Simula Research Laboratory

SARCOS Report

Machine Learning Models for the Selection of Security Test Scenarios

Authors: Arnaud Gotlieb, Pierre Bernabé, Dusica Marijan, Helge Spieker

Delivery Date: 15 December 2019

Abstract

Transport infrastructures are vulnerable to cyber-attacks and the detection and prevention of these attacks represent a big challenge in software security. According to our knowledge, general traffic surveillance and control in the maritime domain have not yet reached the same level of maturity, in terms of cyber-security, than it has reached in the domain of air traffic management and control. So, there is a pressing need to analyse the dangers in this area and propose advanced means of cyber-attacks detection and prevention. Vessel traffic monitoring and control rely mostly on unauthenticated and unencrypted messages transfer that renders these services vulnerable to cyber-attacks. Typical attacks such as false data injection attack are difficult to detect as they alter the semantics of the data (e.g., by adding/removing/multiplying elements on a real-time control equipment), while preserving the syntactical correctness of the messages. Identifying these attacks and classifying them as malicious anomalies or unintentional false data, has become a major challenge for traffic monitoring authorities.

This report aims at analyzing existing threats and anomaly detection methods based on the standard AIS (Automatic Identification System) communication protocol. It explores the usage of Machine Learning techniques that can be leveraged in the automatic detection of false data injection attack in AIS. By focusing on the case of maritime surveillance, and by making use of AIS datasets and other sources of data, the report draws perspectives on the benefice of Machine Learning -based detection of false data injection attacks.

The report is organized in five chapters: Chap.1 briefly introduce the maritime traffic surveillance context; Chap.2 introduces the necessary background to understand the report; Chap.3 reports on five known data-driven attacks on AIS; Chap.4 briefly present a logical structure for a future automatic anomalies detection system; Chap.5 examines three research leads in the area of software security for the traffic surveillance area.

Contents

1 Context **3**

2 Background **4**

2.1 Automatic Identification System (AIS) 4

2.2 Satellite-based AIS (S-AIS) 5

2.3 Data Access 5

3 Known Data-Driven Attacks **7**

3.1 Switch Off AIS (SOA) 7

3.2 Ship Teleportation (TEL) 8

3.3 Ships Spoofing (SPO) 8

3.4 Illegal Ship Rendez-Vous (IRV) 9

3.5 Ship Piracy (PIR) 9

4 Using Machine Learning to Detect Data-Driven Attacks **10**

4.1 Message-Based Detection 10

4.2 Ship-Based Detection 11

4.3 Global Traffic Awareness Detection 12

5 Perspectives **15**

5.1 Exploiting Multi-Sources AIS Data 15

5.2 Exploiting an AIS Anomaly Generation Tool (T-FDI) 16

5.3 Exploiting Additional Data Sources 16

Chapter 1

Context

Traffic Surveillance Systems are modern and complex systems which aim at monitoring and controlling traffic of vehicles, ships, aircraft, etc. These systems tackle more and more sources of data coming not only from the traffic observation through radio-based communication, radar information systems, satellite earth observation data, etc. but also from environmental data (weather forecast, maritime currents, etc.), regulation by control authorities, and sometimes also political decisions. The complexity of these systems, which have become more and more intelligent over the last past years, have render them more vulnerable to attacks.

Maritime traffic surveillance has several objectives including ship identification and control, trajectory collision prevention, Search And Rescue (SAR) missions, and general traffic regulation. In addition, modern traffic surveillance systems also aim at detecting illegal activities such as smuggling, illegal fishing, Exclusive Economical Zone (EEZ) intrusion, illegal transshipment, maritime pollution monitoring, etc. Hence, attacks against traffic surveillance systems need to be identified and classified in order to detect and prevent these illegal activities. False data injection attacks, already mentioned in the report, are particularly difficult to detect when they are performed and implemented by using realistic scenarios.

Chapter 2

Background

This chapter introduces the necessary background information about available sources of data and about Automatic Identification System, which is mainstream in vessel traffic surveillance.

2.1 Automatic Identification System (AIS)

AIS is an automatic identification and tracking system used in the maritime domain, to collect information and identify ships. It is used by ship crew and coast guards to monitor maritime activities. For this, ships must be equipped with a transceiver that can send and receive AIS messages. This equipment is mandatory for all vessels with more than 300GT and above, for all ships which are engaged into international journeys, and for all passenger ships. For all other vessels, cheaper and less powerful transmitters are suggested, but not required.

AIS message are composed of static and dynamic information. Static fields include the **MMSI** number[1], which provides an international standardized number for vessel identification, the **IMO** identification number, the **vessel name**, the **call sign**, the **length**, **width** and **vessel type**. These items are entered manually by the ship captain into the AIS transmitter and static information is automatically transmitted on the broadcast channel, every 6 minutes. AIS transmitters also send dynamic information every 2 to 10 seconds depending on the vessel speed, or every 3 minutes if the vessel is at anchor. Dynamic information includes **navigation status**

[1] Maritime Mobile Service Identity

(e.g., "at anchor", "fishing", etc.), **vessel position** (latitude LAT, longitude LON); **vessel speed** (SOG), course over ground (COG) which is the vessel direction w.r.t. the ground (relative to the north pole), **heading** which is the direction (relative to the magnetic north pole or the geographic north pole), and timestamps. All AIS messages do not contain the same information and they are not always sent at regular time-stamps. Typically, AIS messages have a range of about 20km to 40km. The limitation of this range is due to the earth curvature and the height at which the antenna is installed on ships.

2.2 Satellite-based AIS (S-AIS)

For about ten years, coastal guards and operating companies have performed AIS-messages detection with satellites. The main advantage is the ability to reach the whole globe. The fundamental challenge for AIS satellite operators is to manage a very large number of individual AIS messages simultaneously.

There is an inherent issue within the AIS standard: The radio access scheme defined in the AIS standard, creates 4,500 available time-slots in each minute. However it can be easily overwhelmed by the large satellite reception footprints and the increasing numbers of AIS transceivers, resulting in message collisions, which the satellite receiver cannot process. Some ports in the Mediterranean sea, in the North sea and on the Chinese coasts are difficult to supervise with this technique.

Interestingly, data fusion between traditional AIS and S-AIS allows coastal guards to increase their level of confidence into the monitored zone. However, other source of data may also be relevant to complement or correct AIS-based vessel tracking, such as maritime current streams, weather observation and forecast, satellite geophysical observational pictures, etc. This is debated in the last chapter of this report.

2.3 Data Access

Accessing to AIS data can be performed using various sources. Some websites let users access to real-time observation of AIS messages. Among these websites, the USA coastal guard is the most complete with historical data from 2009 to now, across the entire US coastline [2]. Interestingly, this open dataset has been cleaned up from spurious or incoherent messages.

In maritime surveillance, sea and coastal beacons collect AIS message and transmit them to the coastal guard. This is the main source of information for the guards, even though the scope of surveillance is necessarily limited due to the earth curvature. As said above, satellite-based AIS message can also be collected by deployed satellites. For instance, Norway operates four such satellites, which have, as principal mission, to collect AIS message for the coastal guard.

[2] `https://marinecadastre.gov/ais/`

Chapter 3

Known Data-Driven Attacks

This chapter reviews scenarios[1] that are of interest for coastal guards. As the entire ocean cannot be scrutinized, the maritime traffic surveillance is usually limited to specific areas of interest (e.g., ports, maritime boarders or fishery zones). Even by limiting the surveillance to these zones, mobilizing the sufficient man-power for a systematic surveillance is impossible as just too costly. Therefore, there is an increasing pressure to find automated means and tools for detecting these attacks.

The data-driven attacks presented below can be classified as false data injection attacks, which consist in altering the semantics of the data while preserving their syntactic correctness and logical coherence. They are data-driven as their goal is to falsify one or several sources of data in order to fool the maritime surveillance traffic.

After having reviewed the existing literature and exchanged with domain experts, we identified five main data-driven attacks.

3.1 Switch Off AIS (SOA)

By voluntarily switching off the AIS (SOA) transmitter, a ship captain can perform an illegal maneuver with the wish to stay invisible. The objective is multiple and can, for instance, be motivated by the willingness to enter illegally into a forbidden fishing zone, or to meet another ship for trans-boarding illegal cargo. For the observer, the ship has just missed to send one or several AIS messages, which happens sometimes and cannot be systematically classified as illegal activity without further investigations.

[1] Scenarios are called ”attacks” in this report for the sake of clarity.

Often, in these scenarios, AIS messages are re-emitted after a while, and the missing messages remain totally undetected.

The systematic detection of SOA vs missing AIS messages turns out to be difficult and requires a thorough analysis. However, some indicators can be used to classify a situation as potential malicious SOA behavior:

- The route followed by the ship, as indicated by his last messages, has dramatically changed during the period when no message were emitted ;
- The ship has stopped to emit messages while entering into a very active zone, where several other ships are located ;
- The ship is entering a fishing zone, or an exclusive economic zone (EEZ).

These indicators, that can be converted into detection rules, are of course not always sufficient to classify a situation as illegal activity (SOA) but they are used as signals to launch further investigations.

3.2 Ship Teleportation (TEL)

This attack is characterized by a trajectory anomaly: A ship sends a new message indicating a position that is totally unrelated with the previous known trajectory of the ship, i.e., the ship trajectory contains a "teleportation". This attack differs from SOA because it entails a voluntary action to modify the sent message. It is malicious action which aims to voluntarily obfuscate the actual behavior of the ship and/or to fool the traffic monitoring and control.

3.3 Ships Spoofing (SPO)

Ships spoofing consists in sending multiple AIS messages from a given zone to fabricate a situation where multiple ships are indicated, while they are actually much less. This attack intends to fool the traffic control and monitoring, and can possibly lead to interrupt the traffic control in a entire area. The motivation behind SPO can be to hide an illegal activity, to saturate the access control of a port or to divert the attention of the coastal guards.

The SPO attack can be conducted using multiple channels, ranging from the generation of random AIS message (easy to detect) to the generation of realistic AIS messages showing a viable situation (difficult to detect).

3.4 Illegal Ship Rendez-Vous (IRV)

A ship *rendez-vous* at sea is a maneuver where two ships meet so that they can exchange cargos and/or passengers. These maneuvers and cargo transfers are usually forbidden at sea and may result from criminal activities. When these rendez-vous have not been declared, they are classified as illegal ship rendez-vous (IRV). Depending on the ship navigation ranges, these maneuvers are usually held at different distances from the shore. Two types of rendez-vous maneuvers are observed:

- **Non-obfuscated:** The ships keep their AIS transmitters on, in normal operating mode. The rendez-vous can be observed;
- **Obfuscated:** The ships send falsified trajectories such that they tend to avoid each other. The amount of effort spent to define the appropriate AIS messages to send in order to fool the maritime surveillance, is proportional to the difficulty to detect these IRV;

3.5 Ship Piracy (PIR)

Previous attacks aimed at fooling the maritime traffic surveillance, but other attacks are directed towards ships. Ship piracy (PIR) aims at fooling the the surveillance capacity of a ship in order to take control over it. A typical motivation behind this attack is to ransom shipowners or countries with hostages capture, but there are other motivations such as cargo robbery or hacker fame. Note however that on-board surveillance systems are difficult to fool with AIS message because they use radar and eye-control in parallel to AIS monitoring. Even though analysing the trajectory of a ship only with visual control is very uncertain, the multiplication of detection sources (AIS, S-AIS, radars, contacts with the maritime surveillance, etc.) renders this attack more difficult to execute.

The implementation of these data-driven attacks by using false data injection attack into AIS and S-AIS needs to be investigated and is currently outside the scope of this report. However, our initial investigation of using Machine Learning to detect false data injection attack lead us to a description, given in the next chapter.

Chapter 4

Using Machine Learning to Detect Data-Driven Attacks

Here we examine the usage of Machine Learning to detect some data-driven attacks. Our analysis is preliminary as it only presents initial ideas of the models to use for detecting specific data-driven attacks. We structure the presentation at three distinct levels:

1. **Message-Based Detection:** Falsified data are injected into a single AIS or S-AIS message and the process only aims at detecting the falsified message and data ;
2. **Ship-Based Detection:** A sequence of AIS messages related to a single ship is falsified in a consistent way ;
3. **Global Traffic Awareness Detection:** Messages related to several ships are fabricated, which lead to the modification of a complete situation ;

In the following, we detail these three types of detection.

4.1 Message-Based Detection

Detecting a falsified message is not really difficult when it is considered in a succession of existing messages and within a coherent environment. For example, one can detect a falsified message by checking:

- the position, type or status of the vessel as given by the AIS message, with respect to existing maps. A typical example is to find a vessel located in the middle of a city. It is not totally impossible but very unlikely ;

- the vessel speed as given by the message, as related to its type. It is very unlikely that a container-ship ship will run at 50knots ;
- ...

In a recent NATO report [3], a full list of anomalies in AIS message is proposed and hard-coded rules are indicated to detect them. All these rules are based on incoherent value detection with respect to general knowledge or domain-specific ontology. They can obviously be fooled by attackers who create realistic messages. Nothing prevents an attacker from using the actual characteristics of a vessel and use those to falsify AIS messages to ensure that no aberrant speed value or position is given. Another difficulty lies in the lack of rigor of some ship captains while initializing the AIS transceiver. So, an appropriate detector should be able to perform a triage in between messages with anomalies due to lack of rigor and malicious intention.

Anomalies detection in a single message could be better handled by data-driven methods such as Machine Learning. Indeed, by training appropriately models, it might be possible to classify anomalies in a much better way than using hard-coded rules.

4.2 Ship-Based Detection

Here, anomalies are considered in the context of a ship journey, which means that it is a sequence of AIS messages which is analyzed. For example, falsified scenarios can be detected by checking:

- the position. Potentially detected anomalies include SOA and TEL. Of course, combining the data with maps showing ports, EEZ, marine borders, etc is advantageous here and can bring great precision in the anomaly detection and priority ;
- the time-schedule. Potential anomalies that can be detected include SOA and TEL and are related to possible interruptions in the transmission of messages ;
- the trajectory. Here, the potential anomalies include TEL, IRV and PIR. By looking at a single vessel trajectory, it seems difficult to detect attacks such as IRV and PIR which involve at least two ships. But, by crossing models and anomalies, one can possibly give signals of IRV

and PIR in some cases. By combining the analysis of trajectories with known models of vessels, the anomaly detection can be very precise and give interesting results ;

- the status. The anomalies that can be detected include TEL and PIR, even though the complexity of PIR is probably too high for a precise detection. However, status anomalies indicate potential problems that can be voluntarily sent by ship captains, leading to their potential detection.

The usage of ML to detect these anomalies is pertinent as models can be trained with coherent data and in particular, with algorithms such as Long Short-Term Memory (LSTM) that take into account the time or a probability map to follow the evolution of positions during a trip.

Many research work have been devoted to the detection of anomalies at that level. In these works, one finds the following algorithms: Cell Grid Architecture [5], Bayesian network [1], Gaussian model [1], Trajectory prediction [2], Convolution Neural Network (CNN) [6], Predicting Destinations by Nearest Neighbor [8].

However, even if these algorithms and the deployed methods are quite advanced, they do not focus on the detection of attacks but rather, on the detection of anomalies. An experienced attacker would have no trouble creating scenarios where automatic detection is ineffective. By recording all the trip messages beforehand, adding some noise and sending the data corresponding to recorded trip, one can easily create an attack scenario where an actual ship perform illegal actions. Note however that using only AIS renders this attack impossible to detect without using other source of data or data coming from other ships.

4.3 Global Traffic Awareness Detection

The type of anomalies which are sought here depends on a global awareness of the maritime traffic. By examining all the AIS message available for a given zone and time slot, it becomes possible to detect attacks such as IRV, SPO and PIR, i.e., anomalies which involve several ships and trajectories. However, the volume of AIS data which is necessary for training multi-tasks

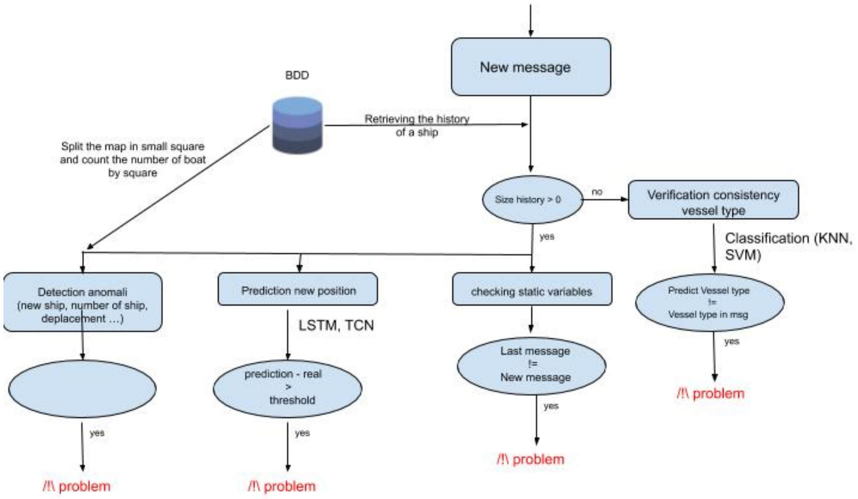
New message
BDD
Retrieving the history of a ship
Split the map in small square and count the number of boat by square
Size history > 0
no
Verification consistency vessel type
Classification (KNN, SVM)
yes
Detection anomali (new ship, number of ship, deplacement ...)
Prediction new position
checking static variables
Predict Vessel type != Vessel type in msg
LSTM, TCN
prediction - real > threshold
Last message != New message
yes
yes
yes
yes
/!\ problem
/!\ problem
/!\ problem
/!\ problem

Figure 4.1: AIS Message Anomaly Detection

models, is currently considered too high for producing models of reasonable accuracy [7]. Here, only Deep Learning models are appropriate and their training requires computational resources and data storage which are not yet accessible to a regular development environment. We are unaware of any approach and work which has reported the usage of Machine Learning for operational traffic surveillance systems which scale-up to global traffic awareness issue detection [9].

In Fig.4.1, we draft a logical structure for an anomaly detection system based on machine learning. Processing a new message would lead to four distinct verification and potential anomaly (or "`/!\ problem`") detection using simple checks against historical data. By checking the vessel type which is produced by ML classification models such as KNN or SVM, against the reported type in the AIS message, one could operate a first anomaly detection. This detection will not used any other data than the one contained in the AIS message. If (sufficient) historical data about the concerned ship is available, then other checks could be performed such as the ones that target

SOA and TEL. Here, only the static data information items of the AIS message would be considered. Based on dynamic items, such as position and speed, another more sophisticated model based on LSTM could be used to predict the next expected position of the concerned vessel and compare it the current AIS message. This would allow us to detect potential anomalies such as IRV and PIR. Finally, a global traffic awareness model, if properly trained, could be eventually exploited to detect SPO. Fig.4.1 depicts a simple architecture for which initial experiments have been undertaken and have produced initial results reported in the presentation joint to this report.

Chapter 5

Perspectives

This chapter discusses of four perspectives that, according to our initial experiments, could be of interest to be explored.

5.1 Exploiting Multi-Sources AIS Data

By examining satellite data provided by StatSat A.S., which is the Norwegian operator of AIS satellites, we observed that this dataset is incomplete, especially in Europe and China near the coast and the ports, where traffic is relatively important. This is due to the inherent range constraints coming from earth-observation satellites. Hopefully, this source of data can be advantageously combined with AIS data collected by the coastal guards, through terrestrial and sea beacons. Even though, these beacons can only have a range analysis up to about 40km from the coast, they offer us a relatively comprehensive database of billions of AIS messages. So, combining these two sources of data (earth-observation satellite S-AIS data and beacon-based AIS), we believe that the scope of anomalies detection could scale up to an acceptable level.

However, there are several challenges related to data quality, such as the irregularity of AIS messages sent by vessels, the noise in AIS messages non-voluntarily introduced by ship captains which leads to confound ships, etc. These challenges are known in the literature and initial solution are depicted, e.g., by interpolating the missing AIS messages [4]. Dealing with these issues is crucial, but could lead, in case of success, to a deployed verification system.

5.2 Exploiting an AIS Anomaly Generation Tool (T-FDI)

Most of anomaly detection tools for AIS that we have reviewed in the literature exploit only unsupervised ML. This is due to the absence of available labels on AIS messages or sequence of messages. To the best of knowledge, there is no publicly available datasets containing anomaly-labeled information.

By exploiting an existing AIS anomaly generation tool, e.g., T-FDI distributed by University of Bourgogne Franche-Comté, we could train models with massively generated falsified scenarios. The availability of this tool is a strong asset because it would allow us to train models (also DL models) for the effective and efficient detection of AIS anomalies. Note however that a typical bias using such a tool is to train the detector to detect anomalies generated by the tool and not real issues. A mitigating solution could be to set up a General Adversarial Network which is train hand-by-hand with the detector. This approach is an interesting research lead in this context which, according to our knowledge, has not yet been explored.

5.3 Exploiting Additional Data Sources

Using only AIS and S-AIS data may not be enough to create an effective anomaly detector. Since attacks are expected to be as realistic as possible, at some point, there is a need to use additional data sources to detect complex attack scenarios. For example, a subtle attack as the one we mentioned above, consisting in recording a previous trajectory in order to replay it with small, almost undetectable modifications, may not easily be detected by using only AIS data. Indeed, in this case, all AIS messages would be consistent with respect to some previously observed trajectories. However, an interesting lead is to detect such anomalies by using other source of data, such as:

- Meteorological data regarding maritime currents. Indeed these data would be interesting to improve the precision of position prediction as they have an impact on the ship trajectories. Though, the availability of these data needs to be investigated ;

- Earth-observation images from satellites. The European Union's earth observation program "Copernicus"[1] is a potential source of such data.

[1] https://www.copernicus.eu/en

Images of the ports and zones at sea could form an interesting source of data to complement anomaly detection based on AIS. However, it should be noticed that anomaly detection must be operative 24-hours a day and these images are not always available due to weather conditions (clouds, storms, etc.) ;

- Video surveillance cameras in ports could form an interesting source of additional data as they could help avoid misclassification related to vessel types. There are accessible source of such data[2] which allow users to collect information about traffic under the form of videos.

Of course, this list is not exhaustive and other source of data can be exploited as well. But, we think that these three sources represent the most promising leads for anomaly detection in the maritime surveillance traffic area.

[2] `https://datafromsky.com/`

Bibliography

[1] Lane R. O. et al. “Maritime Anomaly Detection and Threat Assessment”. In: *13th Intl. Conf. on Information Fusion (FUSION)*. Edinburgh, UK, 2010, pp. 1–8.

[2] Lokukaluge P. Perera et al. “Maritime Traffic Monitoring Based on Vessel Detection, Traking, State Estimation, and Trajectory Prediction”. In: *IEEE Transaction on Intelligent Transportation Systems* 13.3 (2012).

[3] Dominik Filipiak et al. “Anomaly Detection in the Maritime Domain: Comparison of Traditional and Big Data Approach”. In: *NATO Science and Technology Organization (STO)* (2015).

[4] Van-Suong Nguyen, Nam-kyun Im, and Sang-min Lee. “The Interpolation Method for the missing AIS Data of Ship”. In: *Journal of Navigation and Port Research* (2015).

[5] Ciprian Amariei et al. “Cell Grid Architecture for Maritime Route Prediction on AIS Data Streams”. In: *arXiv:1810.00090v1* (2018).

[6] Kwang-Il Kim and Keon Myung Lee. “Deep Learning-Based Caution Area Traffic Prediction with Automatic Identification System Sensor Data”. In: *Sensors (Basel)* 18.9 (Sept. 2018).

[7] Van Duong Nguyen et al. *Multi-task Learning for Maritime Traffic Surveillance from AIS Data Streams*. Tech. rep. arXiv:1806.03972 and hal-01808176. Aug. 2018.

[8] Valentin Roşca et al. *Predicting Destinations by Nearest Neighbor Search on Training Vessel Routes*. Tech. rep. arXiv:1810.00096v1. 2018.

[9] Shilavadra Bhattacharjee. *Automatic Identification System (AIS): Integrating and Identifying Marine Communication Channels*. Tech. rep. 2019. URL: `www.marineinsight.com/marine-navigation/automatic-identification-system-ais-integrating-and-identifying-marine-communication-channels/amp/`.