONROE
58	Point	4248624	366904930	1280586470	0	70	41.8353	-83.1919	8.5	176	175	6	23	210	HERBERT C JACKSON	WL397	TOLEDO
59	Point	4248635	366904930	1280586610	0	70	41.8268	83.1914	9	172.8	176	6	23	210	HERBERT C JACKSON	WL397	TOLEDO
60	Point	4248646	366904930	1280586740	0	70	41.8244	-83.1927	6	219.9	215	6	23	210	HERBERT C JACKSON	WL397	TOLEDO
61	Point	4248651	366904930	1280586810	0	70	41.8232	-83.1953	7.2	245.1	243	6	22	210	HERBERT C JACKSON	WL397	TOLEDO
62	Point	4248657	366904930	1280586880	0	70	41.8233	-83.1982	5.2	294.6	288	6	23	210	HERBERT C JACKSON	WL397	TOLEDO
63	Point	4248663	366904930	1280586950	0	70	41.8246	-83.1998	5.6	329.7	325	6	23	210	HERBERT C JACKSON	WL397	TOLEDO
64	Point	4248669	366904930	1280587010	0	70	41.8262	-83.2008	6.7	339.9	337	6	23	210	HERBERT C JACKSON	WL397	TOLEDO
65	Point	4248671	366904930	1280587030	0	70	41.8271	-83.2012	7.1	344.8	342	6	23	210	HERBERT C JACKSON	WL397	TOLEDO
66	Point	4248674	366904930	1280587070	0	70	41.8281	-83.2014	7.5	352.1	349	6	23	210	HERBERT C JACKSON	WL397	TOLEDO
67	Point	4248680	366904930	1280587150	0	70	41.831	-83.2021	8.2	348.4	348	6	23	210	HERBERT C JACKSON	WL397	TOLEDO
115	Point	4250686	366904930	1280812370	0	70	41.831	83.1878	10.3	18.3	180	6	23	210	HERBERT C JACKSON	WL397	TOLEDO

Figure 6-9 Selected AIS records

All the applications detailed above uses the distilled AIS data. The distilled AIS data not only enables for faster processing but also quicker results either at the server level or on the desktop, but also ensures important information is not lost. As in the example of the irregular vessel movement that was visually identified in the above application.

Chapter 7

Conclusion & Discussion

This study details the techniques of AIS data acquisition from a non-profit source, which is autonomous and automatic. The data acquired for the GL region is coupled with errors and anomalies. Data-mining for errors using SQL queries revealed the different types of errors the AIS data possess. Remediation of the errors using ancillary data sources help in rectifying errors produced from faulty entering of information. Unusable garbage AIS records are identified and flagged for removal. After the cleansing of the data, the next step is to identify if there are gaps in the data. First, the raw AIS data is converted to a geographical file format. ArcMap is used to convert the AIS data by geocoding individual AIS records using location attributes. Maps are made and areas are visually identified where there are gaps in data due to no antenna being installed to capture the AIS data.

Furthermore, with the high volume of AIS data, efforts is made to develop techniques to distill the AIS data. The distillation process does not compress the data but actually it reduces the number of records within the table. The developed algorithm uses the Spatio-Temporal variables to boil the data down to a more manageable amount without losing important information. Further, Stop-Events are identified using the value of the speed variable which is helpful in calculating vessel calls at a particular Dock. Then, an algorithm is developed in MS SQL Server using the distilled AIS records. This algorithm makes a sweep around the dock facility and grabs all the nearby stop-events. The selected stop-events is analyzed and used to calculate the number of vessel calls for that particular dock. The overall effort in this study helps the research community with better management of AIS data. Few studies have subtly mentioned the nature of the AIS data as over a period of time, it grows in numbers. Satellite AIS and terrestrial AIS are used in tandem to globally monitor and track vessel movements within waterways, which results in a huge amount of data. The size of the data will definitely become an issue; this study helps to address this issue.

There are other potential applications of AIS that can use the distillation process for faster processing. Some of the application might require AIS data to be less distilled or reduced, such as emission estimation around ports, estimating traffic volumes within inland rivers. The entire objective of AIS is to promote safe navigation to mariners may it be commercial or recreational user. Initiatives and efforts made by agencies to improve satellite AIS technology so as to effectively monitor vessel movements in open water. AIS has also opened doors for different countries to get together and collaborate its effort from a global perspective to monitor and track vessels. Initiatives such as LRIT and e-Navigation not only provides an opportunity to integrate disparate data but also possess an issue of managing Big Data. Distillation of AIS Data (Big Data) becomes very necessary for faster processing of data and quicker delivery of information for proper decision making.

References

AAPA. (2013, November 14). *American Association of Port Authorities*. Retrieved May 2014, from General cargo shipment keep US Great Lakes ports busy.: http://www.aapa-ports.org/press/prdetail.cfm?itemnumber=19447

Amato, F., Fiorini, M., Gallone, S., & Golino, G. (2011). e-Navigation and Future Trend in Navigation. *The International Journal on Marine Navigation and Safety of Sea Transportation, 5*(1), 11-14.

Arundale, N. (2014, 5). *AIS Decoder*. Retrieved 5 2014, from nmearouter: http://nmearouter com/docs/ais/ais_decoder.html

Burzigotti, P., Ginesi, A., & Colavolpe, G. (2012). Advanced receiver design for satellite-based automatic identification system signal detection. *INTERNATIONAL JOURNAL OF SATELLITE COMMUNICATIONS AND NETWORKING, 30*, 52-63.

Cairns, B. (2011). e-Navigation: Revolution and Evolution. *The Coast Guard Journal of Safety and Security at Sea Proceeding of The Marine Safety and Security Council, 68*, pp. 48-50. Retrieved 2013, from US Coast Guards: http://www.uscg.mil/proceedings/archive/2011/Vol68_No1_Spr2011 pdf

Cairns, W. R. (2005). AIS and Long Range Identification and Tracking. *The Journal of Navigation, 58*, 181-189. doi:10.1017/S0373463305003267

Calder, B., & Schwehr, K. (2009, May 11-14). Traffic Analysis for the Calibration of Risk Assessment Methods. *U.S. Hydro*. Norfolk, Virginia.

Carson-Jackson, J. (2012). Satellite AIS – Developing Technology or Existing Capability? *THE JOURNAL OF NAVIGATION, 65*, 303-321.

Cervera, M. A., Ginesi, A., & Eckstrin, K. (2011). Satellite-based vessel Automatic Idenfication System: A feasibility and performance analysis. *International Journal of Satellite Communications and Networking, 29*(2), 117-142.

Chen, Y. (2014, June). Satellite-based AIS and Its Comparison with LRIT. *TRANSNAV, International Journal on Marine Navigation and Safety of Sea Transportation, 8*. doi:10.12716/1001.08.02.02

CMTS. (2012, Febuary). *CMTS e-Navigation Strategic Action Plan.* Retrieved 2014, from Committee On The Marine Transportation System: http://www.cmts.gov/downloads/CMTS_e-Navigation_Strategic_Action_Plan_Feb_2012.pdf

Duluth Seaway Port Authority. (2014). *Great Lakes/Seaway System* Retrieved September 2014, from Port of Dulth-Superior: http.//www.duluthport com/port-seaway.php

Harati-Mokhtari, A., Wall, A., Brooks, P., & Wang, J. (2007). Automatic Identification System (AIS): Data Reliability and Human Error Implications. *THE JOURNAL OF NAVIGATION, 60*, 373-389.

Hoye, G. K., Eriksen, T., Meland, B. J., & Narheim, B. T. (2008). Space-basedAIS for global maritime traffic monitoring *Acta Astronautica, 62*, 240-245.

Hu, Q., Yong, J., Shi, C., & Chen, G. (2010, March). Evaluation of Main Traffic Congestion Degree for Restricted Waters with AIS Reports. *TRANSNAV International Journal on Marine Navigation and Safety of Sea Transportation, 4*(1), 55-58.

IMO. (2014). *Long-range identification and tracking (LRIT)*. Retrieved October 2014, from International Maritime Organization: http://www.imo.org/OurWork/Safety/Navigation/Pages/LRIT.aspx

IMO. (2015). *E-Navigation.* Retrieved 2015, from International Maritime Organization: http://www.imo.org/OurWork/Safety/Navigation/Pages/eNavigation.aspx

International Maritime Organization. (2014). *Vessel Traffic Services*. Retrieved 10 12, 2013, from IMO Web Site: http://www.imo.org/OurWork/Safety/Navigation/Pages/VesselTrafficServices.aspx

ITU. (2014, Febuary). *M.1371 : Technical characteristics for an automatic identification system using time-division multiple access in the VHF maritime mobile band* Retrieved May 2014 , from International Telecommunication Union: http://www.itu.int/rec/R-REC-M.1371-5-201402-I/en

ITU-R M.2169. (2010). *Improved Satellite Detection of AIS.* Geneva: Electronic Publication.

Kaiser, M. J. (2014, December). Channel Utilization in South Louisiana Using AIS Data: 2011-2012. *TRANSNAV The International Journal on Marine Navigation and Safety of Sea Transportation, 8*(4), 515-522.

Kao, S.-L., Lee, K.-T., Chang, K.-Y., & Ko, M.-D. (2007). A Fuzzy Logic Method for Collision Avoidance in Vessel Traffic Service. *The Journal of Navigation, 60*, 17-31.

Lin, B., & Huang, C.-H. (2006). COMPARISON BETWEEN ARPA RADAR AND AIS CHARACTERISTICS FOR VESSEL TRAFFIC SERVICES. *Journal of Marine Science and Technology, 14*(3), 182-189.

Martin Associates. (2011). *The Economic Impacts of the Great Lakes- St. Lawrence Seaway System* Retrieved June 2013, from http://www.greatlakesports.org/documents.html

NAVCEN. (2011, March 17). *U. S. Coast Guard Navigation Center*. Retrieved December 18, 2012, from http://www.navcen.uscg.gov

NAVCEN. (2014, September 9). *Types of AIS Systems*. Retrieved October 2014, from Navigation Center - US DHS, USCG: http://www.navcen.uscg.gov/?pageName=typesAIS

NAVCEN USCG. (2014, July). *LONG RANGE IDENTIFICATION AND TRACKING (LRIT) OVERVIEW*. Retrieved October 2014, from Navigation Center USCG USDHS: http://www.navcen.uscg.gov/?pageName=lritMain

Perez, H., Chang, R., Billings, R., & Kosub, T. L. (2009). Automatic Identification Systems (AIS) Data Use in Marine Vessel Emission Estimation. *18th Internatinal EmissionInventory Conference* Baltimore.

Peters, D., & Hammond, T. (2011). Interpolation Between AIS Reports: Probabilistic Inference Over Vessel Path Space. *The Journal Of Navigation, 64*, 595-607.

Silveira, P., Teixeira, A., & Soares, G. C. (2013). Use of AIS Data to CharacteriseMarine Traffic Patterns and Ship Collision Risk offthe Coast of Portugal. *The Journal of Navigation, 66*, 879-898.

Stewart, R. D. (2006). *Great Laes Marine Transportation System.* White Paper.

U. S. Coast Guard Navigation Center. (2014, 30 7). *History of Vessel Traffic Services*. Retrieved 10 12, 2013, from NAVCEN - USCG: http://www.navcen.uscg.gov/?pageName=vtsHistory

US EPA. (2012, 6 25). *Great Lakes An Environmental Atlas and Resource Book.* Retrieved May 2014, from US EPA: http://www.epa.gov/greatlakes/atlas/index.html

USACE. (2013, December). *FACTS.* Retrieved January 2014, from Navigation Data Center: http://www.navigationdatacenter.us/factcard/factcard.htm

USCG - DHS. (2008). *Vessel Requirements for Notices of Arrival and Departures, and Automatic Identification System* Federal Register. Retrieved October 2014

USCG NAVCEN. (2013, June 25) *How AIS Works*. Retrieved May 2014, from Navigation Center USDHS USCG: http://www.navcen.uscg.gov/?pageName=AISworks

USCG NAVCEN. (2014, September). *AIS CLASS A SHIP STATIC AND VOYAGE RELATED DATA (MESSAGE 5)*. Retrieved October 2014, from Navigation Center DHS USCG. http://www.navcen.uscg.gov/?pageName=AISMessagesAStatic

USDOT. (2013). *Freight Facts And Figures 2013.* US FHWA and BTS. Retrieved September 2014, from http://aapa.files.cms-plus.com/Statistics/Freight%20Facts%20and%20Figures%202013.pdf

Yuli, C. (2014). Detection Probabillity Analysis of Space-Based AIS Signal by Using Geometric Model. *Journal of Shipping and Ocean Engineering, 4*, 161-171.

Appendix A

AIS Message Tables

A.1 Navigation Status

Value	Description
0	Under way using engine
1	At anchor
2	Not under command
3	Restricted maneuverability
4	Constrained by her draught
5	Moored
6	Aground
7	Engaged in Fishing
8	Under way sailing
9	Reserved for future amendment of Navigational Status for HSC
10	Reserved for future amendment of Navigational Status for WIG
11	Power-driven vessel towing astern (regional use)
12	power-driven vessel pushing ahead or towing alongside (regional use)
13	Reserved for future use
14	AIS-SART is active
15	Not defined (default)

A.2 Rate of Turn (ROI)

Value	Description
0	Not Turning
1 to +126	turning right at up to 708 degrees per minute or higher
1 to -126	turning left at up to 708 degrees per minute or higher
127	turning right at more than 5 degrees per 30 s (No TI available)
-127	turning left at more than 5 degrees per 30 s (No TI available)
-128 (80 hex)	No turn information available (default)

A.3 Ship Type (ROI)

Identifier No.	Special craft
50	Pilot vessel
51	Search and rescue vessels
52	Tugs
53	Port tenders
54	Vessels with anti-pollution facilities or equipment
55	Law enforcement vessels
56	Spare - for assignments to local vessels
57	Spare - for assignments to local vessels
58	Medical transports (as defined in the 1949 Geneva Conventions and Additional Protocols)
59	*Ships and aircraft of States not parties to an armed conflict*

A.4 Ship and Cargo Classification

Code	Ship & Cargo Classification
50	Pilot vessel
52	Tugs
70	Cargo, All ships of this type
71	Cargo, Hazardous category A
72	Cargo, Hazardous category B
73	Cargo, Hazardous category C
74	Cargo, Hazardous category D
75	Cargo, Reserved for future use
76	Cargo, Reserved for future use
78	Cargo, Reserved for future use
79	Cargo, Reserved for future use
80	Tanker, All Ships of this type
81	Tanker, Hazardous category A
82	Tanker, Hazardous category B
83	Tanker, Hazardous category C
84	Tanker, Hazardous category D
85	Tanker, Reserved for future use
86	Tanker, Reserved for future use
87	Tanker, Reserved for future use
88	Tanker, Reserved for future use
89	Tanker, No additional Information
90	Other Type, All Ships of this type
91	Other Type, Hazardous category A
92	Other Type, Hazardous category B
93	Other Type, Hazardous category C
94	Other Type, Hazardous category D
95	Other Type, Reserved for future use
96	Other Type, Reserved for future use
97	Other Type, Reserved for future use
98	Other Type, Reserved for future use
99	Other Type, Reserved for future use
100	Other Type, No Additional Information

Appendix B

PL SQL Procedure

B.1 AIS DISTILL Procedure

```
====This Procedure distills AIS records to a minimum
Create Procedure AIS_Distill
AS
BEGIN

    SET NOCOUNT ON;
    Set XACT_ABORT ON;

    -- Create A Cursor to Store Unique MMSI Selected from the AIS Table
     Declare cursor_MMSI CURSOR FAST_FORWARD
     For SELECT Distinct MMSI FROM T_AIS
     where Type In (30,31,50,52,70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84,
85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97,98, 99, 100)


    --Declare Variables
    Declare @MMSI int
    Declare @RecToProc int
    Declare @CurRow int

    --Declare variables to hold attribute(s) values
    DECLARE @rowcnt int
    DECLARE @ID int, @lat float, @long float, @speed float, @heading int, @course
int, @timestamp datetime
    Declare @init_lat float, @init_long float, @init_speed int, @init_heading int,
@init_course int, @init_timestamp Datetime
```

```
Open cursor_MMSI
FETCH NEXT FROM cursor_MMSI into @MMSI

--PRINT 'Variable @MMSI = ' + CAST(@MMSI as nVarchar)

 WHILE @@FETCH_STATUS = 0
 Begin

        Execute INSERT_T_MMSI_AIS @MMSI_T = @MMSI

        Set @RecToProc = (Select count(*) from T_MMSI_AIS)
        --PRINT 'Variable @RecToProc = ' + CAST(@RecToProc as nvarchar)
        Set @curRow = 1


        --Select the firt record for the given MMSI
        --Select @init_lat = lat, @init_long = long, @init_speed = speed,
@init_heading = heading, @init_course = course, @init_timestamp = timestamp from
T_MMSI_AIS
        --Where RNum = 1

        --Reading through every record in the T_MMSI_AIS table
        WHILE @CurRow <= @RecToProc
        BEGIN
             --Reads the Next Record From T_MMSI_AIS Table
             Select @ID = ID, @lat = lat, @long = long, @speed = speed, @heading =
heading, @timestamp = timestamp, @course = course from T_MMSI_AIS
                                             where RNum = @CurRow


             IF @CurRow = 1
                Begin

                  --Initialize the Initial Values of Variables
                  set @init_lat = @lat
                  set @init_long = @long
                  set @init_speed = @speed
                  set @init_Course = @course
```

```
set @init_heading = @heading
Set @init_timestamp = @timestamp

Update T_AIS set extracode = 'Y' where Id = @ID --Good Records
Marked as "Y"

END
ELSE IF @lat = @init_lat and @long = @init_long and @timestamp =
@init_timestamp and @CurRow > 1
Begin

Update T_AIS set extracode = 'D' where Id = @ID  --Duplicate Records
Marked as "D"

END
ELSE IF @lat != @init_lat and @init_long != @long
Begin
If @Speed != @init_Speed and @heading != @init_heading
BEGIN
If ABS(@speed - @init_speed) >= 15 OR ABS(@heading -
@init_heading) >= 45 or (@speed IN (0,0.1))
BEGIN
Update T_AIS set extracode = 'Y' where Id = @ID --Good
Records Marked as "Y"

--RE-Initialize the Initial Values of Variables
set @init_lat = @lat
set @init_long = @long
set @init_speed = @speed
set @init_heading = @heading
set @init_Course = @course
Set @init_timestamp = @timestamp
END
--                                  Else
--                                  BEGIN
--          --RE-Initialize the Initial Values of Variables
--          set @init_lat = @lat
--          set @init_long = @long
--          set @init_speed = @speed
--          set @init_heading = @heading
```

```
--                              set @init_Course = @course
--                              Set @init_timestamp = @timestamp
--                                                                END

                                                      END


                END

              Set @curRow = @curRow + 1

            END --Inner While Loop Ends

    DELETE FROM T_MMSI_AIS;
    FETCH NEXT FROM cursor_MMSI into @MMSI

    END  -- Outer While Loops Terminated

    CLOSE cursor_MMSI;
    DEALLOCATE cursor_MMSI;

END
```

B.2 AIS Select Records Procedure

```
Create Procedure INSERT_T_MMSI_AIS(@MMSI_T  INT)
AS
BEGIN

  SET NOCOUNT ON;

  Insert into T_MMSI_AIS (ID,
MMSI,lat,long,speed,heading,Course,timestamp,RNum )
  Select ID, MMSI,lat,lon,speed,heading,course,timestamp, Row_Number()
Over(Order By ID) from T_AIS
  where MMSI = @MMSI_T
        Order by ID, timestamp
```

```
END
GO
```

B.3 Calculate Vessel Calls on Docks Procedure

```
====This Procedure Calculates A Particular Dock's Vessel Calls
Create Procedure AIS_DOCKVESSELCALLS(@d_lat Float(18), @d_Lon Float(18))
AS
BEGIN

    SET NOCOUNT ON,
    Set XACT_ABORT ON;


            DECLARE @ID int,@MMSI int, @lat float, @long float, @speed float,
@heading int, @course int, @timestamp datetime
                                    DECLARE @a_ID int,@a_MMSI int, @a_lat float,
@a_long float, @a_speed float, @a_heading int, @a_course int, @a_timestamp datetime
                                    DECLARE @init_ID int,@init_MMSI int, @init_lat float,
@init_long float, @init_speed float, @init_heading int, @init_course int,
@init_timestamp datetime
            Declare @dock_d float
            Declare @D_MMSI int
            Declare @CurRow int
            Declare @d_CurRow int
            Declare @RecToProc int
            Declare @d_RecToProc int
            Declare @D_Vessel_Call int


            --Select AIS Distilled Records into T_MMSI_AIS table
            Insert into T_MMSI_AIS (ID,
MMSI,lat,long,speed,heading,Course,timestamp,RNum )
            Select ID, MMSI,lat,lon,speed,heading,course,timestamp, Row_Number()
Over(Order By ID) from T_AIS
            where extracode = 'Y'
            Order by ID, timestamp
```

```
Set @RecToProc = (Select count(*) from T_MMSI_AIS)
PRINT 'Variable @RecToProc = ' + CAST(@RecToProc as nvarchar)

Set @curRow = 1

--Reading through every record in the T_MMSI_AIS table
WHILE @CurRow <= @RecToProc
                    BEGIN
  --Reads the Next Record From T_MMSI_AIS Table
  Select @ID = ID, @MMSI = MMSI, @lat = lat, @long = long, @speed =
speed, @heading = heading, @timestamp = timestamp, @course = course from
T_MMSI_AIS where RNum = @CurRow


  --Get the distance between AIS signal and the DOck
  Execute @dock_d = AIS_DIST @lat,@long,@d_lat,@d_Lon

  If  @dock_d <= 120
                         Begin
    Update T_AIS set extracode = 'C' where Id = @ID --Records Near The
Dock Marked as "C"

                         End

  Set @dock_d = 0.00
                    Set @curRow = @curRow + 1

                   END --While Loop Ends


Select * into #Dock_Call_AIS  from T_AIS Where extracode ='C' Order By
MMSI, ID, timestamp

Set @D_Vessel_Call = 0
DELETE T_MMSI_AIS;
---Create A Cursor to Store Unique MMSi from temp table #Dock_Call_AIS
Declare cursor_Dock_Call_MMSI CURSOR FAST_FORWARD
For SELECT Distinct MMSI FROM #Dock_Call_AIS
```

```
Open cursor_Dock_Call_MMSI
FETCH NEXT FROM cursor_Dock_Call_MMSI into @D_MMSI
PRINT 'Variable @D_MMSI = ' + CAST(@D_MMSI as nvarchar)

While @@FETCH_STATUS = 0  --Outer While Loop
Begin

                        Insert into T_MMSI_AIS (ID,
MMSI,lat,long,speed,heading,Course,timestamp,RNum )
                        Select ID,
MMSI,lat,lon,speed,heading,course,timestamp, Row_Number() Over(Order By ID) from
T_AIS
                        where MMSI = @D_MMSI and extracode
='C'
                        Order by ID, timestamp

                        Set @d_RecToProc = (Select count(*) from
T_MMSI_AIS)
                        Set @d_curRow = 1
                        PRINT 'Variable @d_RecToProc = ' +
CAST(@d_RecToProc as nvarchar)
        WHILE @d_CurRow <= @d_RecToProc  --Inner While Loop
        BEGIN
            --Reads the Next Record From T_MMSI_AIS
                                        PRINT
'Inside Inner While Loop '
            Select @a_ID = ID, @a_MMSI = MMSI, @a_lat = lat,
@a_long = long, @a_speed = speed, @a_heading = heading, @a_timestamp =
timestamp, @a_course = course from T_MMSI_AIS
            where RNum = @d_CurRow

            IF @d_curRow = 1
              Begin

Set @D_Vessel_Call = @D_Vessel_Call + 1

set @init_lat = @a_lat

set @init_long = @a_long
```

```
set @init_speed = @a_speed

Set @init_timestamp = @a_timestamp

PRINT 'Variable @d_Vessel_Call = ' + CAST(@D_Vessel_Call as nvarchar)
                              End
                            ELSE IF @lat = @init_lat and @long = @init_long and
DATEDIFF(DAY,@init_timestamp, @timestamp) >3
                              Begin
                                Set @D_Vessel_Call = @D_Vessel_Call + 1
                                set @init_lat = @a_lat
                                set @init_long = @a_long
                                set @init_speed = @a_speed
                                Set @init_timestamp = @a_timestamp
                              End
                                                                          ELSE
IF @a_lat != @init_lat and @a_long != @init_long and
DATEDIFF(DAY,@init_timestamp, @a_timestamp) >3
                              Begin
                                Set @D_Vessel_Call = @D_Vessel_Call + 1
                                set @init_lat = @a_lat
                                set @init_long = @a_long
                                set @init_speed = @a_speed
                                Set @init_timestamp = @a_timestamp
                              End
                            -- END   ----IF Loop Ends

                            Set @d_curRow = @d_curRow + 1

                      END --Inner While Loop Ends

                                              DELETE T_MMSI_AIS;
                                              FETCH NEXT FROM
cursor_Dock_Call_MMSI into @D_MMSI;

                                         END  -- Outer While Loops Terminated


                                         CLOSE cursor_Dock_Call_MMSI;
```

```
                    DEALLOCATE cursor_Dock_Call_MMSI;


    Return @D_Vessel_Call;
END


================================================================
======

Declare @d_V_Call float;
Exec @d_V_Call = AIS_DOCKVESSELCALLS  45.297379,-83 421144
Select 'Total Vessel Calls' = @d_V_Call
GO
```

CFIRE

University of Wisconsin-Madison
Department of Civil and Environmental Engineering
1415 Engineering Drive, Room 2205
Madison, WI 53706
Phone: 608-263-3175
Fax: 608-263-2512
cfire.wistrans.org