Detection of malicious AIS position spoofing by exploiting radar information

Fotios Katsilieris
NATO STO Centre for Maritime Research and Experimentation
La Spezia, Italy

Paolo Braca
NATO STO Centre for Maritime Research and Experimentation
La Spezia, Italy
Email: Braca@cmre.nato.int

Stefano Coraluppi
Compunetix Inc.
2420 Mosside Boulevard
Monroeville, PA 15146 USA
Email: Stefano.Coraluppi@compunetix.com

***Abstract*—The Automatic Identification System (AIS) is an automatic tracking system based on reports provided by the vessels carrying an AIS transponder. The reports contain information on the vessel position, velocity *etc.* and typically have high accuracy. Given that the AIS is a self-reporting system, the trustworthiness of positional information depends on data being reported by the vessel, rather than measured by a sensor. Any self-reporting system is prone to "spoofing" or the intentional reporting on incorrect information. This paper addresses the inference problem of whether the received AIS data are trustworthy with the help of radar measurements and information from the tracking system. This problem can be treated in the hypothesis testing framework where the null hypothesis is that the AIS data are trustworthy and the alternative hypothesis is that the data are spoofed. The proposed solution, the generalized version of the sequential log-likelihood ratio test, is compared to the ideally optimal solution using real and simulated data.**

I. Introduction

Maritime situational awareness, which includes accurate knowledge of moving vessel location, has increased the focus on the development of data fusion algorithms. These algorithms can fuse data from several heterogeneous systems in order to provide a better perception of the activity close to the shores of a nation. For example, data from coastal radars, the Automatic Identification System (AIS), video and infrared surveillance systems and SAR systems can be fused. A goal of these systems, in addition to tracking the present vessels, is to introduce some sort of intelligence in the surveillance systems to automatically identify possibly suspicious (also called anomalous) behavior. Some example applications include the deviation of vessels from the known shipping lanes, the *rendezvous* of vessels at sea, the motion of fast moving vessels close to the shore, methods to automatically detect the switching-off of the AIS transponder and others. Some examples from the relevant bibliography are [1], [2], [3], [4], [5].

A common characteristic among the given examples is that they rely on the high accuracy of the AIS reports in order to derive training patterns or the "ground truth" for the motion of the vessels. An often hidden assumption is that the AIS reports are trustworthy. This paper is based on the assumption that AIS reports can be falsified (or *spoofed*) as suggested in [6], [7]. The trustworthiness of the AIS reports depends on the willingness of the crew of the ship to report their true data.

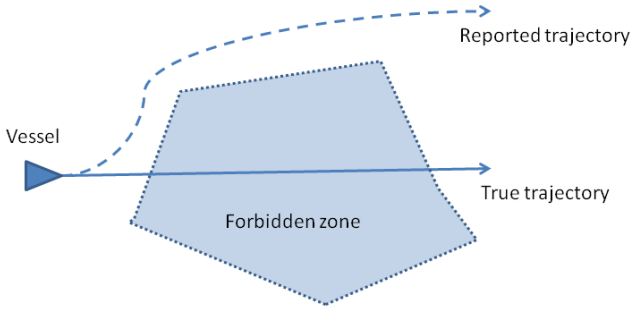

Fig. 1: A fishing boat violates a no-fishing zone but reports a different trajectory.

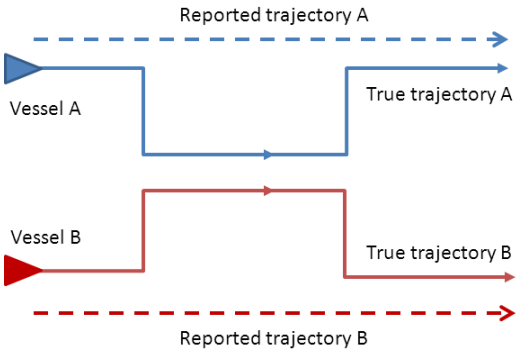

Fig. 2: Two vessels meet in the sea for an illegal transaction and report different trajectories in order to conceal their *rendezvous*.

Fig. 1 shows an example of interest for this problem. Consider an area that specific types of vessels are not supposed to enter, *e.g.*, an area where fishing is not allowed. In this case, a fishing boat could enter the no-fishing zone, following the trajectory denoted by the solid line, while reporting that it is bypassing it, reporting the trajectory denoted by the dashed line. Fig. 2 shows an example where two vessels meet but try to conceal their actions by reporting false trajectories.

In case of correspondence please contact all authors.

F. Katsilieris is currently a Marie Curie fellow at Thales Nederland B.V. in the context of the *MC IMPULSE* project (EU 7^{th} FP under grant agreement n° 238710) and a PhD student at the University of Twente, the Netherlands. This work has been performed at NATO-STO CMRE in the context of the *Visiting Researcher Program* while F. Katsilieris was on temporary leave from Thales. F. Katsilieris can be reached at Fotios.Katsilieris@nl.thalesgroup.com

This paper addresses the problem of determining whether the AIS data received from a vessel are trustworthy or not by using information from additional sensors. The proposed methodology uses radar measurements and prior information from the corresponding tracking system. In the case that the AIS data are indeed trustworthy, they can be safely used in the data fusion algorithms, *e.g.*, for enhancing the tracking accuracy. If the AIS data are estimated to be spoofed, then their fusion with other data can be avoided and an anomaly can be flagged to the operator of the surveillance system.

The rest of the paper is organized as follows: the problem is formulated in Section II; the proposed solution is developed in Sections III and IV. Simulations and experimental results are presented in Section V. Finally, the conclusions are drawn in Section VI.

II. Problem formulation

As explained in the introduction, the considered problem amounts to determining whether the AIS data transmitted by a vessel is spoofed or not by using the measurements from any available radar and the predicted vessel position according to the tracking system of the radar.

First, some notation will be introduced. Subsequently, the fundamental assumptions will be stated and finally, the problem will be posed in the statistical hypothesis testing framework.

A. Notation

The following notation will be used throughout the paper:

- the two-dimensional true target position in Cartesian coordinates will be denoted as $\mathbf{x}_0$, the predicted target position as $\mathbf{x}$ and the AIS reported position as $\mathbf{x}_{AIS}$;
- there is a number of K radars measuring the position of the target at each time instance t and the measurement of each radar is denoted as $\mathbf{z}_j(t), j \in [1, \dots, K]$. From now on, the time index will be suppressed. The radar measurements are usually expressed in polar coordinates, *i.e.*, target range r, bearing θ and possibly target range rate $\dot{r}$. For simplicity reasons, the Cartesian system will be used for the radar measurements with corresponding transformation of the measurement covariance matrix.
- $\mathcal{N}(\mathbf{x}; \mu, \Sigma)$ denotes the two-dimensional Gaussian probability density function (pdf):

$$\mathcal{N}(\mathbf{x}; \mu, \Sigma) = \frac{1}{(2\pi)\sqrt{|\Sigma|}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1} (\mathbf{x}-\mu)} \quad (1)$$

- The symbol $\sim$ stands for "is distributed according to".

B. Assumptions

For formulating and tackling the problem, the following assumptions will be made:

- Zero false alarm rate at the detector of the radar tracker will be assumed because in practice most radar systems operate at low false alarm rate regimes.
- The *a priori* information on the vessel position is a Gaussian with mean $\mathbf{x}_0$ and covariance matrix $\Sigma_{\mathbf{x}}$: $\mathbf{x} \sim \mathcal{N}(\mathbf{x}_0, \Sigma_{\mathbf{x}})$. This represents the predicted position of the vessel according to the tracking system at the same time instance as the radar measurements and the AIS contact are received.
- The measurement from the kth radar is $\mathbf{z}_k = \mathbf{x}_0 + \mathbf{w}_k$, where $\mathbf{w}_k$ is additive Gaussian noise with zero mean and covariance matrix Σ_R: $\mathbf{z}_k \sim \mathcal{N}(\mathbf{x}_0, \Sigma_R)$. Furthermore, it is assumed that the measurements are conditionally independent from radar to radar.
- The trustworthy AIS data $\mathbf{x}_{AIS}$ follow a Gaussian distribution with mean $\mathbf{x}_0$ and covariance matrix Σ_{AIS}. When the AIS data are spoofed, it is assumed that an arbitrary bias $\mathbf{d} = [d_x, d_y]^T$ is added to the true position of the vessel, *i.e.*, the mean is now $\mathbf{x}_0 + \mathbf{d}$.
- The perfect reception of the AIS report is assumed. In other words, the transmitted AIS report is always received by the corresponding tracking system.
- Typically, it holds that the elements of Σ_{AIS} are smaller than the elements of Σ_R.
- Only the single-target case is considered. This approximation is valid in multitarget scenaria when there is a perfect data association scheme or when the targets are sufficiently separated and the single-target case can be reconstructed. In the maritime domain, this approximation is valid when the vessels are outside from a port, which is the case where the AIS spoofing is of interest.
- The mathematical derivation of the joint likelihood of the radar measurements and AIS contacts is done for the static case, *i.e.*, in a snapshot time. When the sequential detection of spoofing will be considered, the measurements will be correlated in time [8] but they will be approximated to be independent. The validity of this approximation will be demonstrated by three examples using real and simulated data.

Note that the assumptions of Gaussian prior and Gaussian measurement noise are very common in the target tracking context [8].

C. Statistical hypothesis testing of AIS spoofing

The AIS spoofing detection problem can be formed as a statistical hypothesis testing problem:

- $H_0 : (\mathbf{z}, \mathbf{x}_{AIS}) \sim p(\mathbf{z}, \mathbf{x}_{AIS}|H_0)$ is the *simple* null hypothesis that the AIS data are trustworthy, versus
- $H_1 : (\mathbf{z}, \mathbf{x}_{AIS}) \sim p_{\mathbf{d}}(\mathbf{z}, \mathbf{x}_{AIS}|H_1)$ is the *composite* alternative hypothesis that the AIS data are spoofed,

where $p_{\mathbf{d}}(\mathbf{z}, \mathbf{x}_{AIS}|H_1)$ is the joint distribution of the radar measurements and AIS contacts, parameterized by the spoofing distance $\mathbf{d}$, and $p(\mathbf{z}, \mathbf{x}_{AIS}|H_0)$ is the joint distribution of the radar measurements and trustworthy AIS contacts.

In testing H_0 versus H_1, there are two types of error that can be made: H_0 can be falsely rejected or H_1 can be falsely rejected. The first type of error is called a *false alarm*, meaning that trustworthy AIS data are classified as spoofed, and the false alarm probability is denoted as P_{FA}. The second type of error is called a *missed detection* and it means that the spoofing of the AIS data has not been detected. The missed detection probability P_{MD} is equal to one minus the probability of detection, or $P_{MD} = 1 - P_D$.

During the design process of the test for H_0 versus H_1, one has to find a good trade-off between the two error probabilities, since one error can become arbitrarily small at the expense of the other error becoming unacceptably large. It is very common in the radar community to follow the Neyman-Pearson paradigm [9]. Accordingly, a low upper bound on the false alarm probability is set and the miss detection probability is minimized, or equivalently the detection probability is maximized. For a better discussion on the fundamentals of hypothesis testing see [9], [10].

III. Single sample detectors

As a first step, the expression for the clairvoyant[1] likelihood ratio test (from now on called C-LRT) is derived for the case of one radar and K radars. Subsequently, the generalized version of the likelihood ratio test is introduced in order to deal with the unknown spoofing distance.

A. Clairvoyant likelihood ratio test

In the beginning, the pdfs of receiving a given radar measurement and a given AIS report under each hypothesis need to be calculated.

The radar measurements and the AIS data are conditionally independent given $\mathbf{x}$:

$$p(\mathbf{z}, \mathbf{x}_{AIS}|H_i) = \int_{-\infty}^{\infty} p(\mathbf{z}|\mathbf{x})\ p(\mathbf{x}_{AIS}|H_i, \mathbf{x})\ p(\mathbf{x})\ d\mathbf{x} \quad (2)$$

where $i = 0, 1$.

The integral in Eq. (2) can be evaluated analytically in the case of Gaussian measurements and prior because the product of two Gaussian pdfs is an unormalized Gaussian [11]:

$$\mathcal{N}(\mathbf{x}; \mu_1, \Sigma_1) \cdot \mathcal{N}(\mathbf{x}; \mu_2, \Sigma_2) = \varepsilon\ \mathcal{N}(\mathbf{x}; \mu_3, \Sigma_3) \quad (3)$$

where

$$\varepsilon = \mathcal{N}(\mu_1; \mu_2, \Sigma_1 + \Sigma_2) = \mathcal{N}(\mu_2; \mu_1, \Sigma_1 + \Sigma_2) \quad (4)$$

$$\Sigma_3 = \left(\Sigma_1^{-1} + \Sigma_2^{-1}\right)^{-1} \quad , \quad \mu_3 = \Sigma_3 \left(\Sigma_1^{-1}\mu_1 + \Sigma_2^{-1}\mu_2\right) \quad (5)$$

Accordingly, the C-LRT is

$$\begin{aligned}\Lambda(&\mathbf{z},\mathbf{x}_{AIS}, \mathbf{d}) \\ &= \frac{\mathcal{N}(\mathbf{z}; \mathbf{x}_{AIS} - \mathbf{d}, \Sigma_R + \Sigma_{AIS})}{\mathcal{N}(\mathbf{z}; \mathbf{x}_{AIS}, \Sigma_R + \Sigma_{AIS})} \\ &\quad \times \frac{\mathcal{N}(\mathbf{x}_0; \mu_1(\mathbf{z}, \mathbf{x}_{AIS}) - \Delta\mathbf{d}, \Sigma_1 + \Sigma_{\mathbf{x}})}{\mathcal{N}(\mathbf{x}_0; \mu_1(\mathbf{z}, \mathbf{x}_{AIS}), \Sigma_1 + \Sigma_{\mathbf{x}})} \quad \gtrless \tau \end{aligned} \quad (6)$$

where

$$\Sigma_1 = (\Sigma_R^{-1} + \Sigma_{AIS}^{-1})^{-1} \quad (7)$$

$$\mu_1(\mathbf{z}, \mathbf{x}_{AIS}) = \Sigma_1(\Sigma_R^{-1}\mathbf{z} + \Sigma_{AIS}^{-1}\mathbf{x}_{AIS}) \quad (8)$$

$$\Delta = (\Sigma_{AIS}\Sigma_R^{-1} + I_{2\times 2})^{-1} \quad (9)$$

[1]The *clairvoyant* test is an ideal test that knows the true spoofing distance.

In the case of Gaussian measurements and prior, an analytic formula for Eq. (6) can be found, see Appendix A.

The C-LRT can also be generalized to address the case where measurements from several radars are available. The multi-radar likelihood $p(\mathbf{z}|\mathbf{x}_0)$ in the case of Gaussian measurements is evaluated in Appendix B. It can be seen, see Appendix B, that in the case where measurements from similar radars are available, the measurements can be easily aggregated and the C-LRT will have the same form as in Eq. (6).

Eq. (6) is interesting for one more reason. The expected value of the logarithm of Eq. (6) under H_1, *i.e.*, $E[\log(\Lambda(\mathbf{z}, \mathbf{x}_{AIS}, \mathbf{d}))|H_1]$ is equal to the Kullback-Leibler divergence between the nominator and the denominator, see [12, Theorem 11.8.1]. In other words, it measures how much the two hypotheses are disjoint. This expression can also be evaluated analytically, see Appendix A. In the simulations, it is shown that the hypotheses become more disjoint as the spoofing distance increases, which is a desirable property.

B. Generalized likelihood ratio test

In practice, the spoofing distance $\mathbf{d}$ is not known to the system and therefore, the generalized version of the likelihood ratio test (from now on called G-LRT) can be used.

The G-LRT is one of the most powerful tools available for solving composite hypothesis testing problems, such as the problem at hand, where the spoofing distance is not known. For a better discussion on the G-LRT see [10].

When using the G-LRT, one first needs to find the estimate $\hat{\mathbf{d}}$ that maximizes the likelihood under H_1

$$\begin{aligned}\hat{\mathbf{d}} = \underset{|\mathbf{d}|>|\mathbf{d}_{min}|}{\arg\max}\ &[\mathcal{N}(\mathbf{z}; \mathbf{x}_{AIS} - \mathbf{d}, \Sigma_R + \Sigma_{AIS}) \\ &\times \mathcal{N}(\mathbf{x}_0; \mu_1(\mathbf{z}, \mathbf{x}_{AIS}) - \Delta\mathbf{d}, \Sigma_1 + \Sigma_{\mathbf{x}})]\end{aligned} \quad (10)$$

and then use it in Eq. (6) or in the corresponding analytic expression shown in Appendix A.

During the likelihood maximization process, where $\hat{\mathbf{d}}$ is evaluated, a minimum distance $\mathbf{d}_{min}$ is set. If the estimated spoofing distance $\hat{\mathbf{d}}$ has a smaller value than $\mathbf{d}_{min}$ then it is assumed that the vessel is not spoofing. The determination of $\mathbf{d}_{min}$ is important for the performance of the G-LRT and depends on the accuracy of the radar measurements, the AIS data and the prior information. In other words, it depends on $\Sigma_{\mathbf{x}}$, Σ_R and Σ_{AIS}. For instance, if the accuracy is low, then setting $\mathbf{d}_{min}$ to a small value will result to an increased false alarm probability. On the other hand, if the accuracy is high, then setting $\mathbf{d}_{min}$ to a large value will result to an increased miss detection probability.

IV. Sequential detection of AIS spoofing

In the previous subsections, one-sample solutions for the problem at hand were developed. As their name suggests, they make a decision about the trustworthiness of the AIS data using one radar measurement and one AIS report. This approach can also be generalized to using a larger, fixed number of samples and is suitable for applications where the number of

observations is known in advance and no new observations can be made.

In the case studied in this paper, new observations are periodically available according to the radar scanning period and the frequency of AIS reports. Subsequently, it can be noticed that the problem of AIS spoofing detection would ideally be dealt online, as new observations are received and one would like to make a decision with certain error probabilities as fast as possible.

Given the aforementioned discussion, the sequential version of the aforementioned LRTs, from now on called SLRTs, is a more appropriate solution. An SLRT has the property that, in general, it requires a smaller expected number of observations than the fixed number of observations needed by the corresponding fixed sample size test in order to achieve the same error probabilities [13].

The sequential version of the previously described tests will have the general form:

$$\log B < \sum_{t=1}^{N} \log \left[\Lambda(\mathbf{z}(t), \mathbf{x}_{AIS}(t), \hat{\mathbf{d}}) \right] < \log A \tag{11}$$

$$B \geqslant \frac{1-P_D}{1-P_{FA}} \quad , \quad A \leqslant \frac{P_D}{P_{FA}} \tag{12}$$

The test is terminated at the *stopping time* N when $\sum_{t=1}^{N} \log \left[\Lambda(\mathbf{z}(t), \mathbf{x}_{AIS}(t), \hat{\mathbf{d}}) \right]$ exceeds one of the two thresholds or a maximum allowed time has elapsed. In practice, the equalities are used in (12). For a better description of the SLRTs see [13].

Wald's approximations can be used for finding the lower and upper thresholds $-\alpha, h$ for desired error probabilities $P_{FA}, 1-P_D$ under each hypothesis [14]:

$$-\alpha \approx \log \left(\frac{1-P_D}{1-P_{FA}} \right) < 0 \quad , \quad h \approx \log \left(\frac{P_D}{P_{FA}} \right) > 0 \tag{13}$$

Wald's approximations can also be used together with the calculated expected value of the LRTs under the two hypotheses in order to derive the expected number of necessary samples under each hypothesis [14]:

$$E[N|H_0] = \frac{(1-P_{FA}) \log \left(\frac{1-P_{FA}}{1-P_D} \right) - P_{FA} \log \left(\frac{P_D}{P_{FA}} \right)}{E[\log(\Lambda)|H_0]} \tag{14}$$

$$E[N|H_1] = \frac{(P_D) \log \left(\frac{P_D}{P_{FA}} \right) - (1-P_D) \log \left(\frac{1-P_{FA}}{1-P_D} \right)}{E[\log(\Lambda)|H_1]} \tag{15}$$

In the case where a fixed value $\hat{\mathbf{d}}$ is used for all sampling times $t = 1, \ldots, N$, an analytic expression for the expected value of the LRT $E[\log(\Lambda)|H_0], E[\log(\Lambda)|H_1]$ under H_0 and H_1 respectively can be found, see Appendix A.

It should be pointed out that the formulas (11) through (15) are exact only in the static case, where the vessel is not moving and the spoofing distance is constant and known. The reason for the formulas to not be exact in the dynamic case is that the predicted positions of the vessel at every time instance are correlated across time. The motion model used for performing the prediction step in the tracking system introduces the correlation. The motion model typically has the form:

$$\mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t-1), \mathbf{v}(t)) \tag{16}$$

where it can be seen that the state of the target at time instance t is predicted using the state of the vessel at the time instance $(t-1)$ and some noise $\mathbf{v}(t)$. A typical example of a motion model is the *nearly constant velocity model* [15].

In the next section, it will be shown that formulas (11) through (15) are good approximations in the case of a moving vessel as long as the spoofing distance is constant. If the vessel is moving and the spoofing distance is increasing then fewer samples are needed.

V. Experimental results

In subsection V-A, the test statistics for the one-sample LRTs are simulated in order to demonstrate the effect of the spoofing distance and the number of radars on the performance of the various tests. Similarly, the test statistics for the SLRTs are simulated in subsection V-B. Finally, in subsection V-C, the SLRTs are applied to real and simulated data in order to verify the validity of the assumptions and approximations made in the previous sections.

A. Single sample log-likelihood ratio tests

As a first step, it will be demonstrated that the generalized LRT has performance that gets close to the performance of the clairvoyant LRT as the spoofing distance and the number of radars increase. This demonstrates two things: *a*) the benefit of using measurements from multiple radars; and *b*) that as the spoofing distance increases, the two hypotheses become more disjoint and therefore, it is easier to detect the spoofing behavior. The latter phenomenon manifests itself faster when measurements from multiple radars are used.

The settings for the following simulations were $\mathbf{x}_0 = [10^3,\ 10^3]^T$ m, $\Sigma_{\mathbf{x}} = \text{diag}(70^2, 70^2)$ m^2, $\Sigma_R = \text{diag}(50^2, 50^2)$ m^2, $\Sigma_{AIS} = \text{diag}(5^2, 5^2)$ m^2. The spoofing distance was varied as $d_x = [20, 40, \ldots, 160, 175, 200, \ldots, 300], d_y = 0$ m. The minimum distance $\mathbf{d}_{min}$ for the G-LRT has been set to 85 m and the grid points are located in increments of 20 meters in the x and y directions. The statistics were evaluated over 10^4 Monte Carlo runs.

Fig. 3 shows the resulting probability of detection at each value of the spoofing distance for the two tests for varying number of radars and for $P_{FA} = 0.01$ using data in a single time interval. Fig. 4 shows the corresponding ROC curves for $\mathbf{d} = [40, 0]$ m. It can be noticed that the performance of the G-LRT becomes equivalent to the performance of the C-LRT as the spoofing distance and the number of radars increase.

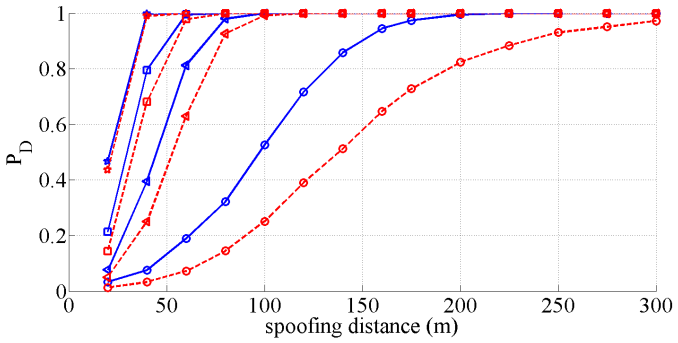

Fig. 3: The probability of detection at different distances using one measurement from a varying number of radars and for $P_{FA} = 0.01$. The C-LRT is denoted by solid blue and the G-LRT by dashed red line. The $\circ$ denotes the use of one, the ∇ of three, the $\square$ of five and the $\star$ of ten radars.

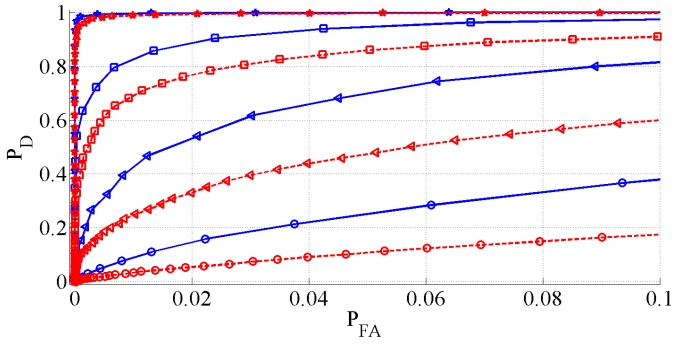

Fig. 4: ROC curves for the two tests for varying number of radars. The spoofing distance is 40 meters in the x direction. The C-LRT is denoted by solid blue and the G-LRT by dashed red line. The $\circ$ denotes the use of one, the ∇ of three, the $\square$ of five and the $\star$ of ten radars.

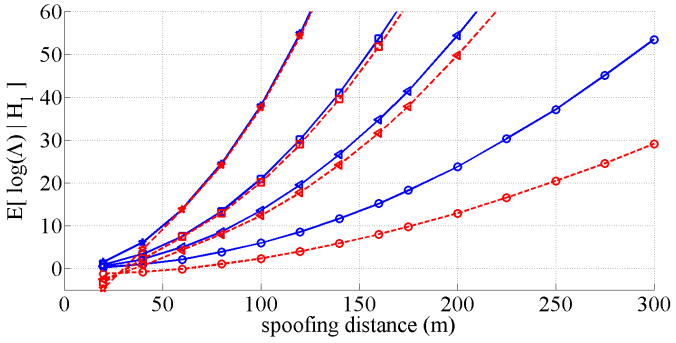

Fig. 5: The performance of the two LRTs under H_1 for varying number of radars. As the spoofing distance and the number of radars increases, the expected value of the tests under H_1 also increases and makes the spoofing detection easier. The C-LRT is denoted by solid blue and the G-LRT by dashed red line. The $\circ$ denotes the use of one, the ∇ of three, the $\square$ of five and the $\star$ of ten radars.

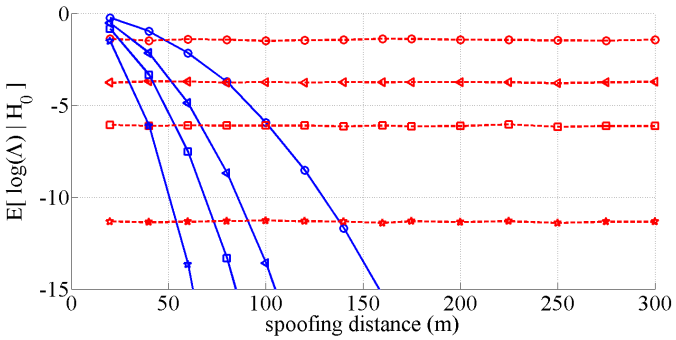

Fig. 6: The performance of the two LRTs under H_0 for varying number of radars. The C-LRT is denoted by solid blue and the G-LRT by dashed red line. The $\circ$ denotes the use of one, the ∇ of three, the $\square$ of five and the $\star$ of ten radars.

Fig. 5 and 6 show how the expected values of the two tests vary under H_1 and H_0 respectively. The settings for these simulations were the same as before. As expected, the longer the spoofing distance and the larger the number of radars that are used, the higher the expected value of the tests under H_1 and therefore, the easier it is to detect the spoofing behavior. In other words, the Kullback-Leibler divergence between the two hypotheses increases and therefore they become more disjoint. It can also be noticed that due to the specific choice of $\mathbf{d}_{min}$, the G-LRT has a negative expected value under H_1 for spoofing distances shorter than 30 to 60 meters, depending on the number of radars used. Using a lower $\mathbf{d}_{min}$ would make the expected value positive at the expense of the expected value of the G-LRT under H_0, which would have smaller absolute value. This effect becomes important when the sequential LRT is used because as the expected value under each hypothesis becomes lower, the corresponding termination time of the test becomes longer for reaching a conclusion with the same error probabilities.

Again, as the number of radars increases under H_1, the G-LRT becomes equivalent to the C-LRT. On the other hand, under H_0 the G-LRT has constant performance that depends on the number of radars.

B. Sequential log-likelihood ratio test statistics

Using the same settings and fixing the spoofing distance at $\mathbf{d} = [80, 80]^T$ m, the thresholds can be varied such that the two tests have the same probability of false alarm. In this way, one can see how long it takes to reach a conclusion with a given probability of detection.

Fig. 7 shows the probability of detection versus the expected termination time of each test. For these simulations, one radar has been used and the probability of false alarm was fixed at $P_{FA} = 10^{-5}$. The results were evaluated over 10^5 Monte Carlo runs. It can be observed that in this case Wald's approximations are not accurate enough because the expected values of the LRTs are too large compared to the evaluated thresholds.

C. Experiments with real and simulated data

In this subsection, the validity of the most important approximations will be checked. In other words, it will be

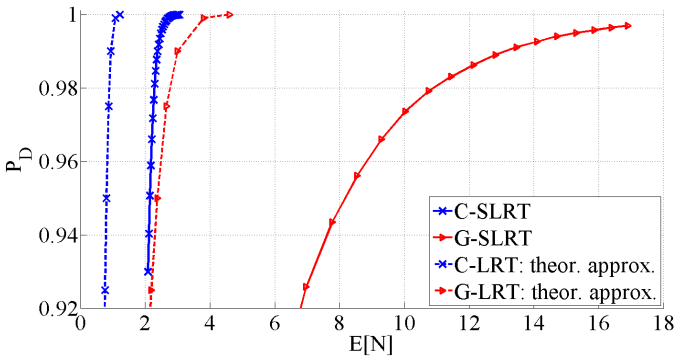

Fig. 7: The probability of detection of the different tests using one radar as a function of the expected number of necessary samples. The false alarm probability has been fixed to $P_{FA} = 10^{-5}$.

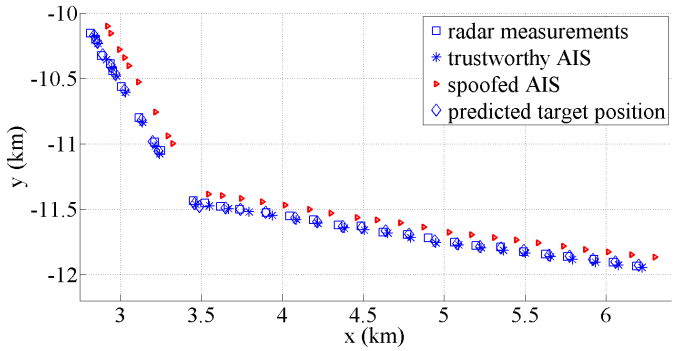

Fig. 8: Example I: Maneuvering target. The collected AIS data are trustworthy and the spoofing is simulated by adding 80 meters in both x and y directions.

shown that the developed SLRTs are valid approximations for and can be used in a dynamic case where the vessel is moving.

The SLRTs were applied to the real data collected from two targets, shown in Fig. 8, 9 and 10. In these scenaria, the two example vessels were sailing in open sea while been observed by a third vessel. The third vessel was recording the radar measurements and tracking the other two vessels. Furthermore, it registered their trustworthy AIS reports. The spoofed AIS data were simulated for the purposes of this work.

The thresholds for the two SRLTs were chosen such that the probability of detection and the false alarm probability would be $P_D = 0.95$ and $P_{FA} = 10^{-5}$ respectively. Subsequently, the termination time of each SLRT was evaluated and compared to the expected number, shown in Fig. 7, in order to validate our assumptions and approximations.

The results are summarized in Table I. In examples I and III, the SRLTs detect the correct behavior of the target.

In the second example, it can be observed that when the vessel tries to report a fake trajectory by increasing the spoofing distance, it is even easier to detect the spoofing behavior. This is a direct consequence of the fact that the two hypotheses become more disjoint as the spoofing distance increases.

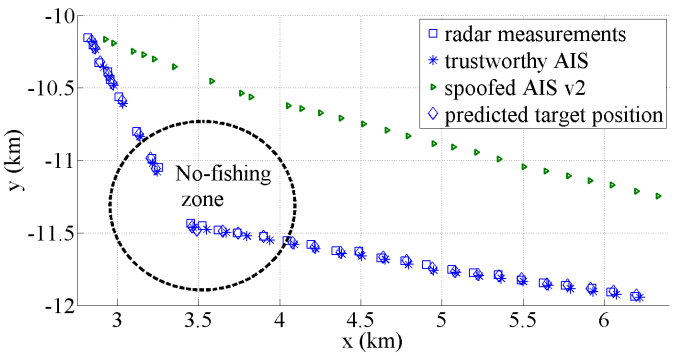

Fig. 9: Example II: The maneuvering target from Example I now spoofs its transmitted AIS data such that it appears that it is not entering a forbidden zone.

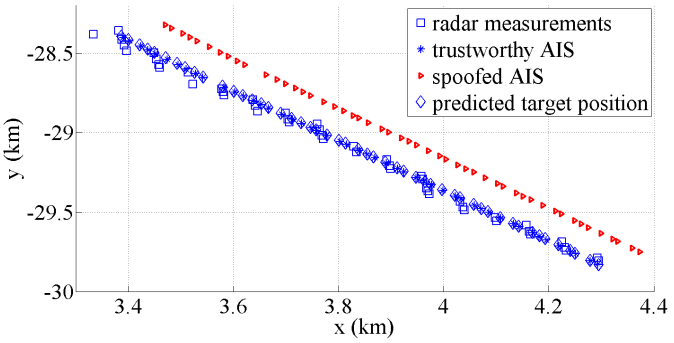

Fig. 10: Example III: A target moving in a straight line. The collected AIS data are trustworthy and the spoofing is simulated by adding 80 meters in both x and y directions.

TABLE I: Necessary number of samples per example such that $P_D = 0.95$ and $P_{FA} = 10^{-5}$

	Example					
	I		II		III	
Test	$E[N]$	N	$E[N]$	N	$E[N]$	N
C-SLRT	2	2	2	2	2	2
G-SLRT	8	22	8	21	8	24

VI. Conclusions

The problem of detecting whether a vessel is transmitting spoofed AIS data was formulated in the context of hypothesis testing. For this scenario, the spoofing distance and the number of radars have been varied in order to obtain the corresponding ROC curves and the expected number of necessary samples for to make a correct decision. The proposed solution was successfully demonstrated using real data with simulated spoofing.

There are also several other problems pertaining to the AIS data transmission, reception and exploitation. Rather than spoofing, vessels could simply turn off their AIS transmitters, possibly periodically, in order to hinder the surveillance systems and their operators from detecting illicit activities. Furthermore, a varying AIS reception probability, for instance due to weather conditions, would further complicate the AIS spoofing detection process [3]. It would be of interest to study

these effects in the context of the solution proposed in the current paper.

Another interesting topic for future research would be to extend the current work such that it can address scenaria with multiple targets. In such cases, collaborative spoofing, for instance swapping of identities, can pose further difficulties in addition to the obvious radar-to-AIS association problem.

APPENDIX A
DERIVATION OF THE CLAIRVOYANT LOG-LRT

The analytic expression of the clairvoyant log-LRT can be found as follows:

$$\begin{aligned}
&p(\mathbf{z}, \mathbf{x}_{AIS}|H_0) \\
&= \int_{\mathbb{R}^2} \mathcal{N}(\mathbf{z}; \mathbf{x}, \Sigma_R)\ \mathcal{N}(\mathbf{x}_{AIS}; \mathbf{x}, \Sigma_{AIS})\ \mathcal{N}(\mathbf{x}; \mathbf{x}_0, \Sigma_{\mathbf{x}})\ d\mathbf{x} \\
&= \int_{\mathbb{R}^2} \mathcal{N}(\mathbf{x}; \mathbf{z}, \Sigma_R)\ \mathcal{N}(\mathbf{x}; \mathbf{x}_{AIS}, \Sigma_{AIS})\ \mathcal{N}(\mathbf{x}; \mathbf{x}_0, \Sigma_{\mathbf{x}})\ d\mathbf{x} \\
&= \mathcal{N}(\mathbf{z}; \mathbf{x}_{AIS}, \Sigma_R + \Sigma_{AIS})\ \mathcal{N}(\mathbf{x}_0; \mu_1(\mathbf{z}, \mathbf{x}_{AIS}), \Sigma_1 + \Sigma_{\mathbf{x}}) \\
&\quad \times \int_{\mathbb{R}^2} \mathcal{N}(\mathbf{x}; \mu^*, \Sigma^*)\ d\mathbf{x} \\
&= \mathcal{N}(\mathbf{z}; \mathbf{x}_{AIS}, \Sigma_R + \Sigma_{AIS})\ \mathcal{N}(\mathbf{x}_0; \mu_1(\mathbf{z}, \mathbf{x}_{AIS}), \Sigma_1 + \Sigma_{\mathbf{x}})
\end{aligned} \tag{17}$$

where $\mathcal{N}(\mathbf{z}; \mathbf{x}_{AIS}, \Sigma_R + \Sigma_{AIS})$ comes from:

$$\begin{aligned}
&\mathcal{N}(\mathbf{x}; \mathbf{z}, \Sigma_R)\ \mathcal{N}(\mathbf{x}; \mathbf{x}_{AIS}, \Sigma_{AIS}) = \\
&\qquad \mathcal{N}(\mathbf{z}; \mathbf{x}_{AIS}, \Sigma_R + \Sigma_{AIS})\ \mathcal{N}(\mathbf{x}; \mu_1(\mathbf{z}, \mathbf{x}_{AIS}), \Sigma_1)
\end{aligned} \tag{18}$$

with

$$\Sigma_1 = (\Sigma_R^{-1} + \Sigma_{AIS}^{-1})^{-1} \tag{19}$$

$$\mu_1(\mathbf{z}, \mathbf{x}_{AIS}) = \Sigma_1(\Sigma_R^{-1}\mathbf{z} + \Sigma_{AIS}^{-1}\mathbf{x}_{AIS}) \tag{20}$$

and $\mathcal{N}(\mathbf{x}_0; \mu_1(\mathbf{z}, \mathbf{x}_{AIS}), \Sigma_1 + \Sigma_{\mathbf{x}})$ comes from the multiplication of the resulting Gaussian $\mathcal{N}(\mathbf{x}; \mu_1(\mathbf{z}, \mathbf{x}_{AIS}), \Sigma_1)$ by $\mathcal{N}(\mathbf{x}; \mathbf{x}_0, \Sigma_{\mathbf{x}})$:

$$\begin{aligned}
&\mathcal{N}(\mathbf{x}; \mu_1(\mathbf{z}, \mathbf{x}_{AIS}), \Sigma_1)\ \mathcal{N}(\mathbf{x}; \mathbf{x}_0, \Sigma_{\mathbf{x}}) \\
&\qquad = \mathcal{N}(\mathbf{x}_0; \mu_1(\mathbf{z}, \mathbf{x}_{AIS}), \Sigma_1 + \Sigma_{\mathbf{x}})\ \mathcal{N}(\mathbf{x}; \mu^*, \Sigma^*)
\end{aligned} \tag{21}$$

and μ^*, Σ^* do not need to be calculated.

Setting $\mathbf{x}_{AIS} = \mathbf{x}_{AIS} - \mathbf{d}$, it holds that under H_1:

$$\begin{aligned}
p(\mathbf{z}, \mathbf{x}_{AIS}|H_1) &= \mathcal{N}(\mathbf{z}; \mathbf{x}_{AIS} - \mathbf{d}, \Sigma_R + \Sigma_{AIS}) \\
&\quad \times \mathcal{N}(\mathbf{x}_0; \mu_1(\mathbf{z}, \mathbf{x}_{AIS}) - \Delta\mathbf{d}, \Sigma_1 + \Sigma_{\mathbf{x}})
\end{aligned} \tag{22}$$

where

$$\Delta = (\Sigma_{AIS}\Sigma_R^{-1} + I_{2\times 2})^{-1} \tag{23}$$

Then, the log-LRT will be

$$\begin{aligned}
&\log\Lambda(\mathbf{z}, \mathbf{x}_{AIS}, \mathbf{d}) \\
&= \log\left(\frac{p(\mathbf{z}, \mathbf{x}_{AIS}|H_1)}{p(\mathbf{z}, \mathbf{x}_{AIS}|H_0)}\right) \\
&= -(\mathbf{x}_{AIS} - \mathbf{z} - \mathbf{d})^T(\Sigma_R + \Sigma_{AIS})^{-1}(\mathbf{x}_{AIS} - \mathbf{z} - \mathbf{d}) \\
&\quad + (\mathbf{x}_{AIS} - \mathbf{z})^T(\Sigma_R + \Sigma_{AIS})^{-1}(\mathbf{x}_{AIS} - \mathbf{z}) \\
&\quad - (\mathbf{x}_0 - \mu_1(\mathbf{z}, \mathbf{x}_{AIS}) + \Delta\mathbf{d})^T(\Sigma_1 + \Sigma_{\mathbf{x}})^{-1} \\
&\qquad \cdot (\mathbf{x}_0 - \mu_1(\mathbf{z}, \mathbf{x}_{AIS}) + \Delta\mathbf{d}) \\
&\quad + (\mathbf{x}_0 - \mu_1(\mathbf{z}, \mathbf{x}_{AIS}))^T(\Sigma_1 + \Sigma_{\mathbf{x}})^{-1} \\
&\qquad \cdot (\mathbf{x}_0 - \mu_1(\mathbf{z}, \mathbf{x}_{AIS}))
\end{aligned} \tag{24}$$

Setting

$$A = (\Sigma_R + \Sigma_{AIS})^{-1}\ ,\qquad B = (\Sigma_1 + \Sigma_{\mathbf{x}})^{-1} \tag{25}$$

$$\mathbf{v} = \mathbf{x}_{AIS} - \mathbf{z}\ ,\qquad \mathbf{w} = \mathbf{x}_0 - \mu_1(\mathbf{z}, \mathbf{x}_{AIS}) \tag{26}$$

the log-likelihood ratio test will have the form

$$\begin{aligned}
&\log\Lambda(\mathbf{z}, \mathbf{x}_{AIS}, \mathbf{d}) = \\
&\quad [a_{xx}d_x(2v_x - d_x) - b_{xx}(\Delta\mathbf{d})_x(2w_x + (\Delta\mathbf{d})_x)] \\
&\quad + [a_{yy}d_y(2v_y - d_y) - b_{yy}(\Delta\mathbf{d})_y(2w_y + (\Delta\mathbf{d})_y)] \\
&\quad + (a_{xy} + a_{yx})\,[v_y d_x + v_x d_y - d_x d_y] \\
&\quad - (b_{xy} + b_{yx})\,[w_y(\Delta\mathbf{d})_x + w_x(\Delta\mathbf{d})_y + (\Delta\mathbf{d})_x(\Delta\mathbf{d})_y]
\end{aligned} \tag{27}$$

where: a_{ij} are the elements of matrix A, b_{ij} the elements of matrix B, v_i the elements of vector $\mathbf{v}$, w_i the elements of vector $\mathbf{w}$ and $(\Delta\mathbf{d})_i$ are the elements of vector $\Delta\mathbf{d}$.

Given the fact that

$$E[\mathbf{v}|H_0] = [0, 0]\ \ ,\ \ E[\mathbf{v}|H_1] = \mathbf{d} \tag{28}$$

$$E[\mathbf{w}|H_0] = [0, 0]\ \ ,\ \ E[\mathbf{w}|H_1] = \Delta\mathbf{d} \tag{29}$$

the expected value of the log-LRT under the two hypotheses can be evaluated:

$$\begin{aligned}
&E[\log\Lambda(\mathbf{z}, \mathbf{x}_{AIS})|H_0] = \\
&\quad - \left[a_{xx}(\hat{d}_x)^2 + b_{xx}(\Delta\hat{\mathbf{d}})_x^2\right] - \left[a_{yy}(\hat{d}_y)^2 + b_{yy}(\Delta\hat{\mathbf{d}})_y^2\right] \\
&\quad - (a_{xy} + a_{yx})\hat{d}_x\hat{d}_y - (b_{xy} + b_{yx})(\Delta\hat{\mathbf{d}})_x(\Delta\hat{\mathbf{d}})_y
\end{aligned} \tag{30}$$

$$\begin{aligned}
&E[\log\Lambda(\mathbf{z}, \mathbf{x}_{AIS}, \mathbf{d})|H_1] = \\
&\quad \left[a_{xx}\hat{d}_x(2d_x - \hat{d}_x) - b_{xx}(\Delta\hat{\mathbf{d}})_x(2(\Delta\mathbf{d})_x + (\Delta\hat{\mathbf{d}})_x)\right] \\
&\quad + \left[a_{yy}\hat{d}_y(2d_y - \hat{d}_y) - b_{yy}(\Delta\hat{\mathbf{d}})_y(2(\Delta\mathbf{d})_y + (\Delta\hat{\mathbf{d}})_y)\right] \\
&\quad + (a_{xy} + a_{yx})\left[d_y\hat{d}_x + d_x\hat{d}_y - \hat{d}_x\hat{d}_y\right] \\
&\quad - (b_{xy} + b_{yx})\left[(\Delta\mathbf{d})_y(\Delta\hat{\mathbf{d}})_x + (\Delta\mathbf{d})_x(\Delta\hat{\mathbf{d}})_y\right. \\
&\quad \left. + (\Delta\hat{\mathbf{d}})_x(\Delta\hat{\mathbf{d}})_y\right]
\end{aligned} \tag{31}$$

where d_i are the elements of the true spoofing distance vector, $\hat{d}_i$ are the elements of the spoofing distance vector used in the log-LRT and $(\Delta\hat{\mathbf{d}})_i$ are the elements of vector $\Delta\hat{\mathbf{d}}$.

APPENDIX B
MULTI-RADAR LIKELIHOOD

When measurements from K radars are available, the multi-radar likelihood will be

$$p([\mathbf{z}_1,\ldots,\mathbf{z}_K],\mathbf{x}_{AIS}|H_i) = \int \prod_{k=1}^{K}\{p(\mathbf{z}_k|\mathbf{x})\}\ p(\mathbf{x}_{AIS}|H_i,\mathbf{x})\ p(\mathbf{x})\ d\mathbf{x} \quad (32)$$

where $i = 0, 1$.

If the measurements of each radar follow a Gaussian distribution and all have the same covariance matrix then using Eq. (3-5) it holds that:

$$\prod_{k=1}^{K}\{p(\mathbf{z}_k|\mathbf{x})\} = \prod_{k=1}^{K}\{\mathcal{N}(\mathbf{z}_k;\mathbf{x},\Sigma_R)\} = \epsilon\ \mathcal{N}(\mathbf{z}'_k;\mathbf{x},\Sigma'_R) \quad (33)$$

with

$$\mathbf{z}'_k = \frac{1}{K}\sum_{k=1}^{K}(\mathbf{z}_k) \quad , \quad \Sigma'_R = \frac{1}{K}\Sigma_R \quad (34)$$

$$\epsilon = \prod_{k=2}^{K}\left[\mathcal{N}\left(\mathbf{z}_k;\mathbf{z}'_{k-1},\left(\frac{k}{k-1}\right)\Sigma_R\right)\right] \quad (35)$$

Therefore, the LRT will have the same form as in the case of one sensor (see Eq. 27, where ϵ in the nominator and denominator have been canceled out) with the differences that *a*) the arithmetic mean of the measurements of all radars is used as a single measurement; and *b*) the common measurement covariance matrix divided by the number of radars K is used as new measurement covariance matrix.

ACKNOWLEDGMENT

The research leading to these results has received funding from the EUs Seventh Framework Programme under grant agreement n° 238710. The research has been carried out in the MC IMPULSE project: https://mcimpulse.isy.liu.se.

REFERENCES

[1] B. Ristic, B. La Scala, M. Morelande, and N. Gordon, "Statistical analysis of motion patterns in AIS data: Anomaly detection and motion prediction," in *Information Fusion, 2008 11th International Conference on*, 30 2008-july 3 2008, pp. 1 –7.

[2] R. Lane, D. Nevell, S. Hayward, and T. Beaney, "Maritime anomaly detection and threat assessment," in *13th International Conference on Information Fusion*, july 2010, pp. 1 –8.

[3] M. Guerriero, S. Coraluppi, C. Carthel, and P. Willett, "Analysis of AIS Intermittency and Vessel Characterization using a Hidden Markov Model," in *Informatik 2010: Service Science - Neue Perspektiven für die Informatik, Beiträge der 40. Jahrestagung der Gesellschaft für Informatik e.V. (GI), Band 2. 5th German Workshop SDF 2010 - Sensor Data Fusion: Trends, Solutions, Applications*, 2010.

[4] K. Kowalska and L. Peel, "Maritime anomaly detection using Gaussian process active learning," in *15th International Conference on Information Fusion*, july 2012, pp. 1164 –1171.

[5] M. Vespe, I. Visentini, K. Bryan, and P. Braca, "Unsupervised learning of maritime traffic patterns for anomaly detection," in *9th IET Data Fusion Target Tracking Conference (DF TT 2012): Algorithms Applications*, may 2012, pp. 1 –5.

[6] Teleplan Globe AS, "MARIA Warship AIS Module," dec 2012. [Online]. Available: http://www.teleplanglobe.com/index.php/products/maria

[7] CNS Systems, "SENTINEL Surveillance and Secure Information System," dec 2012. [Online]. Available: http://www.cns.se/virtupload/content/5/CNSS_11_1667_E_Sentinel.pdf

[8] Y. Bar-Shalom, P. Willet, and X. Tian, *Tracking and Data Fusion: A Handbook of Algorithms*. YBS publishing, 2011.

[9] H. V. Poor, *An Introduction to Signal Detection and Estimation; 2nd edition*. Springer, 1994.

[10] S. Kay, *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory*. Prentice Hall, 1998.

[11] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," nov 2012, version 20121115. [Online]. Available: http://www2.imm.dtu.dk/pubdb/p.php?3274

[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, Inc., 2006.

[13] A. Wald, "Sequential tests of statistical hypotheses," *The Annals of Mathematical Statistics*, vol. 16, no. 2, pp. 117–186, June 1945. [Online]. Available: http://www.jstor.org/stable/2235829

[14] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*. Prentice-Hall, Inc., 1993. [Online]. Available: http://people.irisa.fr/Michele.Basseville/kniga/kniga.pdf

[15] X. R. Li and V. P. Jilkov, "Survey of maneuvering target tracking," *IEEE Transactions on Aerospace and Electronics Systems*, vol. 39, pp. 1333–1400, 2003.