Container Throughput Estimation Leveraging Ship GPS Traces and Open Data

Longbiao Chen[†], Daqing Zhang[‡], Gang Pan*[†,♮], Leye Wang[‡], Xiaojuan Ma[♭], Chao Chen[‡], Shijian Li[†]

[†] Department of Computer Science, Zhejiang University, China

{longbiaochen, gpan, shijianli}@zju.edu.cn

[‡] Institut Mines-TELECOM/TELECOM SudParis, France

{daqing.zhang, leye.wang, chao.chen}@it-sudparis.eu

[♭] Huawei Noah's Ark Lab, Hong Kong, China

xiaojuan.c.ma@gmail.com

[♮] Cyber Innovation Joint Research Center, Zhejiang University, China

ABSTRACT

Traditionally, the port container throughput, a crucial measurement of regional economic development, was manually collected by port authorities. This requires a large amount of human effort and often delays publication of this important figure. In this paper, by leveraging ubiquitous positioning techniques and open data, we propose a two-phase approach to estimation of port container throughput in real-time. First, we obtain the number of container ships arriving at berth by analyzing the ships' GPS traces. Then we estimate the throughput of each ship, in terms of number of containers transshipped, by considering the ship's berthing time, capacity, length, breadth, and crane operation performance, as extracted from different data sources. Evaluation results using real-world datasets from Hong Kong and Singapore show that the proposed approach not only estimates the container throughput quite accurately, but also outperforms the baseline method significantly.

Author Keywords

AIS trace; Open Data; Container throughput estimation

ACM Classification Keywords

H.2.8 Database applications: Data mining.

INTRODUCTION

Container throughput is a crucial measure of port performance and regional economic development, affecting various economic activities such as port planning, liner service optimization, regional policy making, etc. [8,15,21]. Traditional methods of collecting the port container throughput data have centered on manual collection by port authorities, which involves huge human effort and often delays publishing this important figure. Researchers designing models for port logistics and optimization have to estimate container throughput for their analysis [5]. As prior attempts to estimate container throughput were mostly based on historical throughput statistics [16] (which sometimes conflict with each other [6]), these methods had

the potential to introduce risks when the models were implemented into real systems [19].

Therefore, *automated*, *inexpensive* and *real-time* estimation of *port throughput* with high accuracy is in demand. There are two key issues: *estimating the number of ships* arriving at berth for transshipment [22], and *estimating the container throughput of each ship*. Fortunately, with the pervasiveness of AIS (Automatic Identification System [17]) and open maritime data [11,12,20], the dynamic GPS traces of container ships in port as well as information about ships and ports are now available, providing us with new opportunities to address both of these challenges.

In this paper, we attempt to leverage heterogeneous maritime data sources to estimate container throughput. First, we detect *when* a ship arrives at berth and *how long* it stays for transshipment using AIS traces, which contain information about the ship's ID, GPS position, speed, heading direction, etc. However, this is not trivial since a ship may behave differently in a port, such as staying in the anchorage, mooring at berth, etc. These events share similar spatial-temporal patterns and need to be identified accurately. Next, we estimate *how many* containers were handled during each *berthing event*, taking into account openly available information about the corresponding ship and port facilities. Again, it is not trivial to infer the actual container throughput from this information; for instance, although the ship's capacity indicates the maximum throughput of a ship, we do not know how much of the ship's capacity is loaded and unloaded during a specific berthing event.

In summary, the main contributions of this paper include:

1. To the best of our knowledge, this is the first work on *container throughput estimation* leveraging both *dynamic AIS traces* and *open maritime data*.

2. We propose a two-phase approach to estimate container throughput by leveraging heterogeneous maritime data sources. The first phase is *berthing event identification*, in which we consider various features from ship AIS traces (including GPS position, speed, heading direction,
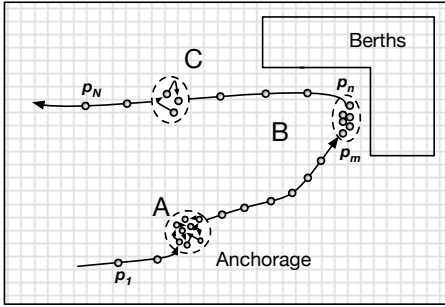
Figure 1. A typical AIS trace of a container ship in port.

etc.) to identify *berthing events*. The second phase is *container throughput estimation*, where we exploit information about ships (e.g., ship capacity, length, and breadth) and port facilities (e.g., container handling performance) from open data sources to estimate container throughput.

3. We evaluate our approach using real-world data collected in two ports – Hong Kong and Singapore. The results show that estimations by our proposed method only differ from the port statistics of Hong Kong and Singapore by 1.0% and 0.4%, respectively. It outperforms the baseline methods significantly.

RELATED WORK

Prior research on port management has led to the development of different container throughput estimation methods. Some methods used historical data to forecast future demand [16], while others leverage macroeconomic conditions and regional competition to model port throughput growth [18]. However, these methods are usually optimized for specific ports and their results sometimes conflict with each other [6].

In the field of event and activity detection using GPS traces, a lot of work has been done with respect to human mobility and taxi operations. For instance, Amini et al. predicted users' destinations from trip GPS trajectories [2], Zheng et al. extracted local attractions from human GPS trajectories [24], and Zhang et al. detected anomalous passenger delivery trips from taxi GPS traces [23]. As for the ship events detection, AIS traces have only been applied to areas such as capturing ship anomaly [17] and assess air pollution in port [13]. There is no work reported for berthing event identification.

Using open data available over the Internet is a common practice in ubiquitous computing. Examples can be found in fields such as public service [4,10] and public safety [9]. To the best of our knowledge, this work is the first to estimate port container throughput with dynamic AIS traces and open maritime data. The following sections provide a detailed description of the proposed *two-phase* approach.

PHASE I - BERTHING EVENT IDENTIFICATION

Figure 1 shows a typical AIS trace of a container ship in a port, where segments A, B, and C correspond to the ship arriving at anchorage, mooring at berth, and stopping at sea, respectively. In order to estimate the number of ships at berth, the research objective in this phase is to *define the berthing event* and develop a general-purpose approach for *identifying berthing events* from ship AIS traces.

Representation of a Berthing Event

A ship AIS trace is composed of a sequence of AIS points, where each point p contains a timestamp (t), the ship's GPS location (lat and lng), speed (v), and heading direction (h), i.e., $p = (t,lat,lng,v,h)$. Hence, a ship AIS trace S can be represented as $S = p_1 \rightarrow p_2 \rightarrow ... \rightarrow p_N$.

Given the above AIS traces, we define a berthing event as a triplet, i.e.,

$$b = (id,s,d)$$

where id is the unique ship identification number, $s = p_m \rightarrow p_{m+1} \rightarrow ... \rightarrow p_n$ refers to the trace segment corresponding to the berthing event, and d denotes the berthing time.

According to port operation conventions [1], when ships are transshipping containers at berth, in principle they are:

1. *stationary* and *stable*, i.e., with speed close to zero, GPS location confined to a fixed point, and heading towards a fixed direction.

2. *densely* located in certain geographical regions.

The above two characteristics of berthing events are used as detecting criteria in this study. Based on these two criteria, we propose a *two-step process* for identifying each berthing event. First, we detect all candidate berthing events (*CBEs*) from a ship's trace that meet criterion 1. Then, we remove noise from the *CBE* set by applying criterion 2.

Adaptive Sliding Window-Based CBE Detection

We utilize an adaptive sliding window-based algorithm to detect all *CBEs* in a ship's trace. Specifically, we start by checking whether a trace segment containing two points (i.e., window size equal to two) meets criterion 1. If not, the algorithm discards the segment and slides to the next one. Otherwise, the algorithm adds the adjacent point to the segment and checks if the newly formed segment still meets the criterion. The window size keeps increasing until the ship starts to move, where the *speed* or *heading* constraint is violated. In this way, the algorithm detects all *CBEs* in a ship's trace. Finally, we apply this method for all ships' AIS traces in a port and obtain a set of *CBEs*.

Clustering Based Noise Removal

The *CBE* set obtained in the previous step may contain non-berthing events, such as temporary stops at sea (segment C in Figure 1). Unlike berthing events that occur densely near terminals, these non-berthing events are distributed more

randomly. Therefore, we adopt the popular density-based clustering method DBSCAN [7] to remove noise.

After the above two steps, the output of this phase is a set of berthing events of all container ships in a port over a period of time, i.e.,

$$B = \{b \mid b = (id, s, d)\}$$

PHASE II - CONTAINER THROUGHPUT ESTIMATION

In this phase, we need to estimate the container throughput for each berthing event identified in Phase I. Based on existing literature [21,22], containers are transshipped by quay cranes (QCs), and the throughput of the transshipment is mainly affected by three factors: 1) the quantity of QCs working on a ship, 2) the performance of a single QC, and 3) the transshipment time. We also observe that:

- The quantity of QCs working on a ship is related to the ship's length. The longer a ship, the more QCs can be assigned to work on it simultaneously.

- The performance of a single QC is related to the distance to transship a container from ship to shore (and *vice versa*). The wider the ship, the farther a container has to travel.

- The transshipment time can be obtained by subtracting unproductive operation time from berthing time.

Based on these observations, we propose a *QC-performance-based* method to estimate the container throughput of the port, which leverages ship dimensions (i.e., length and breadth) [12] and port specifications [11,20] to characterize the above factors.

Quay Crane Quantity Estimation
Given a ship of length L, and assuming the average width of each QC is A, the maximal quay crane number N can be approximated by:

$$N = \left\lfloor \frac{L}{A} \right\rfloor$$

Quay Crane Performance Estimation
Quay crane performance can be measured by *the number of containers transshipped per hour* [21]. As illustrated in Figure 2, containers are transshipped by a trolley moving along the boom. The distance traveled by the trolley depends on the location of the container. Assuming the average distance from the QC to the nearest container is D', the container width is D, and the trolley movement speed is V, then the time spent to transship n containers in a row can be derived as follows:

$$t = \frac{2 \sum_{i=0}^{n-1} (D' + i \bullet D)}{V} = \frac{2nD' + n(n-1)D}{V}$$
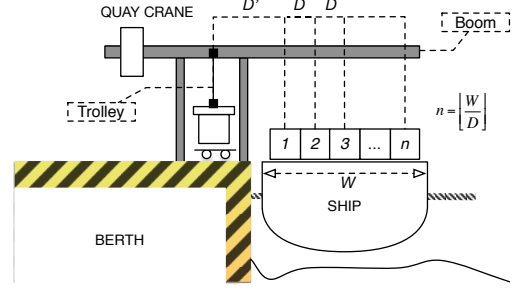

Figure 2. Container transshipment by quay crane at berth.

Thus the quay crane performance Q can be expressed as:

$$Q = \frac{n}{t} = \frac{V}{2D' + (\left\lfloor \frac{W}{D} \right\rfloor - 1)D}$$

Transshipment Time Estimation
To obtain the actual transshipment time, the overhead time ΔT caused by nonproductive operation has to be subtracted from the total berthing time T. There are two main kinds of unproductive operation:

- Before and after transshipment, the booms of QCs have to drop down and lift up, respectively [21].

- During transshipment, QCs have to move horizontally between container stacks (bays) [22].

Total Container Throughput Estimation
For each identified berthing event, given the quantity of QC, the single QC performance, and the transshipment time, we estimate its throughput and sum them up to obtain the total throughput for a port in a given period of time, i.e.,

$$P = \sum P_i = \sum \min(N_i \bullet Q_i \bullet (T_i - \Delta T), 2C_i)$$

$$= \sum \min(\frac{V \left\lfloor \frac{L_i}{A} \right\rfloor}{2D' + (\left\lfloor \frac{W_i}{D} \right\rfloor - 1)D} (T_i - \Delta T), 2C_i)$$

Note that the throughput of a berthing event cannot exceed double a ship's capacity ($2C_i$), which corresponds to a fully unload and reload of containers.

EVALUATION
We conducted an empirical evaluation of our proposed approach on real-world datasets. We used: (1) ship AIS traces collected from Hong Kong and Singapore in 2011, (2) container ship information (e.g., ship capacity, length,
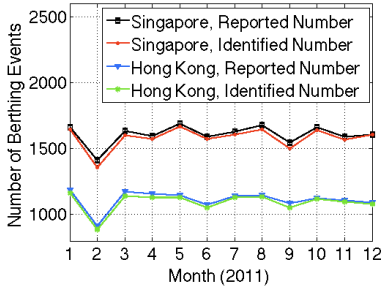
Figure 3. Results of berthing event identification with Port of Hong Kong and Port of Singapore
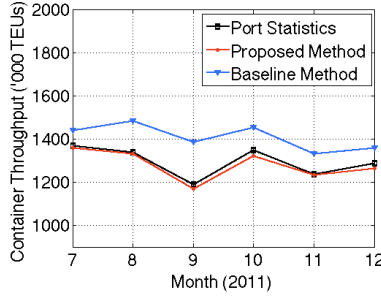
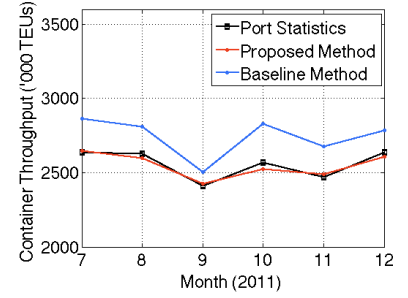**Figure 4. Results of container throughput estimation with Port of Hong Kong.**

Figure 5. Results of container throughput estimation with Port of Singapore.

breadth) obtained from Marine Traffic [12], (3) port facility specifications of Hong Kong [11] and Singapore [20], and (4) ship arrival and container throughput statistics of the two ports [11,20] in 2011. All the datasets were used in our proposed container throughput estimation method, while the fourth dataset also served as ground truth to validate our estimation results.

Evaluation of Berthing Event Identification
Figure 3 shows the number of berthing events identified by our method in the two ports in each month of 2011. Compared to the port authority statistics, the Mean Absolute Percentage Errors [3] of the proposed method for Hong Kong and Singapore are 1.8% and 1.7%, respectively.

Evaluation of Container Throughput Estimation

Parameter Settings
Before conducting the evaluation, we need to determine the four parameters (i.e., *V, A, D', ΔT)* in our proposed method[1]. First, we find the range for each parameter from literature [8,21,22] and port facility specifications [11,20]. Then, we use an expectation-maximization algorithm [14] to select optimal parameter settings for each port, leveraging throughput data from the first six months of 2011. The parameter ranges and selected parameter settings for both ports are shown in Table 1.

	Meaning	**Range**	**HK**	**SG**
V(m/min)	QC trolly speed	80 – 120	83	101
A (m)	QC width	25 – 35	30	32
D' (m)	QC-container dist.	10 – 20	15	17
ΔT (min)	nonproductive time	10 – 20	18	13

Table 1: Parameter ranges and settings for port of Hong Kong (HK) and Singapore (SG)

Baseline Method
For comparison purposes, we take the *berthing-time-based* throughput estimation method as the baseline, which assumes that 1) a ship's throughput for each transshipment

---

[1] The container width *D=2.44m* is the same for all containers.

is proportional to its berthing time, and 2) a ship spends the longest berthing time when it unloads and reloads its full capacity of containers, which is double the ship's capacity (*2C_i*).

Based on the above assumptions, for a specific ship, we first find its maximal berthing time (T_{max}) by searching all its historical berthing events. Then, for each berthing event of that ship, we estimate the throughput p_i as follows:

$$p_i = \frac{T_i}{T_{max}} \bullet 2C_i$$

where T_i is the berthing time. Finally, we sum up the throughput of all berthing events identified in a port over a period of time to obtain the overall throughput.

Evaluation Results
Figure 4 and Figure 5 show the container throughput estimation results using the two methods for Port of Hong Kong and Singapore, respectively, in the last six months of 2011. It can be seen that the *QC-performance-based* method achieves a Mean Absolute Percentage Error of 1.0% with Hong Kong and 0.4% with Singapore, respectively. In contrast, the baseline method produces a Mean Absolute Percentage Error of 8.2% with Hong Kong and 7.1% with Singapore, respectively. Apparently, our proposed method estimates the container throughput of both ports quite accurately, and it outperforms the baseline approach significantly.

CONCLUSION
In this paper, we have proposed a two-phase approach for real-time estimation of container throughput using dynamic ship AIS traces and open maritime data. We first obtain the number of container ships by applying an *adaptive sliding window-based* method to the ship AIS traces. We then propose a *QC-performance-based* method to estimate the container throughput of each ship leveraging open maritime data. Evaluations using real-world datasets from Hong Kong and Singapore show that our proposed approach not only estimates the container throughput quite accurately, but also outperforms the baseline approach significantly.

REFERENCES

1. Alderton, P. *Port Management and Operations*. Taylor & Francis, 2013.

2. Amini, S., Brush, A.J., Krumm, J., Teevan, J., and Karlson, A. Trajectory-aware mobile search. *In Proc. CHI 2012*, ACM Press (2012), 2561–2564.

3. Armstrong, J.S. and Collopy, F. Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting 8*, 1 (1992), 69–80.

4. Bichard, J.-A. and Knight, G. Improving public services through open data: public toilets. *Proceedings of the ICE - Municipal Engineer 165*, 3 (2012), 157–165.

5. Bierwirth, C. and Meisel, F. A survey of berth allocation and quay crane scheduling problems in container terminals. *European Journal of Operational Research 202*, 3 (2010), 615–627.

6. Chen, S. and Chen, J. Forecasting container throughputs at ports using genetic programming. *Expert Systems with Applications 37*, 3 (2010), 2054–2058.

7. Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. *In Proc. KDD 1996*, (1996), 226–231.

8. Fourgeaud, P. *Measuring port performance*. The World Bank, 2000.

9. Garbett, A., Linehan, C., Kirman, B., Wardman, J., and Lawson, S. How dangerous is your life? Personalising Government open crime data. *In Proc. CHI 2012 Adjunct*, ACM Press (2012).

10. Groen, M., Meys, W., and Veenstra, M. Creating Smart Information Services for Tourists by Means of Dynamic Open Data. *In Proc. UbiComp 2013 Adjunct*, ACM (2013), 1329–1330.

11. Hong Kong Marine Department. Port and Maritime Statistics of Hong Kong. http://mardep.gov.hk/en/publication/portstat.html.

12. Marine Traffic. Ship Tracks API. https://www.marinetraffic.com/en/p/api-services.

13. Miola, A. and Ciuffo, B. Estimating air emissions from ships: Meta-analysis of modelling approaches and available data sources. *Atmospheric Environment 45*, 13 (2011), 2242–2251.

14. Moon, T.K. The expectation-maximization algorithm. *Signal processing magazine 13*, 6 (1996), 47–60.

15. Notteboom, T.E. Container Shipping And Ports: An Overview. *Review of Network Economics 3*, 2 (2004).

16. Peng, W.-Y. and Chu, C.-W. A comparison of univariate methods for forecasting container throughput volumes. *Mathematical and Computer Modelling 50*, 7–8 (2009), 1045–1057.

17. Ristic, B., La Scala, B., Morelande, M., and Gordon, N. Statistical analysis of motion patterns in AIS data: Anomaly detection and motion prediction. *In Proc. FUSION 2008*, IEEE (2008), 1–7.

18. Seabrooke, W., Hui, E.C.M., Lam, W.H.K., and Wong, G.K.C. Forecasting cargo growth and regional role of the port of Hong Kong. *Cities 20*, 1 (2003), 51–64.

19. Shabayek, A.A. and Yeung, W.W. A simulation model for the Kwai Chung container terminals in Hong Kong. *European Journal of Operational Research 140*, 1 (2002), 1–11.

20. Singapore Maritime and Port Authority. Port Statistics of Singapore. http://www.mpa.gov.sg/sites/global_navigation/publications/publications.page.

21. Steenken, D., Voß, S., and Stahlbock, R. Container terminal operation and operations research - a classification and literature review. *OR Spectrum 26*, 1 (2004), 3–49.

22. Vis, I.F.A. and de Koster, R. Transshipment of containers at a container terminal: An overview. *European Journal of Operational Research 147*, 1 (2003), 1–16.

23. Zhang, D., Li, N., Zhou, Z.-H., Chen, C., Sun, L., and Li, S. iBAT: detecting anomalous taxi trajectories from GPS traces. *In Proc. UbiComp 2011*, ACM Press (2011), 99–108.

24. Zheng, Y., Zhang, L., Xie, X., and Ma, W.-Y. Mining Interesting Locations and Travel Sequences from GPS Trajectories. *In Proc. WWW 2009*, ACM Press (2009), 791–800.