

Big Data Acquisition and Analysis Platform for Intermodal Transport

Kai Xu¹, Hong Zhen¹, Yan Li² and Luo Yue³

¹(Shanghai International Shipping Institute, Shanghai Maritime University, Shanghai, PR China)

²(Department of Computer Science and Technology, Tongji University, Shanghai, PR China)

³(Shanghai Yiwei Information Technology Co., Ltd., Shanghai, PR China)
kernxu@gmail.com

Abstract

This paper aims for the transparency and visualization of the international intermodal cargo transportation throughout its whole process, achieving a comprehensive monitoring on the multiple transportation means such as by ocean, by air, by land or by rail. Based on Internet-of-Things-based distributed data acquisition technology and the cloud-computing-based big data analysis technology, this paper gives out a Multimodal Monitoring technology that can uniformly solve the comprehensive management of multiple transportation vehicles, which includes a service functionality model, a network hierarchy model and a technology system model. By building a Generic Target Monitoring System, it proves the multimodal monitoring resolution is able to effectively monitor the multiple transportation means and provide a fair good database platform of later analysis, distribution and optimization of those vehicles.

Keywords: Internet of Things; Big Data; Distributed

1. Introduction

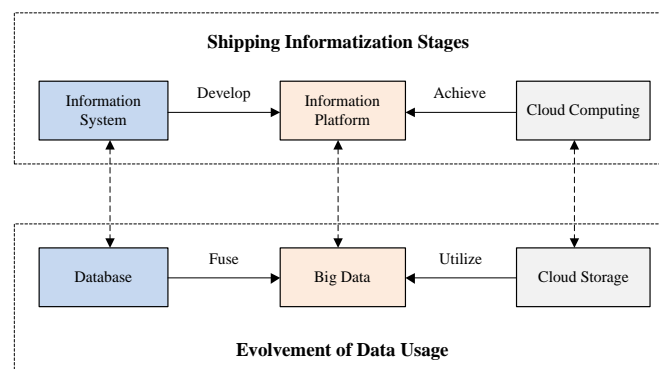


Figure 1. Relationship between Shipping Informatization Stages and Evolvement of Data Usage

The international cargo transportation commonly involves multiple transportation means *e.g.* by sea, by water, by air, by land or by train, *etc.* Such transportation was named as Multimodal Transport, which involves at least two different types of transport from one country to another. The traditional transportation industry information system were mainly designed against the carriers' business demand, which results in isolation of the different businesses, varieties of storage formats and the de-connection between the water, land and air monitoring, dispatching and management systems. Thus, it scarcely

meet the management requirement by the multimodal transport. There is a prophecy in Prospect of China Shipping Development in 2030 that the shipping will be fully transparent throughout the whole procedure; the Information barriers among the multimodal transport will be broken; such relevant technologies will be of great value of research and of importance of application demands [1].

In the big data era, we can rebuild a similar information platform for the multimodal monitoring as presented in Figure I on a mixed viewpoint with the development of the data fusion technology and information platform [2]. Although the multiple transport ways have their own business characters, their business demands are still similar so that an integrated data acquisition and analyzing platform can be built on the data fusion of the multiple transport. The new full-time visualized multimodal transport information platform integrating the monitoring information of the transport by water, by land, by air and by train with the core of oceanic shipping monitoring will be of great significance on improving the management, optimizing transport efficiency and securing the transportation.

2. Relevant Applications

As the intermodal monitoring is a fairly new concept [3], and it involves multiple monitoring on the different transport and together with heterogeneous communication network, there is few direct research on it. However, more researches have done on special transport fields, *e.g.* oceanic and land transportation, in and abroad of China. And some researches have already kicked off on multiple transport monitoring.

2.1. Shipping Monitoring Applications

Oceanic shipping, occupying more than 90% of the multimodal transport, is a section of great significance which has carried on some researches already. The Shared Intermodal Container Information System, *i.e.* SICIS, which was made by the Institute of Shipping Economics and Logistics, ISL in Bremen, Germany, identifies the RFID electrical label at key logistic nodes to track the status of the transported goods. Similar solution was made by Frank Arendt *et al.* (2012) in the research project of INTHGRITY, which provided the carriers, terminals and the customs with visualized container shipping scenes by tapping on the RFID tech [3]. Mr. Bao Qifan *et al.* with Shanghai International Port (Group) Co., Ltd researched on the technology of tracking the cargo shipping through the RFID labels in the delivery points, entry of container yard and the handling nodes [4]. This technology has been first applied to Shanghai-Savannah Sino-American Container Shipping Line [5].

2.2. Other Transportation Monitoring Applications

On other transportation monitoring applications, good practices have been applied mainly to land transports. Vehicle monitoring technologies are there based on automatic identifying technology or vehicle mounted monitoring devices.

Based on automatic identifying technology, cameras and RFID card readers are positioned on the road crosses, entries and monitoring points to monitor the vehicles by the mode identifying technology, *e.g.* vehicle identifying and plates identifying, and RFID technology. Gan Lin *et al.* (2003) adopted Kalman prediction model regional matching algorithm to analyze camera data to achieve tracking the moving vehicles with fair efficiency [6]. Thanks to free of vehicle mounting devices, this technology can be easily and widely applied by the transportation authority while large amount of monitoring blind points are still there.

Based on vehicle mounting technology, GPS positioning devices, Beidou Positioning Module and together with wireless network technologies are mounted on vehicles to achieve real-time monitoring of the vehicles and cargo transport [7]. Wang Baoying *et al.*

(2007) applied ZigBee technology to build a wireless data acquisition network for the intelligent transportation which is of low power consumption, flexible structure, accessibility of extension and connection [8]. Yang Xin *et al.* (2014) researched on the routing strategy of the distributed position service under vehicle mounting Ad Hoc wireless network environment [9]. Zhao Xin *et al.* (2014) researched and made the Vehicle Management System for Hazardous Goods Transportation which needs install an on-vehicle terminal gear that communicating with the servers by wireless data transferring technology to timely report the opening of the carriage, smokes sensors, driver identification and *etc.* so that securing the supervision on the hazardous goods transportation [10]; Bao Yaodong *et al.* (2015) researched a kind of intelligent monitoring device for the environmental transportation of the mud lorries based on the Internet of Things which combines the on-vehicle terminal and the wireless monitoring devices on key logistics nodes [11]. These technologies cost more on the vehicle refurbishment and are uneasy to be promoted. Thus it is often applied in special shipping sectors with intense monitoring demands.

2.3. Monitoring Applications on Comprehensive Transportations

The pre-stage of the demand of researches on comprehensive transport monitoring derives from logistics and supply chain management. Therefore, quite a few researches on logistics comprehensive information platform have raised the issue of transportation monitoring and transparency, but they did not provide a solution for monitoring integration. Giunipero *et al.* (1996) considered the frequency and quality between every supply chain members represented the transparency of the supply chain itself, which greatly influences the success of supply chain management [12]. Mats Abrahamsson *et al.* (2003) defined and described the logistics information platform and considered it as a key part of the entire logistics information system [13].

There have already been several researches that trying to build a comprehensive transportation monitoring platform on basis of those on the transportation by sea, by land and by train. Zhang Rong *et al.* (2006) once analyzed the multimodal transportation and its information system procedure of Yangshan Deep-Water Port, put forth the concept of multimodal transport central information platform and analyzed the business procedure and functionality structure of the said platform [14].

2.4. Distributed Acquisition and Analysis

The data of multimodal monitoring are mainly time-series data. Owing to the mass quantity, liquidity and wide-spread of the monitored objects, the monitoring data acquisition and storage should adopt distributional methodology. Ji Xianbiao *et al.* (2007) researched on key technology of dynamic analysis and conversion of information frame for AIS information distributed acquisition system for ships [15]. Zhang Zhenping *et al.* (2013) raised an AIS database structure model basing on Key-Value type database for overcoming the obstacles of massive AIS data storage [16]. They have referred to the distributed acquisition and storage of the data, but the researches on monitoring data distributive analysis and calculation framework are still in vacancy.

2.5. Remaining Unsettled Problems of Existing Applications

The abovementioned monitoring technologies of the goods transportation still have some defects and short-comings. First, they lack full-time monitoring. Most of them can only monitor the goods transportation on key nodes. Second, the monitoring information is confined to the positioning and geographical information of the goods or vehicles without the physical status sensing during the transporting procedure. Third, any types of the above monitoring systems of the transportation means are independent with each

other, so that they lack a uniform data transferring and storage format to achieve the data sharing and integration.

3. The Proposal of Multimodal Monitoring Technology

A comprehensive monitoring technology that enable to uniformly tackle the multiple transportation problems against their different characteristics of land, water, air and rail are of great value on research and application. The Multimodal Monitoring proposed by this paper is mainly to meet the demands during the multimodal transport as follows:

3.1. Distributed Data Acquisition Based on Internet of Things

The monitoring of multimodal transport should not only be confined to key logistics nodes but also need to cover the full-time monitoring, thus we apply the technology of Internet of Things to achieve full-time and real time data acquisition throughout the different means of transportation such as shipping, water, air, land and rail transportation. This demands all vehicles accessing to internet with their IP address decided by their unique MAC addresses, apply a same communication protocol, automatically acquire data distributed by wireless network or self-organized network and then collect all data to the multimodal monitoring platform.

3.2. Big Data Processing Environment Based on Cloud Computing

The multimodal monitoring technology not only needs information on the longitude and latitude coordinates, moving directions, velocities, but also need to monitor the working power, gas emission, destination, details of the loaded cargos and system malfunctions, in a word, the transportation status information. These kinds of information demands big data technology to support its normal functioning that can integrate themselves stemming from the ships, automobiles, trains and planes owing to its diverse origins, mass of quantity, strict real timing or different storage formats. Such big data processing environment shall be able to solve the key problems on distributed acquisition, storage and indexing as well as those on granularity, format, calibre, distributed computing, real time computing, data mining.

3.3. Visualization Analysis Based on GIS

The applications of multimodal monitoring technology requires huge amount of graphics with a fundamental function of GIS display based on the geographic information system, GIS. First, it shall be able to support real time visualization display of current transportation status. Second, with mass historical data, it shall be able to replay the historical traces of the single vehicle and that of the multiple vehicles in designated geographical areas. Third, it will need to present to the audience with visualized data analysis and prediction results.

4. Building Generic Target Monitoring System, GTMS

Generic Target Monitoring System, GTMS, is designed for the resolution of the requirement of multimodal monitoring based on Internet of Things and big data technology. It is a sum-up concept [1] of technology solution package and information platform that achieves the all-rounded monitoring of the position, information, status and malfunction for all moving targets such as the automobiles, ships, planes, trains and humans by applying Internet of Things, and then integrates, stores, processes, searches and analyses relevant big data to simultaneously display the current and historical information of the monitored objects on a same electronic map.

4.1. Service Functionality Model of GTMS

Figure 2 display the service functionality model of GTMS system. All types of the monitored targets will first be registered in Registration Centre and get a Target ID. The data acquisition devices will then submit the data with its own ID to the storage nodes on distributed database. Based on the same distributed database, the data will be processed on the data analysis platform. In the end, the result of the analysis and search will be displayed to the platform user in graphical form.

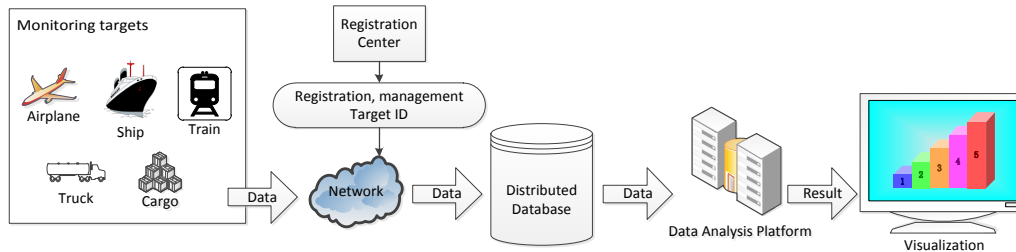


Figure 2. Service Functionality Model of GTMS

4.2. Network Layers Model of GTMS

The GTMS can be divided into six layers. They are sensor layer, transfer layer, storage layer, service layer, application layer and management layer.

Table 1. Description of Six Layers of GTMS

Layer	Description
Sensor layer	data collection on sensing, identification, positioning and <i>etc.</i>
Transfer layer	regulation of the transferring protocol, realizing the data transferring in heterogeneous network
Storage layer	application of the distributed storage framework to integrate, store, process and backup the data
Service layer	services provision for the exchanging, searching and analyzing of the mass data based on parallel computing technology to support the application layer
Application layer	realization of visualized display and traces replay based on GIS to support multiple requirements by different applications
Management layer	terminal registration management, service management, authority safety management, storage distribution management and <i>etc.</i> provided by the registration center

4.3. Technology System Model of GTMS

GTMS system uses a uniform data format - Multi-target Packet (MTP), to achieve the monitoring of automobiles, ships, planes and trains in a whole and to reach a uniformed representation and transferring of the monitoring data on different transportation means.

4.3.1 Intelligent Distributed Acquisition Terminal

The data acquisition for multiple monitored objects is achieved by intelligent distributed acquisition terminals. The hardware structure is shown as in Figure 3 which takes low power consumption intelligent controller as its core and collects the data from its own data bus, satellite positioning module and other sensors. At last it transfers the data to deployed server through communication module. In the Following Figure 4, the vehicle assembles Beidou System as its satellite positioning service, Beidou short message services as the communication module, OBD bus for vehicle status information, and sensors for data acquisition on exhaust pipe as the intelligent distributed acquisition terminal.

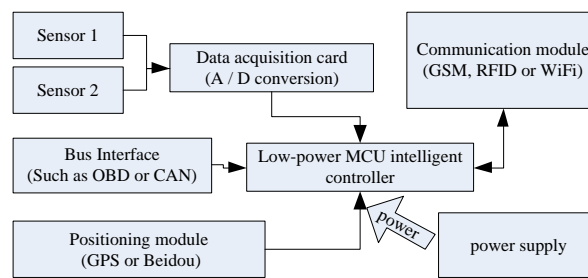


Figure 3. Basic Structure of an Intelligent Distributed Acquisition Terminal

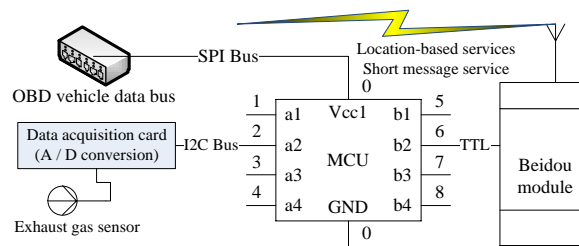


Figure 4. An Assembly Example of Vehicle Mounting Intelligent Distributed Acquisition Terminal

4.3.2. Uniform Data Formats

Multi-target Packet, MTP, is a type of conversation level network protocol data package for GTMS with a default transferring protocol of UDP/IP, TCP/IP or others so that it can adapt to different heterogeneous network environments. There are some basic features of MTP protocol. (1) MTP has two types of packets which are dynamic and static separately. (2) MTP packets support to monitor multiple objects with good customization and extension possibilities. Some of them even can be directly read and can code the MAC address for all monitored objects. (3) MTP not only can show the positioning and status information for the monitored objects, but also can show the moving status, cargo details, malfunction information, energy consumption and *etc.* for the same objects. MTP packet is usually composed of HEAD, DATA and CHECK. (refer to Table 2) in the HEAD, the field of target type is special with a volume of 6 bit, valuing span of 0-63 and 64 types. The current defined types, they are: 0 for ship, 1 for automobile, 2 for crane, 3 for plane and 4 for train. The DATA part in different types of MTP packet shall at least contain public field and extend field. The public fields record their common attributes for different types of the monitored objects, *e.g.* the name, coordinate, velocity and *etc.* The content of the extended field will have different meanings according to the target *_type* field *e.g.* the first field of the extend field of ships' dynamic packet means steering ratio while the same field for the automobiles means RMP (refer to Table 3).

Table 2. MTP Dynamic Packet Structure

Part	Field	Field name	Size
HEAD (37Byte)	Protocol Version	protocol_type	1Byte
	MAC address	mac	8Byte
	Counter	counter	2Byte
	Data Sources	data_source	1Byte
	Packet Type	message_type	2bit
	Data Types	target_type	6bit
	Uniquely identifies	unique_id	10Byte
	Acquisition time	acquisition_time	8Byte
	Extend	extend	4Byte
	Packet length	package_volume	2Byte
DATA (87Byte)	Public data fields	public_field	32Byte
	Extended data field	extend_field	28Byte
	Text Field	text_field	27Byte
CHECK	Check Digit	checksum	4Byte

Table 3. Definition Comparison of the Extension Data Field in MTP Dynamic Packet

<i>Vehicles</i>	<i>Ship</i>	<i>Automobile</i>	<i>Crane</i>
<i>Data type</i>	<i>target_type=0</i>	<i>target_type=1</i>	<i>target_type=2</i>
<i>Field 1</i>	<i>ROT</i>	<i>engine speed</i>	<i>main crane capacity</i>
<i>Field 2</i>	<i>positioning accuracy</i>	<i>coolant temperature</i>	<i>auxiliary crane capacity</i>
<i>Field 3</i>	<i>main engine power</i>	<i>inlet temperature</i>	<i>ascending altitude</i>
<i>Field 4</i>	<i>auxiliary engine power</i>	<i>engine load</i>	<i>main hook ascending speed</i>
<i>Field 5</i>	<i>oil level</i>	<i>remaining oil</i>	<i>gantry speed</i>
<i>Field 6</i>	<i>(reserved)</i>	<i>fuel pressure estimation</i>	<i>crown block speed</i>
<i>Field 7</i>	<i>(reserved)</i>	<i>(reserved)</i>	<i>(reserved)</i>

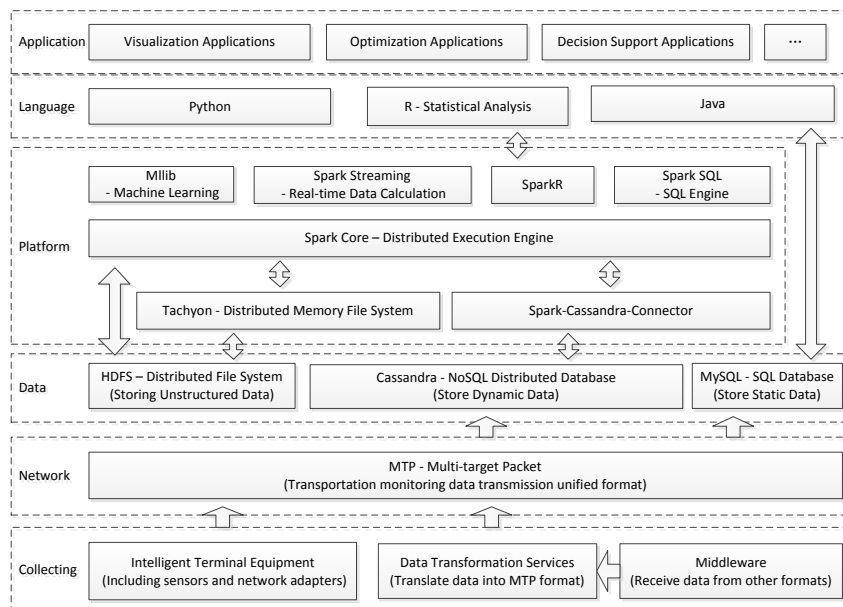


Figure 5. Big Data Analysis Framework Based on Cassandra+Spark Technology

4.3.3. Big Data Analysis Framework

After collecting all data through MTP packet, the system will store the data into the database. The static data will be saved into the MySQL database. At the same time, the dynamic packet updates frequently. The fastest updated dynamic packet will be 2 seconds after previous one. Thus, all the ships over the globe will produce tens of millions of dynamic packets per day. Therefore, the dynamic packets requires distributed database for storage and this paper recommend NoSQL Cassandra which better performs on availability and partition tolerance to achieve the distributed storage of dynamic data. The complete big data analysis framework of GTMS system are showed in Figure 5 with Spark technology primarily. Spark uses memory distribution data set, which optimize the iterative working load as well as providing interactive search service. When the task is complicated enough with numbers of times of iteration, Spark is faster than the MapReduce of the Hadoop technology for one to two orders of magnitude and it excels on real time data analysis. Meanwhile, we can use HDFS for the storage of the non-structural database in distributed file systems and introduce Tachyon memory distributive file system so that Spark can focus on memory computing to improve the memory utilization efficiency. Moreover, the analysis framework provides a machine learning standard library by using MLlib, supports real time data processing by using Spark Streaming and supports structural data search by using Spark SQL. In addition to supporting Python and Java structures by Spark, the SparkR adapter also supports R language analysis process and provides diverse language support for multimodal data analysis.

5. Experiments and Verification

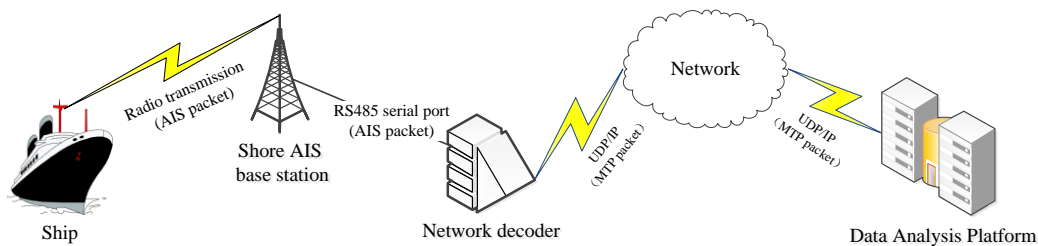


Figure 6. Experimental Prototype System Structure of GTMS

The International Maritime Organization's International Convention for the Safety of Life at Sea requires AIS to be fitted aboard international voyaging ships with gross tonnage (GT) of 300 or more, and to be fitted aboard domestic voyaging ships of the member state with gross tonnage (GT) of more than 500. AIS will broadcast the current driving status of the ship up to a radius of 20 nautical miles for every other second. We can easily collect these AIS broadcastings throughout base station on the shore and receive the AIS signals offshore by low earth orbit mobile communication satellite. Therefore, we can build a prototype system (refer to Figure 6) by using AIS data to verify the feasibility of GTMS system.

5.1. Build GTMS Prototype System

Our prototype system receives two types of AIS data. One is what we purchased. The other type is the data acquired from the base station we built. The network decoder in Figure6, will decode the AIS message from RS485 serial port, repackage and send the MTP packet to MTP receiver by network. The MTP receiver will save the static packets into MySQL and dynamic packets into Cassandra. We build 3+1 type redundant storage structure comprised of 4 nodes in this experiment.

The distributed calculation framework used by the distributed calculation platform is Spark which abstracts Cassandra into RDD data set to achieve the data reading and writing of distributed calculation. The distributed algorithm is taken by R language and Scala programming.

5.2. Trial Results of the Prototype System

The prototype system whose saved tracks of monitoring objects mounting to 13 billion pieces and data size to 1.5TB has run for one year smoothly, increasing more than 20 million pieces of dynamic data day by day.

On this prototype environment, this paper takes the method of length adaptive sliding window to achieve the automatic identification for the anchoring points during the voyaging tracks. The highlighted points on the voyaging tracks in Figure 7 are the anchoring points of the container ship Tianshenghe.



Figure 7. The Voyaging Tracks and Anchoring Events of the Container Ship Tianshenghe (MMSI: 412658000) on early September, 2014

The said algorithm can also calculate global ships' anchoring points by distributed calculation. Figure 8 is the highlighted result of global ships' anchoring events with different gray scale according to their frequency.

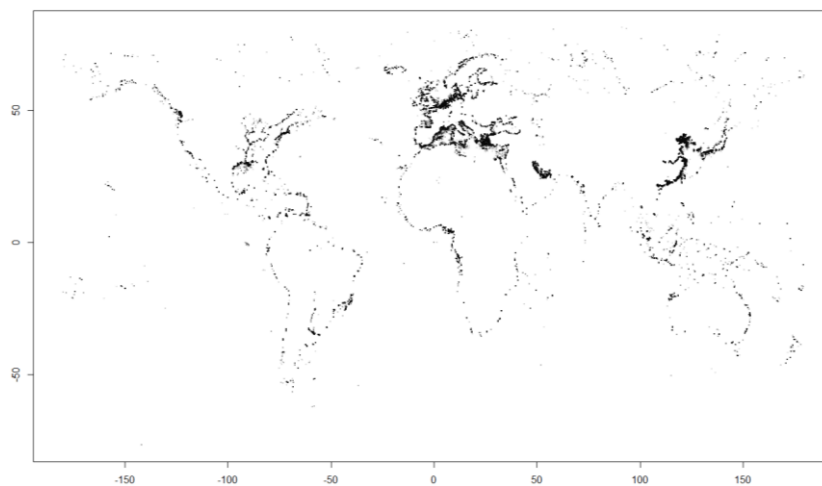


Figure 8. Global Ships' Anchoring Events on Sep. 1st, 2014

6. Conclusion

This paper creates the distributed transport data acquisition and analysis platform GTMS against the demand by multimodal monitoring, proposes a uniform transferring format MTP data package for the moving objects monitoring data, builds the six layers model for GTMS and designs a big data analysis framework based on Cassandra+Spark structure. The GTMS prototype system build in this paper has verified the efficiency of the said technology. It experimented on the oceanic shipping data which is the most seen and also the most difficult-to-monitor to achieve monitoring the global ships. Then it analyzes the ships' anchoring events on big data analysis framework. GTMS technology can achieve monitoring the multiple transport means and provide efficient data platform for analysis, dispatching and optimization of the transportation procedure. It is believed that in the near future, this technology will be of great significance in the multimodal transport fields.

References

- [1] F. Arendt, N. M. Larsen, R. Mueller and A. W. Veenstra, "International journal of shipping and transport logistics", *IJSTL*, vol. 4, no. 2, (2012).
- [2] B. Qifan, "Container RFID information management system", *Chinese Electronic Business*, vol. 5, (2008), pp. 47-48.
- [3] X. Kai and L. Jianli, "Impact of Things technology for port productivity", *Shipping Management*, vol. 33, no. 4, (2011), pp. 28-30.
- [4] G. Ling and P. Xiaolei, "One kind used in the transport environment moving vehicle tracking method", *Journal of Chongqing University of Posts and Telecommunications (Natural Science)*, vol. 25, no. 3, (2013), pp. 408-411.
- [5] B. Jian and Z. Qi, "GIS technology-based distributed mobile LBS applications", *Journal of Chongqing University of Posts and Telecommunications (Natural Science)*, vol. 16, no. 1, (2004), pp. 105-107.
- [6] W. Baoying, C. Xuemei, M. Chunyan and L. Peng, "Intelligent Transportation Network Based on ZigBee Technology", *Journal of Chongqing University of Posts and Telecommunications (Natural Science)*, vol. 19, no. 6, (2007), pp. 1-4.
- [7] Y. Xin, K. Yujun and K. J. John, "Distributed routing policy to a vehicle location service Ad Hoc Networks", *Journal of Chongqing University of Posts and Telecommunications (Natural Science)*, vol. 26, no. 5, (2014), pp. 598-604.
- [8] L. C. Giunipero and R. R. Brand, "Purchasing's Role in Supply Chain Management", *International Journal of Logistics Management*, vol. 7, no. 1, (1996), pp. 29-38.
- [9] M. Abrahamsson, N. Aldin and F. Stahre, "Logistics platforms for improved strategic flexibility", *International Journal of Logistics*, vol. 6, no. 3, (2003), pp. 85-106.
- [10] Z. Rong and L. Zhiqi, "Container Intermodal Terminal Information Platform Framework", *Journal of Tongji University(Natural Science)*, vol. 34, no. 2, (2006), pp. 201-206.
- [11] J. Xianbiao, S. Zheping, P. Jiakai and S. Tengda, "Development and Key Techniques AIS information Distributed Acquisition System", *Journal of Shanghai Maritime University*, vol. 28, no. 1, (2007), pp. 28-31.
- [12] Z. Zhengping, Y. Chun and H. Qinyou, "Automatic Identification System distributed data storage method", *Chinese sailing*, vol. 36, no. 3, (2013), pp. 10-13.
- [13] X. Kai, "Shipping Information Platform in the Age of Big Data", Beijing, China: People's Traffic Press Co., Ltd., (2015).
- [14] "Shanghai International Shipping Institute. China Shipping Development Outlook 2030", Beijing, China: People's Traffic Press Co., Ltd., (2013), pp. 113-114.
- [15] H. Di, "Computer Group Co., Ltd. management system of dangerous goods transport vehicles", Chinese Patent: 2014100084420, (2014).
- [16] "Nantong Polytechnic College. A muck cars environmentally friendly transport means intelligent monitoring based on Internet of Things", Chinese Patent: 201510156072X, (2015).

Authors



Kai Xu, Male, born in 1983, PhD, Engineer, China Computer Society (CCF) members (16854M), director of Shipping & Port Big Data Laboratory, Mainly engaged in the fields of Big Data, Internet of Things, machine learning and these applications in the ports and shipping areas, E-mail: kernxu@gmail.com.

