<table>
<thead>
<tr>
<th>Title</th>
<th>Mining and cleaning real maritime data for security in the defence of Singapore's waters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Jonathan, Cherian Samraj</td>
</tr>
<tr>
<td>Citation</td>
<td>Jonathan, C. S. (2012, March). Mining and cleaning real maritime data for security in the defence of Singapore's waters. Presented at Discover URECA @ NTU poster exhibition and competition, Nanyang Technological University, Singapore.</td>
</tr>
<tr>
<td>Date</td>
<td>2012</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10220/9034">http://hdl.handle.net/10220/9034</a></td>
</tr>
<tr>
<td>Rights</td>
<td>© 2012 The Author(s).</td>
</tr>
</tbody>
</table>
At any one time, there are about 1,000 vessels in the Singapore port and every 3-4 minutes sees 1 ship arrive at or leave Singapore.

Today, there are about 5,000 maritime companies contributing more than 7% to Singapore’s gross domestic product, and employs more than 170,000 personnel.

About a million visitors cruise into Singapore annually in a continuous cycle.

As Singapore is an island state, maritime security is an important factor as maritime based threats represent a very real threat against our island. Furthermore, as Singapore’s ports are one of the busiest in the world, confidence in the security of our waters is of economic importance to Singapore.

Data Analysis

A sample dataset of 2 million Automatic Identification System (AIS) records was provided, consisting of data on ferries that ply our waterways.

- 80 ferries (Feb 09 – Mar 09)
- Data on callsign, course over ground, speed, coordinates, etc

Problems Identified

- MMSI identifiers missing
- Data fusion inconsistencies
- Variable update rates
- Isolating individual ‘trips’

Methods

Trip Segmentation

Ship movements were split based on trips from point to point. Various thresholds are considered to decide at which points to split the time series of movements.

Mining Algorithm

A sequential pattern mining algorithm is employed to discover the movement patterns of ships.

Key Tasks

1. Determine and improve data quality of the input sources
2. Discover movement patterns of vessels in and around Singapore’s waters
3. Flag abnormalities by comparing behavior of ships to established patterns that have been mined