International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2019 IJSRCSEIT | Volume 5 | Issue 2 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT1952326

A Survey on Dynamic Toll Charge

Prof. Mounica B, Merlin Mathew, Aishwarya Balakrishnan, Bikky Kumar Goit
Department of Information Science and Engineering, New Horizon College of Engineering, Bengaluru, Karnataka, India

ABSTRACT

With continuing economic growth, the demand of traffic rises continuously, especially in modern cities like Bangalore. However, the space available in cities is strictly limited. In order to cope with the challenge of serving a rising traffic demand on these limited capacities, traffic management becomes mandatory. One aspect is managing the traffic by charging dynamic toll depends on conditions like occupancy of cab services, previous data. The focus of this project is a dynamic toll pricing scheme that alleviates congestion and maintains an optimized traffic density during peak hour traffic.

Keywords : Big data eco-system, Intelligent transportation system, NoSQL Database, feature extraction and feature selection.

I. INTRODUCTION

The Silicon city of the country which houses more than 2.2 million IT workers is now popularly known for its traffic. This crawling traffic has caused businesses to slow down and showcases its worst behavior during monsoon. One of the major reasons for this traffic is the star pattern road structure in the city which causes extreme congestion. Another crucial factor is that there is inefficient use of four wheelers where there is a single occupant i.e. the driver. This takes up more space per person on the already narrow roads. Is there a way to ease the traffic through tolls which is an already working system? Can a short term exploitation be found to abate the traffic volume? Static traffic management was one solution looked into to monitor continuous traffic with traffic police deployed at every busy junction, although the commuters were guided by road signs. This became arduous and inefficient as the population of the city began to swell. As one solution for this problem, traffic lights were installed to help manage traffic. This system is still prevalent and automated but the problem arose when the length of light was fixed irrespective of traffic making it ineffective in reducing the traffic. To overcome this, intelligent transportation systems were introduced.

II. BIG DATA ECO-SYSTEM

Big data is a field of engineering that deals with large datasets which follows the ETL method of capturing the data from one data environment, extracts it, transforms and loads it into another data environment in an effective manner. Big data is an essential technology when efficiency and accuracy is of highest concern. It integrates and analyses massive amounts of data to reduce the cost of failure and improve the system. It is also a database for correlating the collected data and the user can access a part of the stored information by querying the nosql databases.

A. Apache Hadoop

Hadoop is an open source java based framework designed to store shared data. Hadoop MapReduce works on master-slave architecture where master

node manages all slave nodes. Hadoop is a combination of HDFS and MR. MapReduce(MR) has a mapper function that creates key-value pairs and a reduce function that first gathers input from the mapper functions and reduces the output based on Shuffle and Sort. HDFS is used to split the enormous data into smaller sections, replicate and manage them on multiple nodes. HDFS streams data to user applications. It contains two types of nodes- JobTacker(master) and Task Tracker (slave).

B. Apache Spark

Hadoop Map Reduce is a technique used to process historic data and analyse it. But it is not ideal for real time data as it is slow. This led to the development of Apache Spark which is a distributed processing framework which analyses large amounts of data (both structured and unstructured) at very high speeds. It uses RDD i.e. resilient distributed dataset that is fault tolerant. Spark uses iterative algorithms that checks the dataset multiple times in a continuous loop. The latency of applications that use the iterative approach and repeatedly query the database is greatly reduced in Apache Spark as compared to Apache Hadoop. Hadoop performs batch processing of data whereas Spark facilitates real time processing which is a requirement in dynamic toll generation.

C. Kafka

To enhance the project by obtaining real time data from various sources like traffic congestion level from Intelligent Traffic Systems, we will be using Apache Kafka. Apache Kafka is an open-source stream-processing software stage created by Linkedin and gave to the Apache Software Foundation, written in Scala and Java. It means to give a bound together, high-throughput, low-dormancy stage for dealing with continuous information nourishes. Its stockpiling layer is basically a "highly adaptable bar/sub message line planned as a distributed exchange log, "making it exceptionally significant for big business foundations to process gushing information. Furthermore, Kafka associates with outside frameworks (for information import/send out) by means of Kafka Connect and gives Kafka Streams, a Java stream preparing library. The structure is vigorously affected by exchange logs.

Talking about the integration of Kafka with Apache Spark Streaming, we have Kafka to be a potential forecasting and reconciliation stage for Spark Streaming. Kafka goes about to be the focal center point for ongoing floods of information and is handled utilizing complex algorithms in Spark Streaming. When the information is handled, Spark Streaming could be injecting the results into another Kafka point or store in HDFS, databases or dashboards. The accompanying chart delineates the speculative flow.

D. Type of NoSQL databases

1. Key Values Store: This stores the big hash table with keys and values. Example: Amazon S3
2. Document based store: It stores the document in the form of tagged element. Example CouchDB.
3. Column Based Store: Each block of storage contains data from only one column. Example: Cassandra.
4. Graph Based Store: A network database that uses nodes to store data and edges to represent relationship between data. Example: Neo4J.

E. Data Analysis

Data analysis is further divided as quantitative and qualitative analysis. Quantitative analysis techniques are:

1) Regression analysis:Regression measures the relationship between dependent and independent variable. It helps identify trends and optimizes the links between factors.
2) Hypothesis testing: This compares the data against the assumptions and hypotheses made for

the particular task. It compares two variables and finds the relation between them which further helps make decisions.

Qualitative analysis techniques are,

1) Content Analysis: It helps identify themes from the qualitative data. This method works best when user feedback is collected and based on the result, work on improvement.
2) Narrative Analysis: Focuses on ideas and narratives of employees from organizations.

III. Toll Charge System

The idea of toll charges was invented to allow the commuters to pay for transport infrastructure such as the construction of bridges and tunnels.

A. *Manual Toll Charge:* Tolls were initially collected by hand i.e. toll collectors would be deployed at the various toll collection areas and the vehicles would enter the toll lanes and pay the toll by cash. They would receive an acknowledgement receipt. Paying by cash is still an option, although everyone is switching to electronic toll collection system.

B. *Electronic Toll Collection:* ETC tackles the problem of losing time in paying toll as this system aims to allow payment of toll through RFID transponders and is coupled with the advantage of not stopping at the toll collection area. Dynamic toll charging is based on the real time traffic scenarios. Several factors can be considered while charging toll dynamically, for example, travel speed, seat occupancy, delay in traffic etc. All these factors are measured dynamically and the toll rates will be updated within short intervals of time. The most effective implementation would be charging toll as a response to changing traffic patterns. The first deployment of RFID electronic toll system in the US in 1989 by Amtech was on Dallas North Toll way. This technology was used by Amtech initially to tag and track livestock. The first all-automated toll highway in the world was Highway 407 in Ontario, Canada. It has no toll booths ,but reads the transponder mounted on the windshields of the vehicles. The vehicles that do not have the toll tag will be photographed (i.e the license plate will be photographed) automatically and a monthly bill will be sent to the owner of the vehicle which is higher than the toll of vehicles that use the toll tag. Now EZpass is widely used at all tollgates in US which deducts toll from RFIDs linked to e-wallets.

C. *Mobile Tolling Platforms:* Although the technology used to collect toll through RFID was invented in 1970s, it was implemented in 1980s and 1990s.Now, the states in US use this but have facilitated the toll payment through smart phones. There are various applications available to deduct money from prepaid accounts like Tollmate, TEXpress, BancPass, FastToll, Pass Manager, PayToll. These applications allow the users to add money to the account, view the previous transactions, add vehicle information, there are even reminders on these applications to pay tolls on time.

D. *Previously Proposed Models:* Recently, the macroscopic fundamental diagram (MFD) developed by both field experiments and simulation tests illustrates a unimodal low-scatter relationship between density network and the mean flow indicating that the state of network traffic is homogeneous. It unveils the traffic flow properties at a clustered level and emphasizes on dynamic traffic management of a massive network. A twin programming toll model, incorporating MFD to solve the unbalanced flow

distribution problem within the two-layer transportation networks was proposed by Bangyang Wei and Daniel Sun. Their primary focus is to minimize total travel time and assign link based traffic assignment to the network level. Drivers are often perplexed in deciding to choose between arterial road network and the loop expressway, and there are situations where one of them will be congested while the other observing low traffic density. Dynamic pricing has proved to be an effective scheme helping drivers choose lanes according to the optimal demand distribution on the network. Hence, there is practical need to develop toll strategy at regional level by means of macroscopic traffic flow theory which is consistent with the network traffic dynamics.

E. *Proposed System:* In 2003, London, in order to reduce traffic, established the London Congestion Charge, which is a fee charged on motor vehicles in the central city during weekdays. On the first day of implementation, 1, 90,000 vehicles moved around and within the congestion charge zone and a decrease of 25% in traffic levels was observed. Keeping this idea in mind, the project has been proposed to dynamically charge toll reducing waiting time and the toll charged is based on the seat occupancy (which is a major factor in reducing pollution).

Digital image processing was invented in the 1600s and slowly emerged as the cheapest form of image processing. Over the years, it gained better performance and is used as a key technology for classification, feature extraction, pattern recognition etc. One of the most challenging tasks is to recognize if any faces are present in the image and if so, record the face count. To extract the facial features, an algorithm may use parameters such as size and shape of permanent features like the eyes, nose and lips. After the feature extraction, a classification algorithm is used such as k-nearest neighbor or support vector machine classifiers. In this project, to determine the seat occupancy near the toll gates, this technology has been used and implemented in python. Python is a high level interpreted language which provides code readability. It has less and easily understandable code. It also has various support libraries (like pandas for data analytics), presence of third party module (openCV-an open source machine learning software library for image processing, face capture, face recognition etc.), high portability etc. Although openCV can be integrated with C, C++ or Java, we prefer to use python as openCV is a wrapper around the C++ code and when run in python, we achieve the best features of both the languages i.e. performance of C++ and the simplicity of python. To store the data that is used for processing the toll and structure them in a way convenient for future analysis, we chose to use Cassandra, a NoSQL database over other NoSQL databases such as MongoDb. One of the major reasons to opt Cassandra is its decentralized characteristic.

1. Decentralized – Every hub in the bunch has a similar job. There is no single purpose of disappointment. Information is disseminated over the group (so every hub contains distinctive information), however there is no ace as each hub can support any demand.
2. Scalability – Read and write throughput both increment straightly as new machines are included, with no downtime or interference to applications.
3. Blame tolerant – Data is naturally reproduced to various hubs for adaptation to non-critical failure. Replication over various server farms is upheld. Fizzled hubs can be supplanted with no downtime.
4. Tunable consistency – Writes and reads offer a tunable dimension of consistency, the whole distance from "writes never fall flat" to "obstruct for all replications to be meaningful", with the majority/quorum level in the center.

5. MapReduce support – Cassandra has Hadoop joining, with MapReduce support. There is support additionally for Apache Pig and Apache Hive.
6. Query language – CQL (Cassandra Query Language) was presented, a SQL-like option to the conventional RPC interface. Language drivers are accessible for Java (JDBC), Python (DBAPI2) and Node.JS (Helenus).

Now, talking about the integration of Cassandra with Python, there is already a 'cassandra-driver' a python client driver specially for Cassandra. The driver works with the Cassandra Query Language v3 and Cassandra's native protocol. For obtaining real time data and its monitoring, Apache Kafka is to be used. Apache Spark is used for real time data streaming which is useful for dynamically charging toll rather than Hadoop which can process only batches of data at a time.

To store the data that is used for processing the toll and structure them in a way convenient for future analysis, we chose to use Cassandra, a NoSQL database over other NoSQL databases such as MongoDB.

IV. CONCLUSION

A dynamic powerful toll estimating plan can possibly decrease traffic congestion without a huge framework costs that different choices, for example, street extension and open transportation advancement, require. In this paper, a dynamic toll-pricing procedure is proposed dependent on the inhabitance/occupancy and can be expanded by influencing the toll cost based on the traffic congestion through real-time information. A critical part of any new toll valuing framework is the impact it will have on income generation. Since the proposed pricing procedure can be implemented within the current framework with no adjustments to the foundation, any differences in income will result from the changes in toll prices and toll lane users. The methodology of utilizing dynamic toll pricing for the executives of interstate traffic can be seen as one of the essential building blocks of a Savvy City.

V. REFERENCES

[1]. G. Guerreiro, P. Figueiras, R. Silva, R. Costa e R. Jardim-Gonçalves, "An architecture for Big Data processing on Intelligent Transportation Systems: An application scenario on highway traffic flows," IEEE Intelligent Systems 2016, Sofia, Bulgaria, 2016.

[2]. J. Caserta and E. Cordo, "Big ETL: The Next 'Big' Thing," 9 February 2015. Online]. Available: http://data-informed.com/big-etl-next-bigthing/.

[3]. European Comission, "OPTIMUM," 2015. Online]. Available: http://www.optimumproject.eu/.

[4]. R. Lindsey e E. Verhoef, "Traffic Congestion and Congestion Pricing"Tinbergen Institute, Amsterdam, 2000.

[5]. Y. Lou, Y. Yin e J. Laval, "Optimal dynamic pricing strategies for high occupancy/toll lanes," Transportation Research Part C, vol.19, pp. 64-74, 2011.

[6]. The Apache Software Foundation, "Apache Kafka - A high-throughput distributed messaging system.," 2014. Online]. Available: http://kafka.apache.org/.

[7]. IBM. (2014). Building a smarter transportation management network. Retrieved August 20, 2015, from http://public.dhe.ibm.com/common/ssi/ecm/lb/en/lbw03019usen/LBW03019USEN.PDF

[8]. 495 — 95 Express Lanes - Pricing. (n.d.). Retrieved August 20, 2015, from https://www.expresslanes.com/pricing

[9]. Yin, Y., & Lou, Y. (2009)," Dynamic Tolling Strategies for Managed Lanes"Journal of Transportation Engineering.

[10]. Zhang, G., Ma, X., & Wang, Y. (2014)" Self-adaptive tolling strategy for enhanced high-occupancy toll lane operations" IEEE Transactions on Intelligent Transportation Systems, 15(1), 306?317.

[11]. USDOT VOT Guidance 2014. (n.d.). Retrieved August 20, 2015, from https://www.transportation.gov/sites/dot.gov/files/docs/USDOT VOT Guidance 2014.pdf

[12]. Thao Phan, Anuradha M. Annaswamy, Diana Yanakiev, and Eric Tseng(2016)," A Model based Dynamic Toll Pricing Strategy for Controlling Highway Traffic",from https://www.researchgate.net/publication/282663933

Cite this article as :

Merlin Mathew, Aishwarya Balakrishnan, Bikky Kumar Goit, Mounica. B, "Dynamic Toll Charge using openCV and Spark", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 5 Issue 2, pp. 1255-1260, March-April 2019. Available at doi : https://doi.org/10.32628/CSEIT1952326
Journal URL : http://ijsrcseit.com/CSEIT1952326