

A complex event processing approach to detect abnormal behaviours in the marine environment

Fernando Terroso-Saenz¹ · Mercedes Valdes-Vela¹ · Antonio F. Skarmeta-Gomez¹

Published online: 9 May 2015
© Springer Science+Business Media New York 2015

Abstract Over the last years, many data-sources have become available to monitor the marine traffic. This has motivated the development of support systems to automatically detect vessels' behaviours of interest. The present work states a novel approach in this domain following the Complex Event Processing (CEP) paradigm. As a proof of concept, a CEP-based system has been developed to timely detect a set of vessel's abnormal behaviours by performing an event-based processing of Automatic Identification System data. Experiments based on real-world and synthetic data proved the suitability and feasibility of the proposal.

Keywords Complex event processing · Abnormal behaviour · Automated identification system · Vessel traffic service

1 Introduction

In this day and age, a huge amount of vessel data is available due to a varied range of existing sources in the marine domain such as the *Automated Identification System* (AIS), radars, satellites or video-cameras. As a result, *Vessel Traffic*

Service (VTS) entities in charge of high-traffic areas have to deal with such incoming information from all these sources. This has led to an increasing necessity to develop detection systems which focus on perceiving and informing about vessels' abnormal behaviours in order to support a VTS staff (Nuutinen et al. 2007).

These behaviours may give insight into illegal and/or dangerous activities (i.e. collisions, smuggling or human trafficking). Therefore, one important requirement for these systems is the timely detection of their target behaviours. However, most of the current anomaly-detection system in the marine environment hardly address time constraints to detect their target abnormal behaviours as they usually follow either a data-based or a rule-based approach.

Regarding time-constrained scenarios, the *Complex Event Processing* (CEP) paradigm has arisen as a well-known solution to develop event-based systems in this type of environments for the last years (Etzion and Niblett 2010). A CEP system mainly performs an online processing of primary events from different and distributed data sources and, on the basis of those events, makes up derived events representing pre-defined situations of interest. For that purpose, a CEP system usually relies on different filtering, correlation and pattern-based operators so as to perform its processing.

In this frame, most of the data received by a VTS, like the AIS data, is naturally event-based as each AIS location-message from a vessel can be seen as an event informing about a velocity and/or location change of such vessel. Moreover, most of the vessels' behaviours and activities that are interesting for a VTS can be extracted from its received data streams. Consequently, the present work states that CEP is a suitable approach to develop anomaly-detection systems in the marine environment.

As a proof of concept, the present work puts forward a novel CEP-based system able to timely perceive different

✉ Fernando Terroso-Saenz
fterroso@um.es

Mercedes Valdes-Vela
mdvaldes@um.es

Antonio F. Skarmeta-Gomez
skarmeta@um.es

¹ Department of Information and Communication Engineering, Computer Science Faculty, University of Murcia, 30100 Murcia, Spain

abnormal behaviours related to the vessels moving in a VTS's area of interest. In order to do that, it performs an event-based processing to the information periodically reported by the AISs installed in the vessels. Furthermore, the system also takes into account other *contextual* events related to the VTS's area of interest itself, like the current weather conditions, in order to improve the detection accuracy. As a result, the system delivers a set of *early alerts* each time it detects a target behaviour with the aim of informing the VTS staff by means of some type of dashboard. In that sense, such dashboard is out of the scope of the present paper.

Concerning the target behaviours, they have been chosen on the basis of the conclusions of several workshops where the most interesting situations for the VTS staff were defined (Roy and Davenport 2009; van Laere and Nilsson 2009; Roy 2008). More specifically, the system centres on two types of abnormal behaviours,

- The first one comprises those behaviours in which only a single vessel is involved. In particular, the system focuses on detecting whether a boat is moving abnormally fast or slow. Perceiving these situations is fairly important because, under certain circumstances, a too high or slow vessel might be a sign that such vessel is carrying out some type of illegal or dangerous activity.
- The second type of behaviour involves more than one vessel. In more detail, the present system focuses on detecting situations where two different vessels might be about to collide with each other. Detecting this type of dangerous situations is paramount to improve the safety of the marine traffic.

The adopted CEP approach makes no longer necessary a previous training or data-gathering step as in former solutions which usually rely on some type of in-disk processing. On the contrary, the CEP paradigm focuses on processing the incoming data in an asynchronously and fast way so as to timely detect the target activities. This feature is specially important in the VTS domain where it is paramount to detect the behaviours of interest as soon as possible.

On the whole, the core contribution of this paper is the description of a novel event-based mechanism capable of detecting different behaviours of interest involving either one or more vessels. This has implied the definition of different event-based patterns.

The remainder of the paper is structured as follows, an overview of the state of the art of the anomaly detection systems in the marine environment and the CEP domain is put forward in Section 2. Next, a detailed explanation of the CEP anomaly-detection system is stated in Section 3. Then, Section 4 discusses the results of the different experiments to test the system. Finally, the main conclusions and the future work are summed up in Section 5.

2 Background

2.1 Maritime abnormal-behaviour detection

As far as anomaly-detection systems in the marine environment are concerned, it is possible to distinguish two broad trends.

On one hand, the large amount of available AIS data has motivated the development of data-driven (aka bottom-up) solutions. This type of systems focuses on learning *normal* behaviours from historical AIS data. Thus, it is possible to detect situations which are *abnormal* with respect to the training data set where expert knowledge is not necessary. In this frame, several methods have been proposed in order to generate a *model of normality* like gaussian mixture models (Kowalska and Peel 2012; Will et al. 2011; Garagic et al. 2009), kernel density estimators (Ristic et al. 2008) or bayesian networks (Kruger et al. 2012; Lane et al. 2010; Mascaro et al. 2010). Nevertheless, a common issue of those solutions is that they rely on a previous training step, so the accuracy and reliability of the model depends on the *representativeness* of the used dataset. Hence, during the last years various *on-line* solutions have arisen to learn normality models *on the fly* to avoid the aforementioned training step (Vespe et al. 2012; Bomberger et al. 2006). Since the event-based patterns of the CEP approach introduced in the present paper have been defined by means of expert knowledge, the aforementioned training techniques have not been used. However, these data-driven solutions can be regarded as *complementary* to the CEP paradigm because they could be used to define event-based patterns. Moreover, the proposed system is capable of detecting a more varied range of behaviours than the aforementioned on-line solutions as they usually focus on abnormal velocity variations of a boat. On the contrary, the introduced CEP-system is capable of detecting activities of interest not only related to one single vessel but also those involving more than one boat.

On the other hand, the second line of work follows a rule-based (aka model-driven) approach. In this case, the different solutions comprise a set of rules or patterns to detect a pre-defined group of anomalies on the basis of the received AIS data (Idiri and Napoli 2012; Roy 2010). Our proposal could also be viewed as a kind of rule-based system as it comprises a set of pre-defined patterns whose goal is to detect various target behaviours. Nonetheless, most of the previous methods do not perform an online processing of the incoming AIS data. Still, this data is previously stored and indexed in a database before being processed by the system afterwards. In that sense, the proposed CEP approach allows to timely processing the incoming AIS data without a previous storage stage and, as a result, it generates alerts informing about abnormal behaviours with short delay.

2.2 Complex event processing

In this day and age, the *information flow processing models* have become an important approach so as to cope with time constraints in a wide range of environments (Cugola and Margara 2012). In this frame, CEP has played a key rol.

An overriding line of work in the CEP domain has been the deployment of event-based systems in the business field (Luckham 2011). Nevertheless, several CEP-based proposals have gone beyond that field and have widened the CEP's usage range in several scopes such as advertisement management (Evensen and Meling 2012), road-traffic monitoring (Terroso-Saenz et al. 2012), context-aware services (Terroso-Saenz et al. 2012) or telemedical systems (Meister 2012). Regarding the marine domain, little event-based efforts have been undertaken so far. In that sense, some papers have put forward, as illustrative examples of their core contributions, event-based solutions to detect certain situations involving one or more boats like unusual low speed (Verginadis et al. 2012; Patiniotakis et al. 2013). Unlike these proposals, the present work introduces a more detailed solution to perceive marine situations of interest comprising a more varied range of patterns.

3 CEP anomaly detection system

As Fig. 2 depicts, the proposed system intends to connect a set of event producers, which in the present domain are the vessels in a VTS's area of interest, and the event consumer, which in this case is some type of back-end service used to inform the VTS staff of the detected behaviours. The following subsections explain in detail the different elements of the system.

3.1 Target behaviours

During the last years, various workshops have been held to define the most interesting activities, situations and/or behaviours that should be detected so as to improve the marine traffic's safety (Roy and Davenport 2009; van Laere and Nilsson 2009; Roy 2008). A recurrent point in those meetings was to detect abnormal vessels' speeds. Another point of interest was the early detection of possible collisions between different vessels. Keeping in mind these results, the developed system is able to detect three different vessels' activities,

- The first one occurs when a vessel is moving at abnormally high speed during a certain period of time. Perceiving this situation is fairly important because a too high value of a vessel's speed may be a sign that such

vessel is carrying out some type of illegal or dangerous activity which can put in danger the safety of the surrounding boats specially in areas with high density of traffic.

- The second behaviour arises when a single boat is doing abnormally low speed. This situation may be also a sign of certain suspicious activities. For example, a vessel loitering during a long period of time might indicate an illegal-fishing situation.
- The third type of target behaviour involves more than one vessel. Specifically, the present system focuses on detecting situations where two different vessels might be about to collide with each other. Detecting this type of dangerous situation is paramount in order to improve the safety of the marine traffic.

Each time the system detects any of these behaviours, it delivers an alert to the event consumer.

3.2 Event model

One of the most important duties when it comes to develop a CEP system is to properly define the different events that the system is intended to process. In CEP, an event can be defined as "*an occurrence within a particular system or domain; it is something that has happened, or is contemplated as having happened in that domain*" (Etzion and Niblett 2010). Figure 1 shows the hierarchy of events reporting the *occurrences* that are interesting in the system's domain.

As we can see, all the system's event types inherit from *root event* that contains the attributes that are common to the rest of events. Those event types can be classified in two different groups, namely the ones related to the target vessels (*vessel event*) and the ones related to the area of interest where these vessels move (*context event*).

Regarding the *vessel event* type, the leftmost sub-group comprises the types *location event* and *filtered location event*. The former unifies the information from the vessels's AISs. For that purpose, an adaptation process is performed to the incoming AIS data as it is explained in Section 3.3.1. In addition to that, the *filtered location event* represents those location events that have been filtered-in by a filter mechanism to *smooth* the stream of location events as it is explained in Section 3.3.2.

The second sub-group of event types is compound of the *velocity event* and its two subtypes, *current velocity* and *average velocity event*. While the current velocity event indicates a boat's velocity during a recent period of time, the average velocity event reports a vessel's velocity during a longer one. Hence, this group of events reports information about the movement of the vessels as it is described in Section 3.3.3

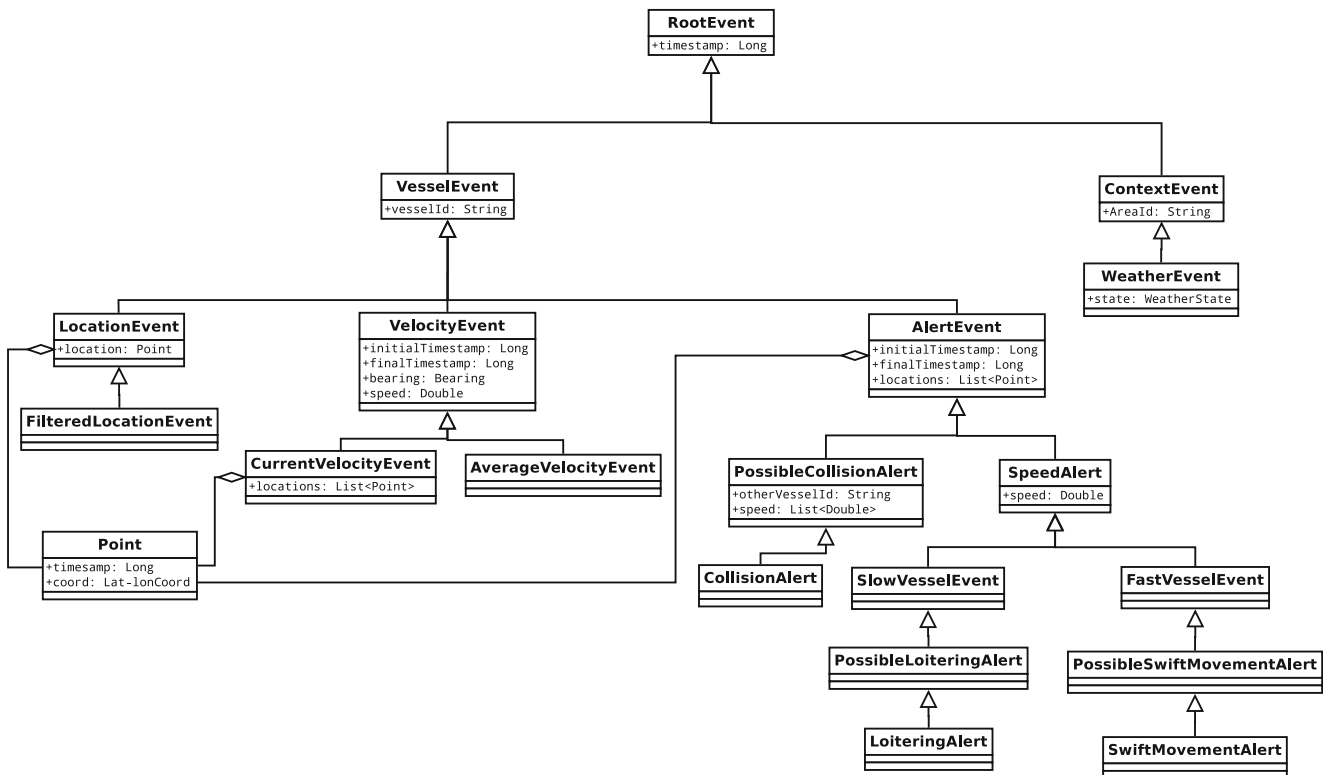


Fig. 1 Event model for the proposed system

Next, the third *vessel event* sub-group comprises the alerts for the three target behaviours. Thus, the *collision alert* represents a collision that is about to happen between two different vessels. The *swift movement alert* indicates a situation where a vessel is doing too high speed according to certain parameters. Lastly, the *loitering alert* reports a situation in which a boat is moving at a too low speed with respect to certain criteria. These three event types have their own super types, *possible collision alert*, *possible loitering alert* and *possible swift movement alert*. These super types represent situations that might develop into the target abnormal behaviours in the future as it is put forward in Section 3.3.5.

Finally, the *context event* group represents meaningful events about the area of interest which may affect the vessels' behaviour. In this first version, this group is basically compound of the *weather event*. This event type reports the current weather conditions of the VTS's area of interest.

3.3 System architecture

As Fig. 2 shows, the introduced CEP system takes as incoming *raw* events the data reported by the AIS in each target boat and it performs an event-based processing of them afterwards by means of its Event Processing Agents (EPAs).

An EPA can be defined as a CEP component in charge of processing events at a certain abstraction level. As a result, it generates different *derived* events (or alerts) reporting the abnormal behaviours listed in Section 3.1 that are eventually sent to a VTS dashboard.

For the sake of clarity, Table 1 lists notations used throughout the paper.

3.3.1 Adaptor EPA

As Fig. 2 shows, this agent is responsible for processing the vessel's AIS messages received by the system and unify them in a unique representation. In particular, the agent only processes the AIS location reports in spite of the fact that the AIS description comprises other types of messages (Navigation Center-United States Coast Guard 2013). Those location messages contains, among other data, the unique identifier of the sender vessel, the location's coordinates, the current speed of movement and the timestamp at which the message was generated (Navigation Center-United States Coast Guard 2013). Therefore, the adaptor agent maps each location report to a new *location event* comprising the aforementioned fields and it discards the other types of messages. Next, the generated events are processed by the upstream agents.

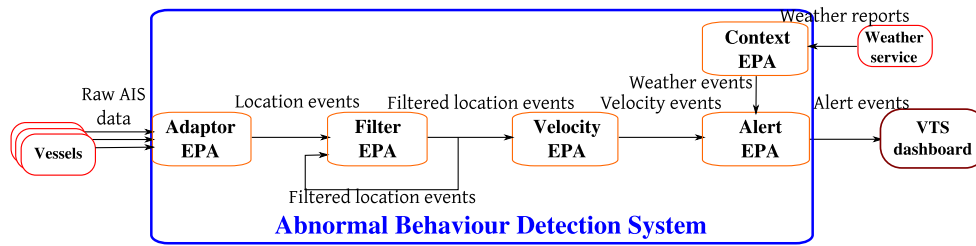


Fig. 2 Abnormal detection system schema. The event channels have been labelled with the stream of events that flows through it

3.3.2 Filter EPA

The stream of location events generated by the Adaptor EPA might comprise a huge amount of events. Therefore, it is necessary to clean up this stream by discarding the irrelevant events.

As a result this EPA discards those location events that do not imply a meaningful movement, otherwise a *filtered location event* is created. Furthermore, a time-based watchdog (t_{max}) is also included to avoid event starvation in the upstream EPAs. All in all, the filtering mechanism can be described as follows,

$$\begin{aligned}
 &le_i^j \rightarrow fle_k^j(\\
 &fle_k^j.timestamp = le_i^j.timestamp, \\
 &fle_k^j.location = le_i^j.location \\
 &fle_k^j.speed = le_i^j.speed \\
 &) \text{ iff}
 \end{aligned}$$

$$\begin{aligned}
 dist \left(le_i^j, fle_{k-1}^j \right) &\geq d_{min} \vee (le_i^j.timestamp \\
 &- fle_{k-1}^j.timestamp) \geq t_{max}
 \end{aligned}$$

Lastly, Fig. 3a depicts an example of the aforementioned filtering process.

3.3.3 Velocity EPA

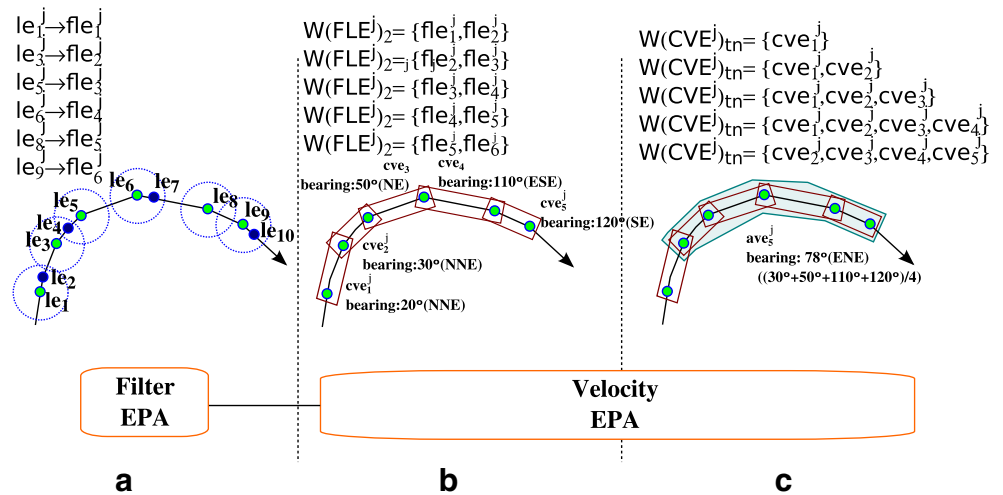
The main goal of this agent is to calculate certain details of the movement of each vessel by aggregating the filtered location events coming from the filter EPA.

First of all, this agent calculates the current velocity (bearing and speed) of each vessel by means of a length-based sliding window that stores the last two filtered locations events of each vessel. On the basis of these two events, the agent makes up a *current velocity event* instance which comprises different details of the current movement of its

Table 1 Table of notations

Symbol	Meaning
le_i^j	The i-th <i>location event</i> of the vessel j.
fle_k^j	The k-th <i>filtered location event</i> of the vessel j.
cve_k^j, ave_k^j	The k-th <i>current velocity</i> and <i>average velocity event</i> of the vessel j.
sve_k^j	The k-th <i>slow vessel event</i> of the vessel j.
pla_k^j, la_k^j	The j-th <i>possible loitering</i> and <i>loitering alert</i> of the vessel j.
pca_k^{ij}, ca_k^{ij}	The k-th <i>possible collision</i> and <i>collision alert</i> between the vessels i and j.
FLE^j	Stream of <i>filtered location events</i> of the vessel j.
CVE^j	Stream of <i>current velocity events</i> of the vessel j.
$W(X)_n$	Length-based sliding window retaining the last n events of the stream X .
$W(X)_n^i$	The i-th event instance contained in the sliding window $W(X)_n$.
$W(X)_{tn}$	Time-based sliding window retaining the events of the stream X generated during the last n time units.
$W(X)_{tn}^i$	The i-th event instance contained in the sliding window $W(X)_{tn}$.
x_{last}	The most recent event of type x made up by the system.
$x \rightarrow y$	The event instance x gives raise to a new event instance y .

Fig. 3 Example of the event processing made by the system for a particular vessel j . **a** Location event filtering. **b** Current velocity generation. **c** Average velocity event generation



sender vessel. The way these attributes are calculated is specified next,

$$\begin{aligned}
 &W(FLE^j)_2 \rightarrow cve_k^j(\\
 &\quad cve_k.timestamp = current_timestamp, \\
 &\quad cve_k.initial_timestamp = W(FLE^j)_2^1.timestamp, \\
 &\quad cve_k.final_timestamp = W(FLE^j)_2^2.timestamp, \\
 &\quad cve_k.speed = W(FLE^j)_2^2.speed, \\
 &\quad cve_k.bearing = bearing(W(FLE^j)_2^1.location, \\
 &\quad\quad\quad W(FLE^j)_2^2.location), \\
 &\quad cve_k.cell = cell(W(FLE^j)_2^2.location) \\
 &\quad cve_k.locations = \{W(FLE^j)_2^1.location, \\
 &\quad\quad\quad W(FLE^j)_2^2.location\} \\
 &\quad)
 \end{aligned}$$

Regarding the new event’s content, the *bearing* attribute indicates the bearing in radians of going from the point $W(FLE^j)_2^1.location$ to $W(FLE^j)_2^2.location$ following a straight line. This can be calculated by basic mathematics.

Moreover, the *cell* attribute is used to position the vessel in a particular square cell of a *location grid*. This grid is compound of different square cells of the same size. Each cell is labelled with a particular coordinate with respect to a reference point. This reference point is the same for all the cells and could be the location of the VTS where the system is intended to run.

Next, the stream of current velocity events is used by the velocity EPA to generate the *average velocity events*. Unlike the current velocity event type, this type of event intends to report a more general view of a vessel’s movement. In order to do that, a time-based sliding window stores the current velocity events of each target vessel generated during

the last m time units. Hence, each time a new current velocity event is generated, a new average velocity event ave_k^j is composed as follows,

$$\begin{aligned}
 &W(CVE^j)_{tn} \rightarrow ave_k^j(\\
 &\quad ave_k.timestamp = current_timestamp, \\
 &\quad ave_k.initial_timestamp = W(CVE^j)_{tn}^1.initial_timestamp, \\
 &\quad ave_k.final_timestamp = W(CVE^j)_{tn}^{last}.final_timestamp, \\
 &\quad ave_k.speed = avg_speed(W(CVE^j)_{tn}) \\
 &\quad ave_k.bearing = avg_bearing(W(CVE^j)_{tn}) \\
 &\quad)
 \end{aligned}$$

In this definition, *avg_speed* and *avg_bearing* stands for ad-hoc aggregation methods in charge of calculating the average speed and bearing of the events contained in the aforementioned sliding window.

For the sake of clarity, Fig. 3b-c depicts an example of how the current and average velocity events are created.

To sum up, the velocity EPA emits two different event streams each of which informs of the movement of the vessels with different *granularities*. Whilst the current velocity events report details about the recent movement of a vessel, the average velocity events inform about the vessels’ movement in a coarser grain. This two levels of information allow the system to have a comprehensive view of vessel traffic at each moment.

3.3.4 Context EPA

In order to improve the system’s accuracy, certain *contextual* information related to the area of interest has been taken into account. This information has to do with particular conditions or circumstances that may affect the vessels’

behaviour and, thus, should be considered so as to infer whether a vessel is behaving abnormally or not. Such information may be reported by a varied range of data sources, like sensor networks or web/cloud services.

Consequently, the Context EPA is mainly in charge of injecting all the information coming from non-AIS sources to the rest of agents in the form of events. In the present version, this EPA focuses on the current weather conditions of the VTS’s area of interest. This information is a clear example of contextual information that might have an impact on a vessel’s behaviour.

The logic design of this agent is based on the sensor management module presented in Valdés-Vela and Gomez-Skarmeta (2010) which was already adapted in Terroso-Saenz et al. (2012) to generate *weather events* from an external weather-information provider. In brief, such a module endlessly requests new weather information to the provider. On the basis of that information, it periodically emits new *weather events* whose *state* attribute indicates the label-based classification of current weather conditions in the VTS’s area of interest.

3.3.5 Alert EPA

This EPA generates the alert events informing about the three target behaviours and delivers them to the dashboard acting as event consumer.

Abnormal low speed This behaviour is detected by an incremental event processing comprising two different steps. The first step centres on detecting vessels moving slowly by only taking into account its present speed. In this frame, a vessel’s current speed can be regarded as low from two different criteria. The former is related to the average speed of the vessel during a long period of time. In that

sense, if the current speed is meaningfully lower than the average of the previous measurements then this situation indicates that the vessel has sharply decelerate. The second criterion has to do with the vessels’ average speed in the area where the vessel is currently moving. In that sense, a vessel moving much more slowly than its surrounding vessels is also a potentially-dangerous behaviour.

If the two aforementioned criteria are accomplished, then a *slow vessel event* is generated representing the situation that a vessel is moving particularly slow.

$$\begin{aligned}
 & cve_i^j \rightarrow sve_k^j(\\
 & \quad sve_k.timestamp = current_timestamp, \\
 & \quad sve_k.initial_timestamp = cve_i^j.initial_timestamp, \\
 & \quad sve_k.final_timestamp = cve_i^j.final_timestamp, \\
 & \quad sve_k.locations = cve_i^j.locations \\
 & \quad sve_k.speed = cve_i^j.speed \\
 & \quad) \text{ iff} \\
 & \quad cve_i^j.speed < \\
 & \quad min(\delta_{dec}(we_{last}) * ave_{last}^j.speed, \delta_{dec}(we_{last}) \\
 & \quad * avg_speed(cve_i^j.cell)) \\
 & \quad 0 < \delta_{dec}(we_{last}) < 1
 \end{aligned}$$

According to this definition, the current velocity event under consideration (cve_i^j) gives raise to a new slow vehicle event (sve_k) only if its speed is below the average speed of both its vessel ($ave_{last}^j.speed$) and its current cell ($avg_speed(cve_i^j.cell)$) giving the *decreasing* factor δ_{dec} .

The particular value of δ_{dec} is modified depending on the current weather conditions reported by the most recent *weather event* (we_{last}). This assignment logic can be defined as follows,

$$\delta_{dec}(we_{last}) = \begin{cases} dec_factor & we_{last}.state \subset \{sunny, cloudy\}, \\ dec_factor \times 0.2 & we_{last}.state \subset \{foggy, rainy, windy, snowy\} \end{cases}$$

As we can see, δ_{dec} is decreased in case of adverse weather conditions. This way, a certain vessel’s speed, which is inferred as low in case of sunny weather, might be not classified that way in case of rough weather conditions. By means of this approach, the system takes into account the fact that a vessel tends to reduce speed if it faces dangerous weather conditions and, thus, it should not be classified as an abnormal movement.

Next, the slow movements perceived in the form of slow vessel events should be further analysed to detect whether they last enough to be regarded as persistent situations. Con-

sequently, it is necessary to define a time threshold to filter out those low-speed situations that are irrelevant because of their short lifetime.

In terms of events, if a boat moves quite slowly during a meaningful period of time, it will cause the generation of many consecutive slow vessel events. Hence, the first task of this step is to detect long sequences of these events related to the same vessel and group them in a *possible loitering alert* instance. As it was already stated in Section 3.2, this event type is used to represent a situation that is suspicious of being an abnormal low-speed behaviour but

it is not confirmed yet. This gathering process can be described as follow,

$$\begin{aligned} & \{sve_{last}^j, pla_{i-1}^j\} \rightarrow pla_i^j(\\ & \quad pla_i^j.timestamp = current_timestamp, \\ & \quad pla_i^j.initial_timestamp = pla_{i-1}^j.initial_timestamp, \\ & \quad pla_i^j.final_timestamp = sve_{last}^j.final_timestamp, \\ & \quad pla_i^j.locations = \{pla_{i-1}^j.locations, sve_{last}^j.locations\}, \\ & \quad pla_i^j.speed = avg_speed(sve_{last}^j, pla_{i-1}^j) \\ & \quad) \text{ iff} \\ & \quad pla_{i-1}^j.final_timestamp \geq sve_{last}^j.initial_timestamp \wedge \\ & \quad pla_{i-1}^j.final_timestamp < sve_{last}^j.final_timestamp \end{aligned}$$

In short, each new slow vessel sve_{last}^j is fused with the last possible loitering alert (pla_{i-1}^j) of the same vessel to create a new one (pla_i^j) if both events' time intervals overlap. Thus, a possible loitering alert gathers the information of the slow vessel events that are very close in time and, as a result, refer to the same situation.

Lastly, in case the reported time interval of a possible loitering alert is long enough to be considered a meaningful behaviour, a new loitering alert is made up. This new event is emitted by the alert EPA to the event consumer afterwards. This definition is shown next,

$$\begin{aligned} & pla_i^j \rightarrow la_k^j(\\ & \quad la_k.timestamp = pla_i^j.timestamp, \\ & \quad la_k.initial_timestamp = pla_i^j.initial_timestamp, \\ & \quad la_k.final_timestamp = pla_i^j.final_timestamp, \\ & \quad la_k.locations = pla_i^j.locations, \\ & \quad la_k.speed = pla_i^j.speed \\ & \quad) \text{ iff} \\ & \quad pla_i^j.final_timestamp - pla_i^j.initial_timestamp > t_{alert}^{min} \end{aligned}$$

In this definition, t_{alert}^{min} indicates the minimum time interval of a possible loitering alert (pla_i^j) to give raise to a new loitering alert la_k .

In conclusion, this aggregation step intends to deliver only the alerts informing of behaviours that last a certain period of time so that they can not be considered as exceptions but persistent in time. This intends to reduce the cognitive overload of the VTS staff.

Abnormal high speed The approach to detect abnormal high speed values is similar than the one for the abnormal low speed described above.

Firstly, if the system detects that a vessel is moving faster than its own average speed and the one in its surrounding area then it makes up a *fast vessel event* to represent such situation. In this case, the *weather events* are not taken into account.

Secondly, the system aggregates the fast vessel events that overlap in time giving raise to a *possible swift movement alert*. Then, each possible swift movement alert is filtered depending on its reported time length. Provided that this time period is over a certain threshold, the system composes a new *swift movement alert* which is eventually emitted by the alert EPA as a system output.

Possible collision In this third case, the system intends to detect those situations in which two different vessels are about to collide to each other. Intuitively, this situation occurs when two different vessels follow such close and convergent trajectories that, as a result, they will collide to each other if they do not change their current velocities.

In order to formally describe this situation, we will use the term *closest point of approach* (CPA) previously coined in the Moving Object Databases (MODs) domain (Guting and Schneider 2005). Basically, a CPA is the location at which two moving objects have attained their closest possible distance based on their *known* positions. This term has been slightly modified to give raise to the *projected closest point of approach* (pCPA). A pCPA is the forecast location at which two moving objects will attain their closest possible distance if they keep moving with their current velocities. Unlike the former CPA, a pCPA does not only rely on the already-known locations of the objects but also on the future ones according to their current velocities.

As Fig. 4 depicts, the pCPA can be seen as the intersection point of the forecast trajectories of the two involved vessels (i and j). These two trajectories can be calculated by the last two known locations of each vessel ($\{a_i, b_i\}$ and $\{a_j, b_j\}$). Figure 4 also shows the distance between the current position of each vessel and their common pCPA (d_{pCPA}^i and d_{pCPA}^j). On the basis of these distances and the current speed of each vessel, it is possible to know the time at which each vessel will reach their pCPA (t_{pCPA}^i and t_{pCPA}^j) by using basic calculus.

Consequently, two different vessels i, j might collide to each other if the two following conditions occur,

1. A pCPA for the two vessels exists and
2. $|t_{pCPA}^i - t_{pCPA}^j| \leq t_{pCPA}^{max}$

The alert EPA intends to detect these two conditions by means of the movement information reported by the current velocity events.

In the first place, the agent correlates the current velocity events of different pairs of vessels so as to detect whether they are quite close to each other both in space and time. If

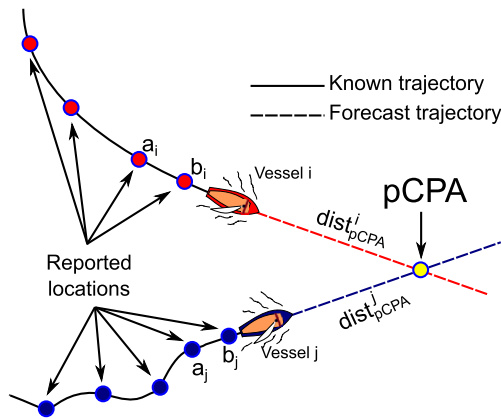


Fig. 4 Example of a pCPA of two different vessels given their real and predicted trajectories

both conditions are accomplished, the alert EPA makes up a new possible collision alert. This first step is defined next,

$$\{cve_{last}^i, cve_{last}^j\} \rightarrow pca_k^{ij} ($$

$$pca_k^{ij}.timestamp = current_timestamp,$$

$$pca_k^{ij}.locations = \{cve_{last}^j.locations, cve_{last}^i.locations\},$$

$$pca_k^{ij}.speeds = \{cve_{last}^j.speed, cve_{last}^i.speed\}$$

) iff

$$i \neq j \wedge$$

$$|cve_{last}^j.cell.x - cve_{last}^i.cell.x| \leq 1 \wedge$$

$$|cve_{last}^j.cell.y - cve_{last}^i.cell.y| \leq 1 \wedge$$

$$|cve_{last}^j.final_timestamp - cve_{last}^i.final_timestamp| \leq t_{pca}^{max}$$

As this definition shows, the alert EPA correlates the most recent current velocity events of each pair of vessels ($cve_{wlast}^j, cve_{wlast}^i$) to create a new possible collision alert pca_k . Two vessels are regarded as close to each other in space if their last current velocity events are in the same or adjacent cells of the location grid.

On the basis of the information of each pca_k^{ij} , the alert EPA calculates the pCPA of the two involved vessels along with their times to reach it. If the difference between those times is below a certain threshold t_{pCPA}^{max} then a new collision alert is made up. This procedure is described as follows,

$$pca_k^{ij} \rightarrow ca_l^{ij} ($$

$$ca_l^{ij}.timestamp = pca_k^{ij}.timestamp,$$

$$ca_l^{ij}.locations = pca_k^{ij}.locations$$

) iff

$$|t_{pCPA}(pca_k^{ij}.locations, pca_k^{ij}.speeds, i) -$$

$$t_{pCPA}(pca_k^{ij}.locations, pca_k^{ij}.speeds, j)| \leq t_{pCPA}^{max}$$

Finally, each ca_l^{ij} is delivered to the dashboard.

4 Experiment results

In order to test the present system, it was fully implemented by means of the CEP platform Esper (Espertech 2013). Esper is a well-established GNU open-source CEP tool that defines its own stream-oriented *Event Processing Language* (EPL). This EPL allows to specify the processing of each EPA by means of a varied collection of built-in or ad-hoc resources such as sliding windows, contexts or aggregation functions.

Next, we evaluated the system in a maritime scenario comprising real-world datasets. Besides, we compared our approach with one state-of-the-art anomaly detection system for the maritime domain. Finally, in order to study the potential feasibility of the proposal for other domains, we tested the system in a road-traffic scenario.

4.1 Maritime case study

4.1.1 Experiment setup

Datasets. Two different datasets were used for this evaluation and both of them were injected to the system at the same time. The former was generated by means of two *rigid-hulled inflatable boats* (RHIBs). Both RHIBs were ordered to perform various movements and follow different abnormal behaviours in order to cover a varied range of situations that may arise in the marine environment during a 7-hour trial departing from Cowes (England). The two RHIBs were equipped with data loggers that recorded the vessels' locations and speed of movement, among other parameters, at 10-second intervals. As a result, a location log comprising 2332 locations for the first boat and another containing 2327 locations for the second RHIB were created. All these locations fell into the square of latitude 50.76 to 50.80 and longitude 1.37 to 1.27 whose length was approximately 6000×5000 m.

The second dataset was the real AIS data collected from an AIS-monitoring web service.¹ Due to the slow refresh date of this type of web services, the data of this second dataset was gathered in 10-minute intervals, and it also covered a 7-hour period. This dataset comprised 200 vessels of different types and sizes and it fell into the square of latitude 50.73 to 50.81 and longitude -1.40 to -1.28.

Settings The system evaluation was conducted on a PC running a Ubuntu 12.04 operating system with 4GiB of memory, Intel(R) Core i5 at 3.10GHz and Java Runtime Environment 7.0 (JRE 7) with 2GiB of allocated memory.

Table 2 lists the default parameters used throughout the experiment set by expert knowledge. In this frame, the size

¹<http://vesseltracker.com>

Table 2 Values of the system's parameters for the experiments

Param.	Description	Holder agent	Value
d_{min}	Min. distance between two consecutive location events.	Filter EPA	100 m
t_{max}	Max. time between two consecutive location events.	Filter EPA	60 s
tn	Length of the time window to generate average velocity events.	Velocity EPA	300 s
dec_factor	Decreasing factor to detect abnormally-low speeds.	Alert EPA	0.3
δ_{inc}	Increasing factor to detect abnormally-high speeds.	Alert EPA	1.4
t_{alert}^{min}	Min. time length of an alert event.	Alert EPA	120 s
t_{pca}^{max}	Max. time gap between to current velocity events to give raise to a possible collision event.	Alert EPA	3 s
t_{pCPA}^{max}	Max. time for two vessels to reach their $pCPA$ to give raise to a collision event.	Alert EPA	10 s

of the square cells was 1000×1000 m length and the reference point to make up the location grid was the port from which the two RHIBs departed. This size allowed to enclose this port (where the vessels must move quite slow) in a unique square whereas other open-sea areas (where the boats are allowed to move more freely) were contained in other different squares. As a result, the location grid comprised 2234 cells, 31 of which were enclosed in the RHIBs' dataset area.

Reference framework (RF) Our CEP-based solution has been compared with the maritime anomaly detection system described in Kowalska and Peel (2012). In brief, this framework creates a model of normality from historical AIS data using Gaussian Processes which do not require expert knowledge. Then, on the basis of such a model, an anomaly score for each observation is obtained. For its evaluation, authors in Kowalska and Peel (2012) make use of the same

RHIB dataset than the present work, so the outcome of both solutions can be easily compared.

Methodology Since the RHIBs' dataset was labelled with the behaviour of each RHIB, it was used to compare the system's alerts with the vessels' actual behaviour so as to study their accuracy and reliability. Thus, the main goal of the second dataset was to inject real data to the system so that it was capable of calculating the cells' average speed of the location grid in a realistic way. However, the system's alerts related to the vessels of this dataset were not taken into account in this case study.

4.1.2 Speed-based alerts

Figure 5a shows the evolution of the real speed of the vessel RHIB 1 along with its average speed during the last tn seconds (see Table 2). These average values are the ones reported by the average velocity events. The figure also

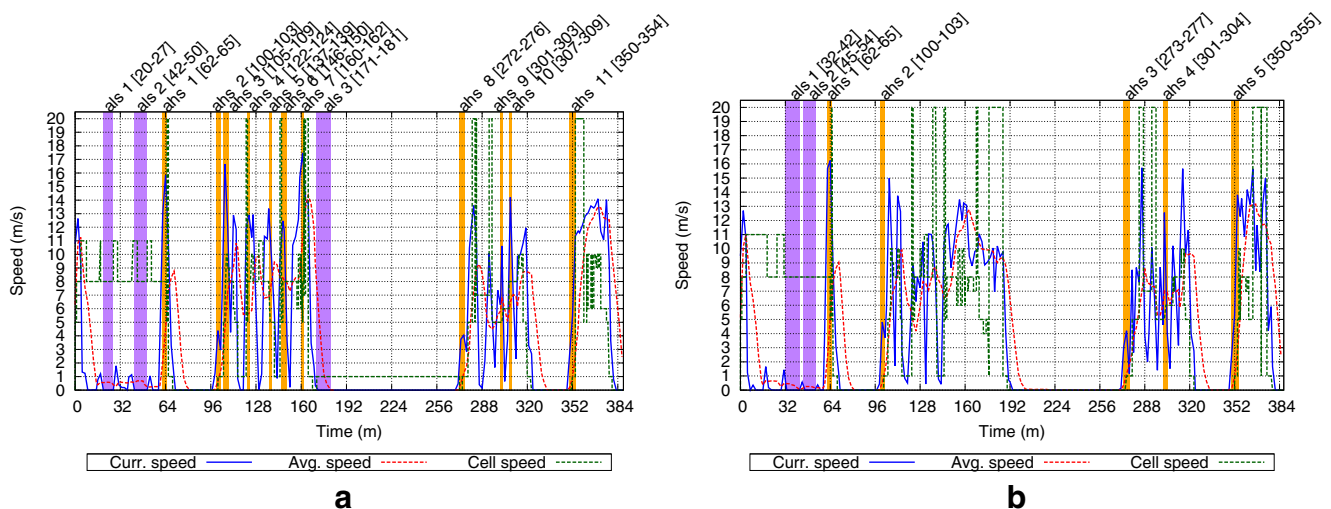


Fig. 5 RHIB 1's (a) and RHIB 2's (b) speed evolution throughout the trial and time periods during which the vessel did either abnormal low speed (als_i , purple areas) or abnormal high speed (ahs_i , orange areas). The time interval of each period is shown in brackets

depicts the speed of the cells in the location grid crossed by the RHIB 1 over the trial. Besides, the x-axis projection of the purple areas represent the time intervals during which the RHIB moved at an abnormally low speed, whereas the projection of the orange ones are the periods during which the boat moved abnormally fast. These periods covered a varied range of abnormal behaviours that may occur in a real environment differing to each other both in length and the reason why they were considered abnormal.

Table 3 shows the detection rate and the mean time to detect each of the aforementioned abnormal periods in x/y format. While x represents the detection rate, y indicates the detection time in seconds. In that sense, the different sampling rates included in the table were generated by sub-sampling the original RHIBs dataset.

As such a table shows, the system detected all the target abnormal situations given low sampling rates (10 s). This is because the current velocity and average velocity events made up by the system with these sampling rates reported the vessel’s speed in a fairly fine grain which allowed to keep track of the vessel’s speed with a great level of detail. This gave raise to high detection rates. In particular, the detection rate of the system was over the 88 % in all the cases as long as the sampling rate was below 60 s as Table 3 depicts.

As the sampling rate was increased, the detection rate of the system decreased. For instance, increasing the sampling rate to 480s made the system to detect only 1 out of 13 abnormal behaviours according to Table 3. This decrease is because, given high sampling rates, the incoming AIS-based

location and speed data is less detailed than with lower sampling rates. This caused that certain parts of the movement of the vessel become *invisible* for the system as they were reported neither in the incoming AIS data nor, as a result, in the velocity events created by the system.

Table 3 also shows that the shorter in time an abnormal behaviour is, the less likely to be detected by the system if the sampling rate is increased. For example, the behaviour ahs_{10} , whose time length was 120 s (from the minute 307 to the minute 309 of trial according to Fig. 5a), was not detected by the system given the 60-s sampling rate. This is because such rate was so large that the system did not to receive enough AIS data related this part of the RHIB 1’s trajectory so it became invisible for the system. Moreover, given the 480-s rate, the system only detected the behaviour als_3 . This was due to the fact that this behaviour’s length was about 10 minutes (from minute 171 to minute 181) which was large enough to be reported (at least partially) by the AIS data.

Apart from that, Table 3 also shows the mean time to detect (MTTD) each of the target speed-based abnormal behaviours. According to this table, the system required at least 120 seconds to detect any of the RHIB 1’s target behaviours even with low sampling rates. This is because of the t_{alert}^{min} parameter. As it was explained in Section 3.3.5, the system needs to aggregate possible-alert events during at least t_{alert}^{min} time units before delivering an alert. Thus, taking into account that for the present experiments this parameter was set to 120 seconds as Table 2 shows, the system needed at least that amount of time to deliver an alert informing of a behaviour. Furthermore, increasing the sampling rate also increased the MTTD. This is because a lower rate of incoming AIS data led to a slower generation of derived events which, in turn, affected to the time required to emit an alert.

As for RHIB 2, Fig. 5b depicts the current and average speed evolution of the vessel along with the speed of the crossed cells. In this case, the vessel moved abnormally slow at two periods of the trial (als_1 and als_2) because it remained stationary in an area where the boats usually moved faster (as the cells speed indicate in Fig. 5b). Regarding the abnormally-high speed periods (ahs_{1-5}), they were labelled that way because they involved sudden and unexpected accelerations of the vessel in areas where, in addition, vessels used to move much more slowly.

Table 4 shows the system’s detection rate and time of the behaviours listed above for various sampling rates. In this case, the system detected the two abnormally-slow behaviours (als_{1-2}) given all the sampling rates. As it was explained before, this was because the time length of these behaviours was large enough to be included in the AIS data delivered to the system and, in turn, in the derived velocity events used to generate the alerts. On the contrary, the system was not able to detect any of the high-speed abnormal

Table 3 Detection rate and mean time to detect (in seconds) the abnormal-speed periods for RHIB 1

Period	Sampling rate (s)			
	10	60	120	480
als_1	1.0/121	1.0/140	1.0/168	0.0/x
als_2	1.0/124	0.95/145	0.56/170	0.0/x
als_3	1.0/128	1.0/151	1.0/164	0.86/312
ahs_1	1.0/124	0.98/149	0.51/166	0.0/x
ahs_2	1.0/124	0.92/138	0.49/154	0.0/x
ahs_3	1.0/126	1.0/146	0.77/179	0.0/x
ahs_4	1.0/124	0.87/165	0.05/189	0.0/x
ahs_5	1.0/123	0.89/164	0.01/178	0.0/x
ahs_6	1.0/127	1.0/139	0.75/161	0.0/x
ahs_7	1.0/126	1.0/169	0.09/162	0.0/x
ahs_8	1.0/121	1.0/146	0.90/154	0.0/x
ahs_9	1.0/127	0.88/148	0.2/186	0.0/x
ahs_{10}	1.0/121	0.79/156	0.0/x	0.0/x
ahs_{11}	1.0/120	0.99/163	0.85/197	0.0/x
Avg.	1.0/122	0.88/151	0.51/171	0.06/312

Table 4 Detection rate and mean time to detect (in seconds) the abnormal-speed periods for RHIB 2

Period	Sampling rate (s)			
	10	60	120	480
<i>als</i> ₁	1.0/122	1.0/154	1.0/168	1.0/204
<i>als</i> ₂	1.0/124	1.0/157	1.0/175	1.0/208
<i>ahs</i> ₁	1.0/128	1.0/145	0.42/163	0.0/x
<i>ahs</i> ₂	1.0/121	1.0/149	0.39/168	0.0/x
<i>ahs</i> ₃	1.0/122	1.0/151	0.66/170	0.0/x
<i>ahs</i> ₄	1.0/129	0.95/165	0.47/181	0.0/x
<i>ahs</i> ₅	1.0/127	1.0/178	1.0/186	0.0/x
Avg.	1.0/125	0.99/157	0.71/173	0.29/206

behaviours (*ahs*_{1–5}) given the 480-s rate as the system did not receive enough AIS data of the vessel related to these behaviours to detect them.

Finally, Table 4 also depicts that the system needed at least 120 seconds in all the cases. As it was previously explained, this has to do with the t_{alert}^{min} parameter. Nevertheless, the system was able to detect most of the behaviours in less than 200 seconds.

4.1.3 Collision alerts

Over the trial, the two RHIBs were ordered to approach to each other at different locations and following several trajectories and speeds in order to simulate risky situations where the two vessels might collide to each other. As a result, Table 5 depicts these situations along with their time lengths and occurrence time.

Table 6 shows the detection rate achieved by the system for those possible-collision behaviours given different sampling rates of the incoming AIS data. Since the suspicious behaviours only lasted short time periods (only a few seconds) the system achieved a high detection rate given a low sampling rate (10s). The rationale of this is that a low sampling rate allows the system to control the location,

Table 5 List of possible-collision behaviours involving RHIB 1 and RHIB 2 generated in the trial

Collision	Length (s)	Trial's time instant
C-I	20	3 m
C-II	24	111 m
C-III	23	117 m
C-IV	19	157 m
C-V	14	176 m
C-VI	20	193 m
C-VII	8	197 m

Table 6 Detection rate and mean time to detect (in seconds) the possible-collision behaviours involving RHIB 1 and RHIB 2

Collision	Sampling rate (s)			
	10	15	20	30
C-I	1.0/4	0.75/12	0.01/19	0.0/x
C-II	1.0/6	0.81/13	0.07/17	0.0/x
C-III	1.0/8	0.76/16	0.06/21	0.0/x
C-IV	1.0/6	0.45/11	0.0/x	0.0/x
C-V	0.91/11	0.03/13	0.0/x	0.0/x
C-VI	1.0/9	0.79/15	0.0/x	0.0/x
C-VII	0.42/7	0.0/x	0.0/x	0.0/x
Avg.	0.90/7.3	0.51/13.7	0.02/19	0.0/x

speed and bearing of each vessel in a quite accurate way. Therefore, the more accurate and detailed these parameters are, the more likely the system to detect a possible collision is.

For the same reason, like the speed-based alerts, increasing the sampling rate made certain parts of the vessels' trajectory *invisible* for the system as they were not included in the AIS data that the system took as input. Considering that each of the possible collisions only lasted a few seconds, this lack of data remarkably reduced the collision-detection capabilities of the system.

Lastly, Table 6 also depicts the time the system required to detect each of the possible-collision behaviours. Although the time was affected by the sampling rate, the system was able to emit an alert only a few seconds after the suspicious behaviour had started. More specially, given a 10-s sampling rate, the system emitted a collision alert in less than 10 seconds after the situation started in most of the cases. Despite the fact that this amount of time might not be short enough to avoid the collision, at least it would be useful for a VTS staff to send its emergency resources to the alerted collision point as soon as possible.

4.1.4 Scalability study

The capability of the system to process different rates of AIS data was also studied given the memory constrains of the deployment platform. For that purpose, the two RHIB's traces were cloned to make up new pairs of *artificial* vessels. This was done by translating the original traces' points. As a result, it was possible to configure the number of vessels reporting their AIS data to the system by generating more or fewer artificial vessels.

Table 7 shows the maximum number of vessels the system was able to deal with given different sampling rates. As expected, the sampling rate meaningfully influenced the achieved values as it has a direct impact on the flow of

Table 7 Maximum number of vessels that the system was able to process given different sampling rates

Sampling rate (s)	Max. number of vessels
10	8943
60	9321
120	10476
480	12057
Avg.	10194

incoming events of the system. In more detail, system was able to process, on average, the raw events of roughly 10000 vessels at the same time. These results are quite promising if we consider the memory restrictions defined in the execution environment.

4.1.5 Comparative with reference framework

According to Kowalska and Peel (2012), the RF was able to detect three tracks of *unusual* behaviour from the RHIB dataset representing 1) drug smuggling, 2) human smuggling and 3) terrorism. In all the cases the involved vessel was the RHIB 1.

Figure 6 shows as coloured areas the time period covered by the aforementioned behaviours. For example, the drug smuggling behaviour detected by the RF started at the minute 120 of the trial and ended at minute 124. Figure 6 also depicts as dots the time instants at which our CEP system generated any speed-based alert involving RHIB 1. For each of the three RF behaviours we can make the following remarks,

- During the drug smuggling period (120 m-124 m), the CEP system firstly made up a set of loitering alerts between minutes 118-122. Next, it detected an abnormal high speed giving rise to several swift-movement

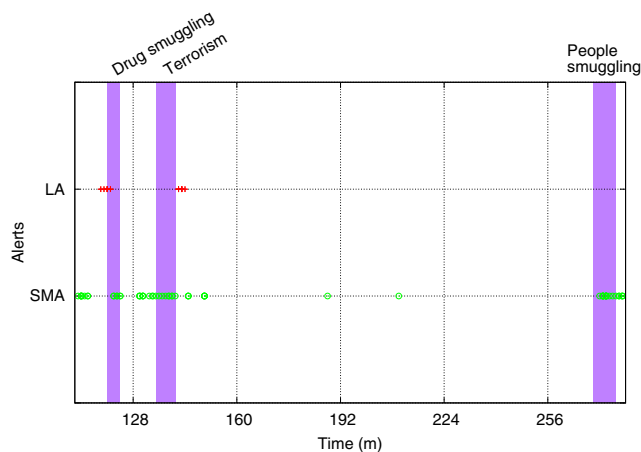


Fig. 6 Abnormal behaviours detected by the RF along with the loitering alerts (LA) and the swift movement alerts (SMA) generated by the system at the same time period

alerts during 123-125 m. This sequence of alerts is consistent with the RF outcome because it identified the potential drug smuggling scenario as period during with RHIB was moving very slowly followed by a sudden acceleration of the vessel to pick up some packages dropped into the sea by another vessel.

- The potential terrorism scenario (135 m-141 m) is perceived by the RF as a sudden acceleration of RHIB to drop some explosives near a ferry. As we can see from Fig. 6, during the same period the CEP system generated a set of consecutive swift-movement alerts informing of such an acceleration of RHIB 1.
- Finally, the people smuggling case (270 m-277 m) is detected by the RF when RHIB speeded to the shore to pick up some people and quickly returned to its original path. At the same instant, our CEP solution also made up a set of consecutive swift-movement alerts informing of the RHIB 1’s speed-up to approach the shore.

All in all, we can see that there exists a strong correlation between the three abnormal behaviours perceived by the RF and the alerts generated by CEP system. In that sense, the CEP system generated a set of alerts informing of abnormal speed of the target vessel during the same time periods covered by the behaviours. Nevertheless, a major difference exists between the two approaches. Whilst the RF needs to process long parts of a vessel’s track to perceive the abnormality, the CEP system generates earlier alerts as it only needs to process a few locations to perceive an abnormal speed.

Consequently, both approaches complement one another. Firstly, the CEP system can be used to provide early alerts about potential abnormal behaviours and, such suspicious situations can be confirmed by the RF later on. Hence, a VTS staff can be informer in a more timely and detailed manner.

4.2 Road-traffic case study

Dataset In order to test our solution in a completely different environment, the brinkhoff simulator (Brinkhoff 2002) was used to make up a synthetic dataset containing different trajectories on the road map of San Fransisco (USA). In this dataset, 5000 moving objects of 5 different types were simulated for a 400-time-unit period. At each time step, each moving object generated one location. In addition to that, the simulator was modified so that the speed of each moving object was always the maximum allowed of each edge. Finally, we used the default time and distance units of the generator.

Settings For this study, we used the same configuration for the system than the maritime case study (see Table 2).

However, for the cell grill, its reference point was the center of the San Francisco map and its cell size was set to 100x100m. This way, the number of cells in this case study was similar than the maritime one.

Methodology In order to evaluate our solution, we defined three measurements from the trajectories generated by the brinkhoff simulator.

- Firstly, the number of speed increments (NSI) indicates the number of pairs of consecutive edges of a vehicle's trajectory whose assigned maximum speed increase at least 40 %.
- Secondly, the number of speed decrements (NSD) counts the number of pairs of consecutive edges of a trajectory whose assigned maximum speed decrease at least 30 %. This measurement, along with the NSI, will be used to test the accuracy of the speed-based alerts.
- Lastly, the number of potential collisions (NPC) of a vehicle's trajectory indicates the number of times such a vehicle goes across a node (acting as crossroads) an another different vehicle also crosses the same node with a time difference less than t_{PCPA}^{max} time units. Thus, this measurement evaluates the accuracy of the collision alerts of the system.

4.2.1 Speed-based alerts

Figure 7 depicts the number of swift-movement and loitering alerts generated by the system given the different NSIs and NSDs of the experiment. For instance, for those trajectories comprising 10 NSIs the system generated, on average, 15 swift-movement alerts.

As we can see from Fig. 7, there exists a strong correlation between the NSI/NSD of a vehicle's trajectory and

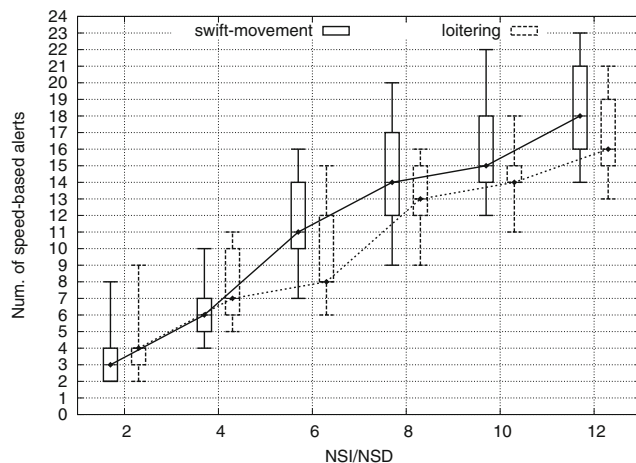


Fig. 7 Number of swift-movement and loitering alerts generated by system with respect to the NSI/NSD of a vehicle's trajectory

the number of speed-based alerts it gave rise. Nonetheless, the system usually generated more alerts than NSI/NSD. This is mainly because of the cell grill. As it was stated in Section 3.3.5, the average speed of the vehicles in each cell is taken into account by the system to decide whether a particular vehicle in such a cell is moving too fast or slow. However, in the road-traffic scenario, a particular cell might include several edges (roads) in its spatial area, and the maximum speed of each of these edges might be different to the others in the same cell. This might cause the average speed of the cell to be much higher or slower than the maximum speed of some of its edges. Consequently, all the vehicles driving along these edges will give rise to extra loitering of swift-movement alerts. These extra alerts explain why the system generates more alerts for a vehicle than its NSI/NSD.

4.2.2 Collision alerts

Regarding the collision alerts, Fig. 8 shows the number of these alerts generated by the system with respect to the NPC of each trajectory. For example, for those trajectories comprising 3 NPCs, the system generated, on average, 6 different collision alerts.

From Fig. 8 we can see that a strong correlation also exists between the NPC and the number of collision alerts generated by the system. Moreover, the system tended to generate more collision alerts for a vehicle than the NPC of its trajectory. That is because the system perceived certain movements of pairs of vehicles as potential collisions when such vehicles are actually moving along edges that do not intersect at all, and thus, a collision risk does not really exist.

To sum up, this case study has shown the potential feasibility of the proposal in a completely different scenario where the movement of the target objects is constrained by

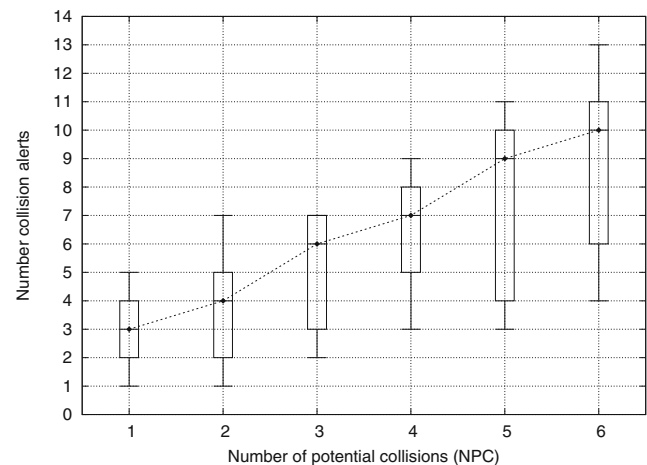


Fig. 8 Number of collision alerts generated by system with respect to the NPC of a vehicle's trajectory

a road network. Despite the fact that the accuracy of the system decreases with respect to the maritime case study, results show that it may be used by certain traffic information services to generate alerts in certain areas of the road network where the early detection of traffic problems is paramount.

5 Conclusion

The present work introduces the CEP paradigm as a novel approach to timely detect vessels' abnormal behaviours. To begin with, a set of three target behaviours has been identified from the results of different marine-surveillance workshops. Secondly, a CEP-based system devoted to detect these behaviours was developed as a proof of concept. This system has been designed to run as part of the infrastructure of a VTS, and it performs an event-based processing of the AIS data and the weather conditions. As output, the CEP system emits various abnormal-behaviour alerts.

The system has been tested in two case studies representing different types of movement. In the maritime one, results have shown, unsurprisingly, that there exists a strong dependency of the system to the sampling rate of the incoming AIS data and the time length of the target behaviours. Moreover, the experiment has shown a potential application of the solution to enrich other machine-learning proposals and provide them with early alerts of dangerous behaviours. In the road traffic case study, despite the network-constrained movement of the vehicles, the proposal has been able to detect several abnormal behaviours. Consequently, it might be used as a lightweight mechanism in critical areas where the detection of traffic problems must be done as soon as possible.

Further work will follow a twofold course of action. On the one hand, the system will be improved so as to process new geographical information related to regions of interest like fishing areas, ports or harbours. The second line of work will focus on the detection of new abnormal behaviours by making use of not only AIS data but also the aforementioned new data sources. This way, it is intended to come up with a CEP-based system that, by processing a varied range of event sources, will be able to offer a useful and almost real-time service to the VTS staff.

Acknowledgment Authors would like to thank BAES systems, and specially Jordi Barr, for kindly allowing them to use their datasets in the experiments section.

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 241598 (SeaBILLA project) and from the

Fundación Séneca Programme for Helping Excellent Research Groups 04552/GERM/0.

References

- Bomberger, N., Rhodes, B., Seibert, M., & Waxman, A. (2006). Associative learning of vessel motion patterns for maritime situation awareness. *Information Fusion, 2006 9th International Conference on*, 1–8.
- Brinkhoff, T. (2002). A framework for generating network-based moving objects. *Geoinformatica*, 6(2), 153–180.
- Cugola, G., & Margara, A. (2012). Processing flows of information: From data stream to complex event processing. *ACM Computing Surveys*, 44(3), 15:1–15:62.
- Esptech (2013). *Esper reference documentation*, version 4.9.
- Etzion, O., & Niblett, P. (2010). *Event processing in action*: Manning Publications.
- Evensen, P., & Meling, H. (2012). AdScorer: an event-based system for near real-time impact analysis of television advertisements. In *Proceedings of the 6th ACM International Conference on Distributed Event-Based Systems; DEBS '12* (pp. 85–94). New York, NY, USA: ACM.
- Garagic, D., Rhodes, B.J., Bomberger, N.A., & Zandipour, M. (2009). Adaptive mixture-based neural network approach for higher-level fusion and automated behavior monitoring. In *Communications, 2009. ICC'09, IEEE International Conference on* (pp. 1–6) IEEE.
- Guting, R.H., & Schneider, M. (2005). *Moving object databases*: Morgan Kaufmann Publishers.
- Idiri, B., & Napoli, A. (2012). The automatic identification system of maritime accident risk using rule-based reasoning. In *System of Systems Engineering (SoSE), 2012 7th International Conference on* (pp. 125–130): IEEE.
- Kowalska, K., & Peel, L. (2012). Maritime anomaly detection using gaussian process active learning. In *Information Fusion (FUSION), 2012 15th International Conference on* (pp. 1164–1171): IEEE.
- Kruger, M., Ziegler, J., & Heller, K. (2012). A generic bayesian network for identification and assessment of objects in maritime surveillance. In *Information Fusion (FUSION), 2012 15th International Conference on* (pp. 2309–2316) IEEE.
- Lane, R.O., Nevell, D.A., Hayward, S.D., & Beaney, T.W. (2010). Maritime anomaly detection and threat assessment. In *Information Fusion (FUSION), 2010 13th Conference on* (pp. 1–8): IEEE.
- Luckham, D. (2011). *Event processing for business: organizing the real-time enterprise*. Wiley & Sons.
- Mascaro, S., Korb, K.B., & Nicholson, A.E. (2010). *Learning abnormal vessel behaviour from ais data with bayesian networks at two time scales*: Tracks A Journal Of Artists Writings.
- Meister, S. (2012). In Niedrite, L., Strazdina, R., & Wangler, B. (Eds.) *Telemedical Events: Intelligent Delivery of Telemedical Values Using CEP and HL7* (Vol. 106 of Lecture Notes in Business Information Processing, pp. 1–13). Berlin: Springer.
- Navigation Center-United States Coast Guard (2013). *AIS messages*.
- Nuutinen, M., Savioja, P., & Sonninen, S. (2007). Challenges of developing the complex socio-technical system: realising the present, acknowledging the past, and envisaging the future of vessel traffic services. *Applied Ergonomics*, 38(5), 513–24.
- Patiniotakis, I., Papageorgiou, N., Verginadis, Y., Apostolou, D., & Mentzas, G. (2013). Dynamic event subscriptions in distributed event based architectures. *Expert Systems with Applications*, 40(6), 1935–1946.

- Ristic, B., La Scala, B., Morelande, M., & Gordon, N. (2008). Statistical analysis of motion patterns in AIS data: Anomaly detection and motion prediction. In *Information Fusion, 2008 11th International Conference on* (pp. 1–7) IEEE.
- Roy, J. (2008). *Anomaly detection in the maritime domain* Vol. 6945. 69450W–69450W–14.
- Roy, J. (2010). Rule-based expert system for maritime anomaly detection. In *Proceedings of SPIE-The International Society for Optical Engineering, USA*. 76662N–1, Vol. 7666.
- Roy, J., & Davenport, M. (2009). Categorization of maritime anomalies for notification and alerting purpose. In *NATO Workshop on Data Fusion and Anomaly Detection for Maritime Situational Awareness* (pp. 15–17).
- Terroso-Saenz, F., Valdes-Vela, M., Campuzano, F., Botia, J., & Skarmeta-Gomez, A.F. (2012). A complex event processing approach to perceive the vehicular context. *Information Fusion, 0*.
- Terroso-Saenz, F., Valdes-Vela, M., Sotomayor-Martinez, C., Toledo-Moreo, R., & Gómez-Skarmeta, A. (2012). A cooperative approach to traffic congestion detection with complex event processing and VANET. *IEEE Transactions on Intelligent Transportation Systems, 13*(2), 914–929.
- Valdés-Vela, M., & Gomez-Skarmeta, A.F. (2010). Unified management of heterogeneous sensor for complex event processing. In *Proceedings of the 4th International Symposium of Ubiquitous Computing & Ambient Intelligence*.
- van Laere, J., & Nilsson, M. (2009). Evaluation of a workshop to capture knowledge from subject matter experts in maritime surveillance. In *Information Fusion, 2009. FUSION'09, 12th International Conference on* (pp. 171–178): IEEE.
- Verginadis, Y., Papageorgiou, N., Patiniotakis, I., Apostolou, D., & Mentzas, G. (2012). A goal driven dynamic event subscription approach. In *Proceedings of the 6th ACM International Conference on Distributed Event-Based Systems; DEBS '12* (pp. 81–84). New York, NY, USA: ACM.
- Vespe, M., Visentini, I., Bryan, K., & Braca, P. (2012). Unsupervised learning of maritime traffic patterns for anomaly detection. In *Data Fusion Target Tracking Conference (DF TT 2012): Algorithms Applications, 9th IET* (pp. 1–5).
- Will, J., Peel, L., & Claxton, C. (2011). *Fast maritime anomaly detection using kd-tree gaussian processes*: IMA Maths in Defence Conference.

Dr. Fernando Terroso-Saenz graduated from University of Murcia with a degree in Computer Science in 2006. He also received the master's degree in Computer Science at the same university in 2010. Then, he finished his PhD at the Department of Communications and Information Engineering in 2013. Since 2009, he has been working as a researcher in this group. His research interests include Complex Event Processing (CEP), Ubiquitous Computing and Intelligent Transportation Systems (ITSs).

Dr. Mercedes Valdes-Vela received the B.S. degree in computer engineering and the Ph.D. (Hons.) Degree in computer science in 2003 from the University of Murcia, Murcia, Spain. In 1999, she was a Research Staff with the department where she has been involved in several European research projects and is currently a full-time Assistant Professor. Her research interests include complex event processing, ambient intelligence, and fuzzy modelling.

Dr. Antonio F. Skarmeta-Gomez received the M.S. degree in Computer Science from the University of Granada and B.S. (Hons.) and the Ph.D. degrees in Computer Science from the University of Murcia Spain. Since 2009 he is Full Professor at the same department and University. He has worked on different research projects in the national and international area, like IoT6, Openlab, SWIFT, GEN6.