Abnormal Trajectory-Gap Detection: A Summary

Arun Sharma[1]
Department of Computer Science & Engineering, University of Minnesota, Minneapolis, MN, USA

Jayant Gupta
Department of Computer Science & Engineering, University of Minnesota, Minneapolis, MN, USA

Shashi Shekhar
Department of Computer Science & Engineering, University of Minnesota, Minneapolis, MN, USA

Abstract

Given trajectories with gaps (i.e., missing data), we investigate algorithms to identify abnormal gaps for testing possible hypotheses of anomalous regions. Here, an abnormal gap within a trajectory is defined as an area where a given moving object did not report its location, but other moving objects did periodically. The problem is important due to its societal applications, such as improving maritime safety and regulatory enforcement for global security concerns such as illegal fishing, illegal oil transfer, and trans-shipments. The problem is challenging due to the difficulty of interpreting missing data within a trajectory gap, and the high computational cost of detecting gaps in such a large volume of location data proves computationally very expensive. The current literature assumes linear interpolation within gaps, which may not be able to detect abnormal gaps since objects within a given region may have traveled away from their shortest path. To overcome this limitation, we propose an abnormal gap detection (AGD) algorithm that leverages the concepts of a space-time prism model where we assume space-time interpolation. We then propose a refined memoized abnormal gap detection (Memo-AGD) algorithm that reduces comparison operations. We validated both algorithms using synthetic and real-world data. The results show that abnormal gaps detected by our algorithms give better estimates of abnormality than linear interpolation and can be used for further investigation from the human analysts.

2012 ACM Subject Classification Information systems → Data mining; Computing methodologies → Spatial and physical reasoning

Keywords and phrases Spatial Data Mining, Trajectory Mining, Time Geography

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.26

Category Short Paper

Funding This research is funded by an academic grant from the National Geospatial-Intelligence Agency (Award No. HM0476-20-1-0009, Project Title: Abnormal Trajectory Gap Detection). Approved for public release, 22-379.

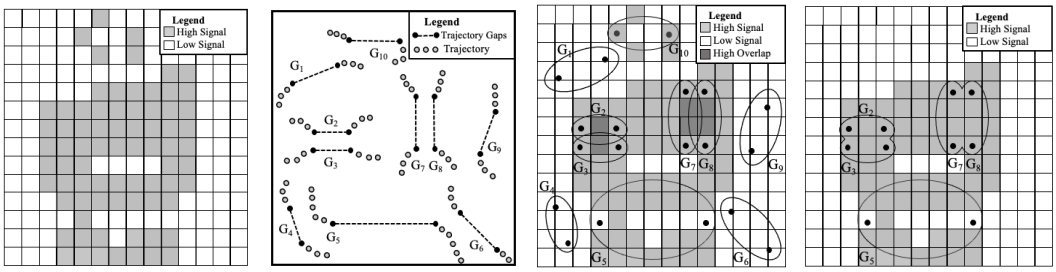
Acknowledgements We also want to thank Kim Koffolt and the spatial computing re- search group for their helpful comments and refinements.

1 Introduction

Given multiple trajectory gaps and a signal coverage map based on historic object activity, we find possible abnormal gaps in activity where moving objects (e.g., ships) may have behaved abnormally such as not reporting their locations in an area where other ships historically did report their location. Figure 1 shows the problem's *input* which includes a map of the signal coverage area (grey cells) for a set of derived historic trajectories (shown in Figure 1a) and trajectory gaps G_1, ... , G_{10} (shown in Figure 1b). Figure 1 also shows the output where

[1] Corresponding author

gaps G_1, G_4 G_6 and G_9 are entirely outside the signal coverage map indicating weak signal coverage. In contrast, the rest of the gaps are overlapping the signal coverage map. The absence of location reporting in an area known to have signal coverage may be interpreted as intentional behavior by a ship that temporarily switched off its location broadcasting device. Figure 1c shows *intermediate output* stage where trajectory gaps (G_1, ... , G_{10}) are modelled in the form of geo-ellipses [12, 10] along with their intersections with the signal coverage map. Figure 1c also shows two-gap pairs, G_2, G_3 and G_7, G_8, that have intersecting regions (i.e., dark grey cells), which suggests two ships in a rendezvous potentially engaging in illegal activity. The gaps entirely outside the signal coverage area have been filtered out. Figure 1d shows the *final output* with gap pairs G_2, G_3 and G_7, G_8 are merged to their overlapping regions and gap G_{10} is filtered out since it did not meet the user defined *priority threshold*.



(a) Input 1: Signal Coverage Map (SCM). **(b) Input 2:** Trajectories with Trajectory Gaps (G_i). **(c) Intermediate Output**: Abnormal Gap Computation. **(d) Final Output:** Summarized Abnormal Gaps.

Figure 1 An illustration of the Abnormal Gap Region Problem (Best in color).

Analyzing trajectory gaps has many societal applications in maritime safety, homeland security, epidemiology, and public safety. Other use cases include tracing comets, tracking marine animals and contact tracing. In this paper, we focus on understanding the potential benefits of signal coverage mapping to improve maritime safety and regulatory enforcement. A signal coverage map helps to reduce false positives by providing historical activity traces for a region, against which abnormal behavior can be detected within a trajectory gap. In addition, current methods assume shortest path in a trajectory gap which leads to many missed patterns since moving objects do not always travel in a straight path. In contrast, our approach is based on a space-time prism, which can accommodate greater movement possibilities around a signal gap where an object of interest could have potentially deviated from the predefined (or known) linear path. We also provide a way to reduce computational cost over a large geographical space.

Related Work. Surveys [2, 18] provide a broad classification of anomaly detection methods in trajectory mining. Riveiro et al. [13] provide a holistic view of maritime anomaly detection but do not cover trajectory gaps. Works that do consider trajectory gaps [11, 3, 8] assume shortest path methods. For instance, the proposed framework in [11] extracts maritime movement patterns assuming shortest path within trajectory gaps.

There are some realistic frameworks [4, 15, 16] that employ reconstruction techniques for modeling uncertainty in trajectory gaps based on space-time prism models [10, 12]. However, they are limited to theoretical simulations and little attention is given to real-world applications [5]. For instance, Winter [16] provide a probabilistic interpretation of a space-time prism. However, the interpretation lacked a real-world validation. A recent work [17]

considers trajectory gaps using a space-time prism but does not consider abnormal behavior within gaps. This paper uses a space-time prism model to capture abnormal gaps based on historical data which allows us to distinguish abnormal gaps for possible anomaly hypothesis.

Contributions. We define an Abnormal Gap Measure (AGM) for modeling abnormal gaps and propose an abnormal gap detection (AGD) algorithm to handle multiple gaps. In addition, we propose a memoized abnormal gap detection (Memo-AGD) algorithm to further improve computational efficiency. We show experimentally that our methods are efficient and have good solution quality. We also present a real-world case study to validate our approach.

Scope and Organization. This work is limited to space-time prisms for computing abnormal gaps and methods such as kinetic prisms [7] are not studied. This paper did not consider acceleration in space-time prism due to data limitations. The use of signal coverage maps based on aerial imagery datasets (e.g., satellite imagery) and it's it's falls outside the scope of this paper. The rest of the paper is organized as follows: Section 2 introduces key concepts and formally defines the abnormal gap detection problem. Section 3 describes the proposed algorithms AGD and refined Memo-AGD. Experiments and Results are presented in Section 4. Finally, Section 5 concludes this work and briefly lists the future work.

2 Problem Definition

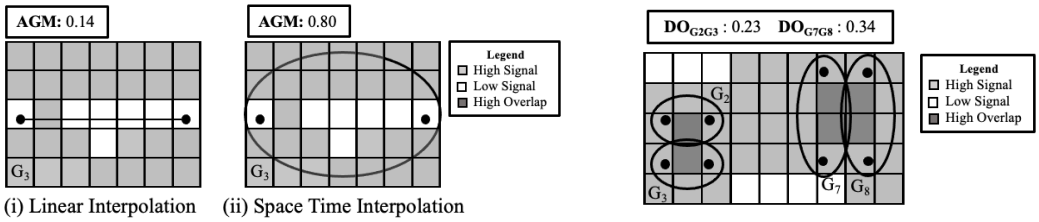
2.1 Basic Concepts

▶ **Definition 1.** *A **signal coverage map (SCM)** is a discretized grid space where cells are color-coded to represent regions with regular historically reported location signals.*

The maps are generated by first computing the total reported ship movement in the area using a set of historic location-traces P_i (i.e., trajectories) within some time interval and then checking if the total is above a given threshold (say θ). Figure 1a shows a synthetic example of signal coverage map based on historical ship movement with binary color-coding. The grey cells in the figure have movement above the threshold (i.e., High Signal), whereas the white cells have low or no reported historical movement (i.e., Low Signal).

▶ **Definition 2.** *An **abnormal gap measure (AGM)** for a gap G_i is the probability that a possible location of the object during the gap (unreported data time interval) has signal coverage. A higher value of AGM indicates anomalous behavior since it means an object is not reporting its location despite having the location signal coverage in the past.*

The probability is computed using interpolated grid cells (GC_{int}) and regions with high historical movement (GC_m). We first compute the overlap between GC_{int} and GC_m, and then normalize the overlap with GC_{int}. Figure 2a provides two examples of computing the abnormal gap measure between two points using cells color-coded based on SCM. The first shows the AGM for linear interpolation, where GC_{int} is the set of cells crossed or touched by a line between the points (i.e., 7) which overlaps with only 1 cell with a high historically reported movement. Thus, the AGM is 0.14 ($\frac{1}{7}$). The second shows the AGM based on (GC_{int}) for space-time interpolation to accommodate additional (movement) possibilities. As shown, GC_{int} is the number of cells crossed or touched by the ellipse (i.e., 35) which overlaps with 28 cells with a high historically reported movement. Thus, the AGM is 0.80 ($\frac{28}{35}$).



(a) Abnormal Gap Measure. **(b)** Degree of Overlap.

Figure 2 Examples of Abnormal Gap Measure (Left) and Degree of Overlap (Right).

▶ **Definition 3.** *Given a pair of gaps (G_A, G_B), the* ***degree of overlap ($DO_{G_AG_B}$)*** *for two pairs is the minimum ratio between the common interpolated cells to the interpolated cells.*

Figure 2b shows two gap pairs G_2, G_3 and G_7, G_8. The degree of overlap for the first pair is $\mathrm{DO}_{G2,G3} = 0.23(min[\frac{2}{9}, \frac{2}{9}])$ and for the second pair $\mathrm{DO}_{G7,G8} = 0.33(min[\frac{4}{12}, \frac{4}{12}])$.

2.2 Problem Formulation

Input.

(1) Trajectory Gaps: Missing location signal(s) between two consecutive points. Fig. 1b represent trajectory gaps G_1 to G_{10} which later modelled as geo-ellipses (Fig. 1c-1d).

(2) Signal Coverage Map (SCM): The primary motivation behind a signal coverage map (Figure 1a) is to improve the accuracy of detecting specific trajectory gaps where a moving object (e.g., ships) may have behaved abnormally. If we consider the entire study area, we may increase the rate of false positives since not reporting locations from moving objects may be due to weak signal coverage in certain regions. Hence, these maps provide a way to narrow down the search space to specific gaps which need further investigation.

(3) Priority Threshold: A threshold value used for extracting abnormal gap based on user's preference. For instance, gaps with AGM scores *above* a priority threshold are extracted and prioritized for further investigation (e.g., G_2, G_3, G_5, G_7, G_8 in Figure 1d).

Output. Summarized abnormal trajectory gaps (as shown in Figure 1d). Here, we first filter out trajectory gaps that are within areas without signal coverage and then use the priority threshold to prioritize gaps with a high AGM score. In addition, we also coalesce the gaps for two or more geo-ellipse intersections, which usually occur in dense regions, and reduce additional scanning of the overlapped area (e.g., Figure 1c) while considering one gap at a time. This results in higher post-processing costs by the human analysts, which can be greatly reduced by merging the common intersection region.

Objective. Our objective is solution quality and computational efficiency. Solution quality can be achieved by reducing the false positive rate and using optimal AGM values to detect abnormal gaps. To enhance computational efficiency, we focus on optimizing gap enumeration during the process of forming clusters of gaps that interact spatially and temporally.

Constraints. The space-time prism do not consider acceleration and deceleration [7] and only consider speed derived from start and end point of a gap. Signal coverage map's stability and effect of weather, environmental conditions, and radio outages are not considered.

3 Proposed Approach

Framework. Our aim is to identify possible abnormal gaps on a given set of trajectory gaps and signal coverage area through a three-phase *Filter* and *Refine* approach. The trajectory signals are first preprocessed to filter out all the trajectory gaps. Then we model gaps as geo-ellipses and apply the proposed algorithms to effectively optimize the spatial interactions by coalescing pair of gaps and reducing redundant linear scans on the overlap as shown in Figure 1c and 2b. Then, the output is a summary of significant abnormal gaps (Figure 1d) which helps a human analyst for ground truth verification via satellite imagery.

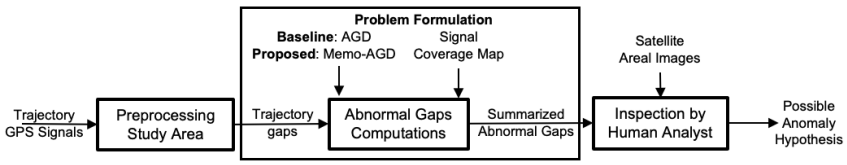


Figure 3 Framework for detecting possible abnormal gaps to reduce manual inspection by analyst.

(1) Abnormal Gap Detection (AGD) Algorithm. Here we describe the abnormal gap detection (AGD) algorithm used for enumerating gaps such that each gap is modeled via a geo-ellipse and an AGM score is computed with a coalescing operation (Figure 1d). First, we sort each gap G_i by time and then check which gap pairs intersect in space and time. We further calculate their Degree of Overlap (DO) and check if DO $\geq \lambda$. Second, we save each qualified gap either as a single ($<G_i>$) or as a subset of gap pairs (e.g., $<G_i,G_j>$). Finally, we merge all subsets of trajectory gaps and compute their AGM scores, which are later extracted via a priority threshold. An execution trace is given in Appendix A.

(2) Memoized Abnormal Gap Detection (Memo-AGD) Algorithm. Since the AGD approach enumerates an exponential number of candidates, we use additional variables such as G_i^{Obs} and G_i^{LU}, where G_i^{Obs} keeps track of the total current elements in an Observed List and G_i^{LU} provides a *lookup table* which allow us to store information which was already involved in a prior intersection with G_i. Such *memoization* avoids unnecessary gap enumeration.

Algorithm 1 Memoized Abnormal Gap Detection (Memo-AGD) Algorithm.

Input : Trajectory Gaps (G_i), Signal Coverage Map (SCM) and DO Threshold λ
Output: Summarized Abnormal Trajectory Gaps

1: **procedure** :
2: **Step 1:** Initialize LookUp Table to G_i^{LU} and copy Observed List to G_i^{Obs}
3: **Step 2:** Check spatiotemporal overlap and avoid gap enumeration
4: **while** $G_i^{Obs} \neq \emptyset \; \forall \; G_j \in G_i^{Obs}$ **do**
5: **while** $G_i \cap G_j \neq \emptyset$ and DO $\geq \lambda$ **do**
6: Save or Update the derived shape from $G_i \cap G_j$ to G_i^{LU} and G_j^{LU}
7: Add $<G_i,G_j>$ to G_i^{LU} and G_j^{LU} and update $G_i^{Obs} \leftarrow G_i^{Obs} - G_j^{LU}$
8: If $G_i \cap G_j = \emptyset$ then update $G_i^{Obs} \leftarrow G_i^{Obs} - G_j^{LU}$
9: **Step 3:** Compute AGM Score $\forall$ elements in LookUp Table G_i^{LU}

First we initialize G_i^{LU} and copy the current Observed List in G_i^{Obs}. After checking for spatiotemporal overlap and the DO $\geq \lambda$ condition, we update and save the resultant shape derived from $G_i \cap G_j$ and subset $<G_i,G_j>$ to G_i^{LU} and G_j^{LU}. For instance, a new gap G_k

only needs to perform one comparison with $G_i \cap G_j$ with subset $<G_i, G_j>$ saved in G_i^{LU} and G_j^{LU}. In addition, G_k will skip a comparison with G_j via G_i^{Obs} - G_j^{LU} providing further computational speedup. An execution of the Algorithm 1 trace is given in Appendix B.

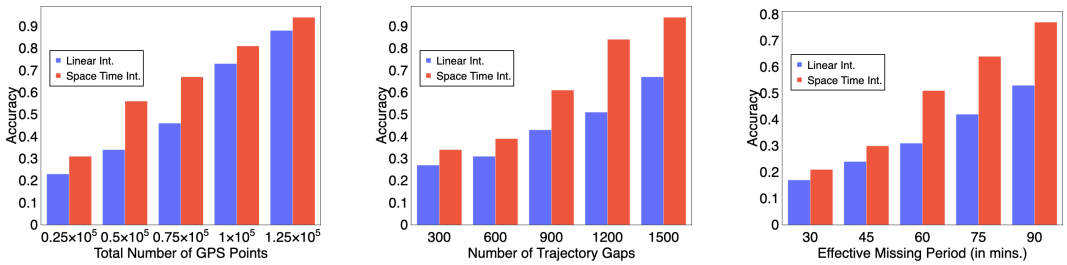
4 Experimental Validation

Synthetic Data Generation. For *solution quality* experiment, we lack ground truth data (i.e., absence of information on whether each gap is abnormal or not). Therefore, we evaluated the proposed algorithms on synthetic data derived from a real-world dataset. First, we gather trajectories gathered on a fixed study area ranging from $179.9W$ to $171W$ degrees in longitude and from $50N$ to $58N$ degrees in latitude in the Bering Sea with 1.25 x 10^5 with 1500 trajectory gaps spanning from 2014 to 2016. Then, for each object, we preprocessed trajectory points with a time-gap duration greater than 30 mins to qualify it as a trajectory gap. Finally, we calculated an AGM score using linear interpolation and proposed methods and classified each gap as abnormal and non-abnormal based on a specific priority threshold (i.e., **0.6**). For instance, gaps with AGM scores greater than **0.6** are considered abnormal.

Real World Data. We used MarineCadastre [1], a real world dataset containing many attributes (e.g., Longitude, Latitude, Speed Over Ground etc.) for 150, 000 objects from 2009 to 2017. In addition, we also used MarineTraffic [9] for a case-study near the Galapagos Islands to verify the effectiveness of the proposed algorithms.

Computing Resources. We performed our experiments on a system with a 2.6 GHz 6-Core Intel Core i7 processor and 16 GB 2667 MHz DDR4 RAM.

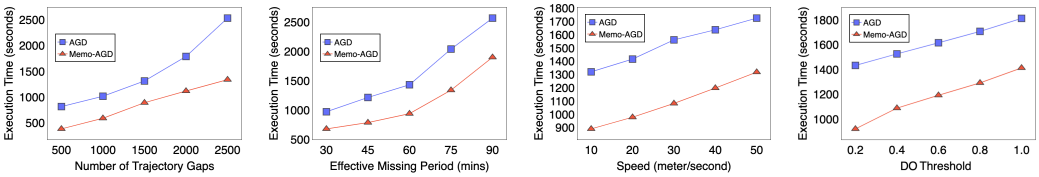
Experiments for Solution Quality. We compared the accuracy of a linear interpolation based method [3, 14] and our space-time interpolation based AGD and Memo-AGD algorithms. The solution quality was based on fixed number of abnormal trajectory gaps resulted from both linear and space-time interpolation methods and a fixed priority threshold. The output gaps are then compared with a predefined normal-abnormal ground truth from synthetic data for computing accuracy. We varied three parameters: number of GPS points, number of trajectory gaps, and effective missing period (EMP), i.e., the total time when a given object was missing. Figure (4a - 4c) shows that space-time interpolation outperforms linear interpolation on all three parameters. The reason is that the AGM scores captured are more accurate in space-time interpolation as compared to linear interpolation based methods.



(a) GPS Points. **(b)** Trajectory Gaps. **(c)** EMP.

Figure 4 Spatio-temporal interpolation is better than linear interpolation for different parameters.

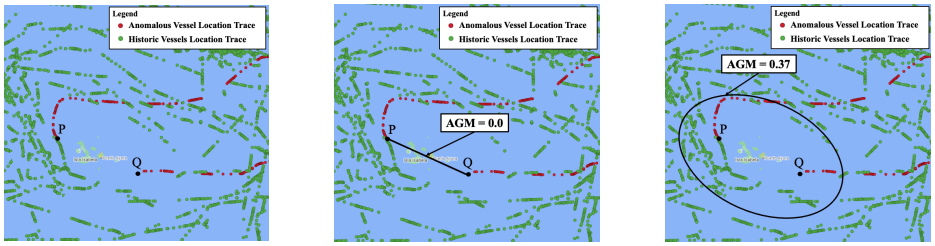
Experiments for Computational Efficiency. Next, we compared Memo-AGD against AGD based on computation efficiency using the MarineCadastre [1] dataset. Figure 5a shows that Memo-AGD consistently outperforms AGD. The reason is that the potential interactions of two or more geo-ellipses increase as we increase the number of trajectory gaps. Figure 5b shows Memo-AGD outperforms AGD as we increase EMP since larger geo-ellipses result in a higher number of potential interactions. Figure 5c also shows MemoAGD is faster than AGD. The reason is that high S_{max} (i.e., maximum possible speed an object can attain during the EMP of its trajectory gap) produces larger geo-ellipses, resulting in more potential interactions. Figure 5d shows that higher DO threshold means gaps are less likely to intersect with each other. This helps in avoiding large coalesced gap pairs which result in higher ground-truth verification cost. Hence, MemoAGD is more efficient than the AGD algorithm.



(a) Trajectory gaps. **(b)** Effect of EMP. **(c)** Effect of S_{max}. **(d)** Effect of λ.

Figure 5 Comparison of execution times of AGD and Memo-AGD under different parameters

Case Study. We conducted a case study of an actual illegal shipping event [6] by applying the proposed Memo-AGD algorithm on a set of real-world data from MarineTraffic [9] data consisting of around 10^5 location trace broadcasted from over 121 fishing vessels for October 2014 (i.e., 31 days) covering the Galapagos marine reserve. Figure 6 shows where a fishing vessel (in red) inside a protected habitat area reported no signals for 15 days because it had switched off its transponder. We applied our Memo-AGD algorithm and a linear interpolation method on the historic location traces (in green). In linear interpolation, the vessels interacted small number of island and did not intersected with any historic location trace (Figure 6b) resulted in an AGM score of **0.0**, indicating no abnormal activity detected. By contrast, the Memo-AGD algorithm returned a geo-elliptical area (Figure 6c) within which some historic trajectories resulted in an AGM score of **0.37**, which accords with the known abnormal activity. The domain experts can set a priority threshold of 0.3 in the flagged area for further investigation by the human analysts.



(a) No Interpolation. **(b)** Linear Interpolation. **(c)** Memo-AGD.

Figure 6 Comparison of Linear Interpolation and Memo-AGD around Galapagos Marine Reserve

Discussion. In this paper, we used a signal coverage map (SCM) as input to significantly reduce the false-positive rate by filtering out gaps in zones with historically weak signal coverage. However, estimating signal coverage is challenging if the vessel enters certain regions (e.g., the arctic) where limited or no historical trace data exists. This result in cold start problem [17]. In addition, the proposed methods results in similar execution times in case of no potential interactions among geo-ellipses (more details described in Appendix C).

In addition, none of the local gaps exceed the time range of the signal coverage map. For instance, Figure 6 shows the time range of the largest signal gap (i.e., 15 days) does not exceed the signal coverage spanned for entire month. Furthermore, given the object disappeared for 15 days in Figure 6, considering speed only at the P and Q Figure 6 while modeling trajectory gaps that may produce relatively smaller ellipse region.

5 Conclusion and Future Work

We investigated Abnormal Gap Measure and proposed AGD and Memo-AGD algorithms. We performed experimental evaluation under varying parameters where the results show that compared to linear interpolation methods, space-time interpolation based methods [12] are better at detecting abnormal gaps in trajectory data. Also, Memo-AGD is computationally more efficient than AGD. In the future, we plan to develop spatiotemporal statistics and conduct a statistical significance test to eliminate chance patterns. We will add other use cases such as a cold-start problem [17] and include acceleration using kinetic prisms [7]. Finally, we plan to explore datasets in network space for other potential societal use-cases.

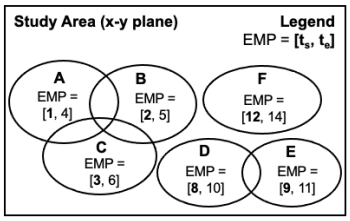
References

1 BOEM and NOAA. Marinecadastre. `https://marinecadastre.gov/ais/`, 2020.
2 V. Chandola et al. Anomaly detection: A survey. *CSUR*, 41(3):1–58, 2009.
3 C. Chen et al. Isolation-based online anomalous trajectory detection. *IEEE TIST*, 2013.
4 R. Cheng et al. Probabilistic verifiers: Evaluating constrained nearest-neighbor queries over uncertain data. In *2008 IEEE 24th ICDE*, pages 973–982. IEEE, 2008.
5 S Dodge et al. Towards a taxonomy of movement patterns. *IV*, 7(3-4):240–252, 2008.
6 S Gibbens. How illegal fishing is being tracked from space. *National Geographic*, 2018.
7 B. Kuijpers et al. Kinetic prisms: incorporating acceleration limits into space time prisms. *IJGIS*, 31(11):2164–2194, 2017.
8 Po-Ruey Lei. A framework for anomaly detection in maritime trajectory behavior. *Knowledge and Information Systems*, 47(1):189–214, 2016.
9 MarineTraffic. Galapagos island case study. `https://www.marinetraffic.com/en/ais/`, 2020.
10 H. J. Miller. Modelling accessibility using space-time prism concepts within geographical information systems. *IJGIS*, 5(3):287–301, 1991.
11 G. Pallotta, M. Vespe, and K. Bryan. Vessel pattern knowledge discovery from ais data: A framework for anomaly detection and route prediction. *Entropy*, 15(6):2218–2245, 2013.
12 D. Pfoser and C. S. Jensen. Capturing the uncertainty of moving-object representations. In *International Symposium on Spatial Databases*, pages 111–131. Springer, 1999.
13 M. Riveiro, G. Pallotta, and M. Vespe. Maritime anomaly detection: A review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(5):e1266, 2018.
14 Robert et. al Skulstad. Dead reckoning of dynamically positioned ships: Using an efficient recurrent neural network. *IEEE Robotics & Automation Magazine*, 26(3):39–51, 2019.
15 Goce Trajcevski et al. Uncertain range queries for necklaces. In *2010 Eleventh International Conference on Mobile Data Management*, pages 199–208. IEEE, 2010.
16 Stephan Winter and Zhang-Cai Yin. Directed movements in probabilistic time geography. *International Journal of Geographical Information Science*, 24(9):1349–1365, 2010.

17 Pengxiang Zhao, David Jonietz, and Martin Raubal. Applying frequent-pattern mining and time geography to impute gaps in smartphone-based human-movement data. *International Journal of Geographical Information Science*, 35(11):2187–2215, 2021.
18 Yu Zheng. Trajectory data mining: an overview. *ACM TIST*, 6(3):1–41, 2015.

A Execution Trace of AGD Algorithm

Figure 7 shows the execution trace of the AGD algorithm with gap (<A>,1)„.,(<F>,1) where each gap set is associated with a single gap (e.g., <A>) and its count (i.e., 1). We then sort each gap G_i based on its start time t_s and at t_s=1 and we add (<A>,1) to the Observed List since the list is empty. At t_s=2, we add (<B>,1) since it satisfies spatiotemporal intersection criteria resulting in (<A>,2),(<B>,2),(<A,B>,2) and incrementing count for each subset by 1. A similar operation is performed at t_s=3, resulting in (<A>,3),(<B>,3),.., (<A,B,C>,3). However at t_s=8, we only add (<D>,1) since it does not interact with any of the elements in the Observed List. At t_s=9, (<E>,1) interacts with (<D>,1) performing similar operations as executed at $t_s = 2$ resulting in (<D>,2),(<E>,2) and (<D,E>,2). At $t_s = 12$, set (<F>,1) is added to the Observed List similar to (<D,1>) since it does not interact with any of the elements in Observed List. Finally we filter out (<A,B,C>,3),(<D,E>,2) and (<F>,1) since the number of elements is equal to their respective counts and we compute AGM via the signal coverage map. We then output the summarized abnormal gaps based on their AGM scores and the priority threshold given by the human analysts.

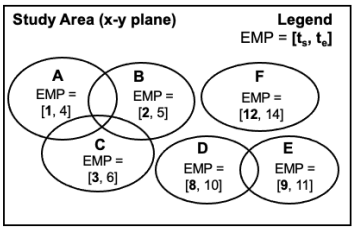


Time	Non-Observed List (G_i)	Observed List (G_j)
t_s = 1	{<B>,1}, {<C>,1}, {<D>,1}, {<E>,1}, {<F>,1}	{<A>,1}
t_s = 2	{<C>,1}, {<D>,1}, {<E>,1}, {<F>,1}	{<A>,2}, {<B>,2}, {<A,B>,2}
t_s = 3	{<D>,1}, {<E>,1}, {<F>,1}	{<A>,3}, {<B>,3}, {<C>,3}, {<A,B>,3} , {<B,C>,3}, {<A,C>,3}, {<A,B,C>,3}
t_s = 8	{<E>,1}, {<F>,1}	{<A>,3}, {<B>,3}, {<C>,3}, {<A,B>,3} , {<B,C>,3}, {<A,C>,3}, {<A,B,C>,3}, {<D>,1}
t_s = 9	{<F>,1}	{<A>,3}, {<B>,3}, {<C>,3}, {<A,B>,3} , {<B,C>,3}, {<A,C>,3}, {<A,B,C>,3}, {<D>,2}, {<E>,2}, {<D,E>,2}
t_s = 12	Empty	{<A>,3}, {<B>,3}, {<C>,3}, {<A,B>,3} , {<B,C>,3}, {<A,C>,3}, **{<A,B,C>,3}**, {<D>,2}, {<E>,2}, **{<D,E>,2}, {<F>,1}**

Figure 7 Execution trace of the Baseline AGD algorithm.

B Execution Trace of Memo-AGD Algorithm

Figure 8 shows the execution trace of the Memo-AGD algorithm. At $t_s = 1$, Step 1 initializes G_i with variables G_i^{LU} and G_i^{Obs} which are later added to the empty Observed List. At $t_s = 2$, $<A^{Obs}>$ copies the *current* Observed List (i.e., only <A>) to $<B^{Obs}>$ after satisfying spatial and temporal overlap conditions. If true, we perform Step 2 where the resultant shape of $<A> \cup <B>$ is saved in both <A> and <B>. The variables LU gets updated with <A,B> for both $<A^{LU}>$ and $<B^{LU}>$. Finally, removing <A> from $<B^{Obs}>$ results in an empty list and the loop terminates. At $t_s = 3$, <C> interacts with the resultant shape of both <A> and <B> (i.e., $<A> \cup <B>$ and result $<A^{LU}>$,$<B^{LU}>$ and $<C^{LU}>$ as <A,B,C> (i.e., maximal sets). However at $t_s = 8$, <D> does not intersect spatially with <A> and will skip intersection operations with other elements of A^{LU} (i.e., <B> and <C>) via G^{Obs} -G^{LU} operation and gets added to the Observed List and result in performance speedup as compared to the AGD algorithm. A similar operation is done at $t_s = 9$, where <E> does not interact with <A> but does interact with <D>, resulting in D^{LU} = <D,E>. Finally, <F> does not interact with any of the elements, causing to be directly saved as maximal set <F>. The rest of the steps remain the same as the baseline except we gather all maximal sets from the lookup table of each G^{LU} in the Observed List.



Time	Non-Observed List (G_i)	Observed List (G_j) with Look Up Table (G_j^{LU})
$t_s = 1$	B^{LU} = {<B>}, C^{LU} = {<B>}, D^{LU} = {<D>}, E^{LU} = {<E>}, F^{LU} = {<F>}	A^{LU} = {<A>}
$t_s = 2$	C^{LU} = {<B>}, D^{LU} = {<D>}, E^{LU} = {<E>}, F^{LU} = {<F>}	A^{LU} = {<A,B>}, B^{LU} = {<A,B>}
$t_s = 3$	D^{LU} = {<D>}, E^{LU} = {<E>}, F^{LU} = {<F>}	A^{LU} = {<A,B,C>}, B^{LU} = {<A,B,C>}, C^{LU} = {<A,B,C>}
$t_s = 8$	E^{LU} = {<E>}, F^{LU} = {<F>}	A^{LU} = {<A,B,C>}, B^{LU} = {<A,B,C>}, C^{LU} = {<A,B,C>}, D^{LU} = {<D>}
$t_s = 9$	F^{LU} = {<F>}	A^{LU} = {<A,B,C>}, B^{LU} = {<A,B,C>}, C^{LU} = {<A,B,C>}, D^{LU} = {<D,E>}, E^{LU} = {<D,E>}
$t_s = 12$	**Empty**	A^{LU} = **{<A,B,C>}**, B^{LU} = **{<A,B,C>}**, C^{LU} = **{<A,B,C>}**, D^{LU} = **{<D,E>}**, F^{LU} = **{<F>}**

Figure 8 Execution Trace of Proposed Memo-AGD Algorithm.

C Asymptotic Time Complexity Analysis

Given N gaps, both algorithms first perform sorting operations in $O(Nlog(N))$ time and perform k operations for k subsets of coalesced gaps. The time complexity of comparison operations for each algorithms are as follows:

AGD Algorithm. Given k gaps that intersect within all the gaps in the Observed List, it is necessary to check the total number of subsets (i.e., $\binom{k}{1} + \binom{k}{2} + \binom{k}{3} + \ldots + \binom{k}{k} = \sum_{i=1}^{k} \binom{k}{i}$). When a new $(k+1)$-th gap is added, it is necessary to check whether it intersects with all the existing subsets. Since checking with the subsets of size-i gap costs i, the cost of each intersection is $\sum_{i=1}^{k} i \times \binom{k}{i} = k \times 2^{k-1}$. Therefore, $|N|$ gaps in the dataset, the total cost for adding is $\sum_{k=1}^{|N|} k \times 2^{k-1} = O(N \times 2^{|N|})$. The best case is when all the gaps are disjoint i.e., $O(N^2)$. The worst case is rare because of the sparsity in real datasets.

Memo-AGD Algorithm. The worst case of Memo-AGD will be similar to *AGD* except we only check maximal sets instead of $\binom{k}{i}$ gaps. Hence, for each N and k, we update their resultant shape and elements in the lookup table (i.e., $O(kN)$) and the cost of updating the lookup table. In contrast, the cost will be similar to best case if no gaps are intersecting.