A STUDY ON SELF SIMILAR VEHICLE ARRIVAL PATTERN AT ISOLATED SIGNALIZED INTERSECTION

By

CHEW CARL JUN

A dissertation submitted to the Department of Civil Engineering,
Lee Kong Chian Faculty of Engineering and Science,
Universiti Tunku Abdul Rahman,
in partial fulfillment of the requirements for the degree of
Master of Science
August 2018

ABSTRACT

A STUDY ON SELF SIMILAR VEHICLE ARRIVAL PATTERN AT ISOLATED SIGNALIZED INTERSECTION

CHEW CARL JUN

In data network, network arrivals are modeled based on Poisson assumptions but later it was discovered that Local Area Network (LAN) traffic possesses self-similar characteristics (Leland et al., 1994). Paxson and Floyd (1995) discovered that Wide Area Network (WAN) traffic is well modeled by using self-similar process rather than Poisson process. Since then, self-similar processes have been studied and implied in data network.

Conventionally, the vehicle arrival pattern is also modeled based on Poisson assumptions in traffic studies. In traffic signal design, vehicle arrival pattern is important in calculating the queue length and cycle length. However, recent researches have shown that Poisson assumptions no longer hold under moderate to heavy traffic conditions. Nagatani (2005) discovered that individual vehicle that passes a series of traffic lights possessed self-similar characteristics. Meng and Khoo (2009) concluded that the vehicle arrival pattern on highway possessed self-similar characteristics. These recent studies have shown that vehicle arrival patterns possess self-similar characteristics. Therefore Poisson assumptions made on vehicle arrival pattern should be reconsidered and the existence of self-similar characteristics in vehicle arrival pattern should be examined.

This research aims at investigating the existence of self-similarity characteristics for vehicle arrival pattern and its impact on traffic signal design. This research emphasizes on the vehicle arrival pattern of isolated signalized intersections within the Kuala Lumpur City Centre. By recording the movement of incoming vehicles, the vehicle arrival patterns and its

corresponding time headway is tabulated and analyzed. Then, statistical analysis such as hypothesis tests was executed to examine the goodness-of-fit between Poisson process and self-similar process. It is discovered that the vehicle arrival pattern in these intersections exhibit self-similarity characteristics and its corresponding time headway distribution is heavy-tailed.

By affirming the self-similarity characteristics in vehicle arrival pattern, this thesis developed a new approach in calculating the average delay of the vehicles under different values of Hurst parameter, which is an indicator for self-similarity. Then, the formulation of queue length and cycle length are derived based on self-similar characteristics. It is aimed that this new approach can provide a better and more accurate alternative in the construction of traffic signal design to optimize the average delay of the vehicles. This is crucial in providing a more precise and accurate queue length and cycle length for traffic signal design.

ACKNOWLEDGEMENT

I would like to acknowledge all the lecturers and supervisors who have been providing guidance in the completion of this dissertation. I would like to express my gratitude to my main supervisor, Ir Dr Khoo Hooi Ling and my co-supervisors, Dr Gloria Teng Ai Hui and Dr Koh Siew Khew for their motivations and dedications in guiding me throughout the whole journey of this research.

Besides that, I appreciate the encouragement and motivation from my dearest family members and friends. Without their support, it would be challenging for me to complete this dissertation successfully.

APPROVAL SHEET

This dissertation entitled "**A STUDY ON SELF SIMILAR VEHICLE ARRIVAL PATTERN AT ISOLATED SIGNALIZED INTERSECTION**" was prepared by CHEW CARL JUN and submitted as partial fulfillment of the requirements for the degree of Master of Science at Universiti Tunku Abdul Rahman.

Approved by:

____
(Ir Dr Khoo Hooi Ling @ Lai Hooi Ling)
Date:28 August 2018
Main Supervisor
Department of Civil Engineering
Lee Kong Chian Faculty of Engineering and Science
Universiti Tunku Abdul Rahman

____
(Dr Koh Siew Khew)
Date: 28 August 2018
Co-supervisor
Department of Mathematical and Actuarial Sciences
Lee Kong Chian Faculty of Engineering and Science
Universiti Tunku Abdul Rahman

LEE KONG CHIAN FACULTY OF ENGINEERING AND SCIENCE

UNIVERSITI TUNKU ABDUL RAHMAN

Date: 28 August 2018

SUBMISSION OF DISSERTATION

It is hereby certified that **CHEW CARL JUN** (ID No: **15UEM06660***)* has completed this dissertation entitled "**A STUDY ON SELF SIMILAR VEHICLE ARRIVAL PATTERN AT ISOLATED SIGNALIZED INTERSECTION**" under the supervision of **Ir Dr Khoo Hooi Ling @ Lai Hooi Ling** (Supervisor) from the Department of Civil Engineering, Lee Kong Chian Faculty of Engineering and Science, and **Dr Koh Siew Khew** (Co-Supervisor) from the Department of Mathematical and Actuarial Sciences, Lee Kong Chian Faculty of Engineering and Science.

I understand that University will upload softcopy of my dissertation in pdf format into UTAR Institutional Repository, which may be made accessible to UTAR community and public.

Yours truly,

____

(*CHEW CARL JUN*)

DECLARATION

I hereby declare that the dissertation is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UTAR or other institutions.

____

(CHEW CARL JUN)

Date: 28 August 2018

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Introduction

In data network, one of the essential measurements for network performance is the average delay to transfer a data packet from the source origin to the destination. An efficient network should have a relatively low average delay. Hence, the development of traffic model is emphasized in data network. This is because a good traffic model that can reflect the scenarios in network traffic is useful in understanding its characteristics. The variables that contribute to a higher average delay can be identified and the performance of the network can be optimized.

Initially, the traffic model in data network is modeled by Poisson process. The preference of using Poisson process is motivated by the traffic model in traditional telephony network, in which the total number of incoming calls is independently and identically distributed (IID), and the average hold time of call is exponentially distributed (Frost and Melamed, 1994). This scenario is similar to data network, in which the total arrival of data packet is IID, and the inter-arrival time of the data packet is exponentially distributed. The Poisson process is widely implied in data network for simplicity in analysis and simulation. However, it is discovered that this deviates from the actual traffic flow and is not appropriate to be used in traffic model.

Later, it was discovered that the traffic models possessed the characteristics of self-similarity (Leland et.al, 1994). Then, the discovery of self-similarity in WAN traffic (Paxson and Floyd, 1995) and WWW traffic (Crovella and Bestavor, 1997) reinforced the development of traffic modeling. Nowadays, the implication of self-similarity in data network is widely established.

On the other hand, the traffic model in road traffic is also modeled by Poisson process. The total number of arriving vehicles is IID, and the inter-arrival time between two vehicles is exponentially distributed. For instance, by counting the number of arriving vehicles per 10 second period, it is observed that this vehicle arrival rate follows Poisson distribution (Adams, 1936). Since then, Poisson process has been used in road traffic studies. Tanner (1951) assumed that the arrival of traffic is Poisson distributed to derive the expected number of incoming road pedestrians in between successive vehicles. Kisi and Hiyoshi (1962) investigated the impact of the distance of traffic light on vehicle flow and the results revealed that the vehicle flow converges to Poisson distribution when the distance of traffic light is closer to the vehicles.

The traffic signal designs developed by Webster was based on the assumption that the arrival of vehicles is Poisson distributed (Webster, 1958). By using this assumption, he derived a series of formulae in calculating the optimal cycle length and the average delay per vehicle in isolated intersections. His formulations that are based on Poisson process serve as the fundamental in the construction of traffic signal.

However, Webster's traffic signal designs are useful when the traffic flow is not saturated. In modern road traffic, the traffic flow can be saturated and over-saturated due to an increasing volume of incoming vehicles on road over the years. Poisson process is not appropriate to be used under these conditions. For instance, it is observed that the average delay computed by Webster's method is inaccurate under highly saturated traffic flow (Dion, 2004). Motivated by this, studies are being carried out to find a better alternative traffic model in the construction of traffic signal designs.

Since the approach in traffic modeling is similar between data network and road traffic, in which Poisson process is initially used, researches are investigating whether road traffic also possesses self-similarity. Nagatani (2004) is the pioneer who used simulation model to discover the travel time of an individual vehicle moving across a series of traffic lights. He realized that the travel time exhibit self-similar characteristics. Besides that, Meng and Khoo (2009) used collected data from highway to conclude that the vehicle arrival pattern on highways do not follow Poisson assumptions but possess self-similar characteristics. These findings are significant to encourage that the incorporation of self-similarity characteristics in traffic signal model might provide insightful information that could eventually lead to optimization of traffic delay. Since the development of self-similarity in road traffic is relatively new as compared to data network, more efforts should be emphasized on understanding the relationship between self-similarity and road traffic.

This thesis continues the journey in investigating the characteristics of self-similarity in road traffic. The research is divided into two main studies. The first section investigates the properties of vehicle arrival pattern and its corresponding time headway in isolated intersections. It analyzes whether Poisson distribution is appropriate in modeling the vehicle arrival pattern of isolated intersections. It also analyzes whether this vehicle arrival pattern possesses self-similarity characteristics or not. Subsequently, its corresponding time headway is analyzed. If the results for the first section affirm that the vehicle arrival pattern possesses self-similarity characteristics and its corresponding time headway is heavy-tailed, this research will proceed to the next step of the studies. The second section of this paper incorporates the Hurst parameter in analyzing and computing the average delay and optimal cycle length of the isolated intersection. Hurst parameter is the parameter to estimate self-similarity characteristics. The results are compared with Webster's optimal cycle length and average delay per vehicle to determine which method is more appropriate to be used in traffic signal designs.

The significance for this study is to provide a basic understanding in incorporation of self-similarity characteristics in traffic signal designs. In data network, self-similarity characteristics have been widely implied in many areas such as WAN network, LAN network and etc. However, in road traffic, the incorporation of self-similarity in traffic modeling is still not developed. Hence, it is aimed that the results of this thesis can helps to optimize the current traffic signal designs. Another significance of this study is to provide a realistic perspective in analyzing the self-similarity characteristics in isolated intersections. This is because most of the researches are focusing on using theoretical assumptions and simulations to understand the characteristics of self-similarity of isolated intersections, which might deviate from the actual situation in road traffic. In this thesis, the vehicle arrival pattern and its corresponding time headway are investigated by using analytical approach, as there is very limited study on this topic by using actual data. Lastly, this study is significant in understanding the characteristic of vehicle arrival pattern and its corresponding time headway in Kuala Lumpur. Due to the rapid development in Kuala Lumpur, the traffic volume increased drastically, it is important to study the characteristics of the road traffic and finding an optimal cycle length to minimize the average delay per vehicle (Hossain, 2006).

1.2 Objectives of Study

The objectives of this study are:

- To examine the existence of self-similar characteristics on vehicle arrival pattern and the properties of its corresponding time headway distributions on isolated signalized intersections.
- To formulate a computation of optimum cycle length and average delay per vehicle by using the vehicle arrival patterns under different Hurst parameters.

- To compare the results obtained from the computation of cycle length and average delay per vehicle with the results obtained from Webster's formulae.

1.3 Scope of Study

This study aims at analyzing the properties of the traffic flow, which are the vehicle arrival patterns and its corresponding time headway. Two isolated intersections which are located in Kuala Lumpur are investigated. Isolated intersection is ideal to be investigated in this research because the incoming traffic flow is not affected by any traffic signal beforehand. Besides that, delay experienced by vehicle due to previous traffic light is minimized. Finally, isolated intersection helps to ensure that the traffic flow is relatively smooth and continuous.

In this study, vehicles that travel straight across the intersections are considered. Left turning traffic and right turning traffic are not considered, as these traffics are always disrupted by vehicles that are cutting into the vehicle queues. This is to reduce the confounding variables in this research and to provide an accurate analysis of vehicle arrival pattern.

1.4 Thesis Organization

The overview in this study will be explained in this section. There are a total of 6 chapters in this thesis.

Chapter 1 gives an introductory idea in regards to the discovery self-similarity characteristics in data network and road traffic. It also explains the significance in carrying out this research.

Chapter 2 gives a detailed explanation of the development and implementation of self-similarity characteristics in data network. Besides that, it provides an extensive explanation on the progress of the modeling of vehicle arrival patterns and its corresponding time headway, and how self-similarity characteristics are discovered and implied in road traffic.

Chapter 3 provides the theoretical background on the properties of self-similarity characteristics and heavy-tailed distribution. In latter part of the Chapter 3, the logic in the construction of optimum cycle length and average delay per vehicle by using Webster's method is explained.

Chapter 4 explains the methodology used in the collection and analysis of data. In latter part of Chapter 4, the computation of optimum cycle length and average delay per vehicle by incorporating the Hurst parameter is discussed.

Chapter 5 explains and elaborates on the results obtained from the collected data.

Chapter 6 concludes the findings of this research and proposes ideas on future works.

CHAPTER 2

LITERATURE REVIEW

2.1 Vehicle Arrival Pattern

One of the factors that affects the delay of traffic is the arrival rate. In data network, the arrival rate of data packet is defined as the average duration of one unit package arriving at the node of the traffic. By aggregating these arrival rates over a fixed period of time, the arrival pattern of data packet is constructed. Similarly, in road traffic, the arrival rate of vehicle is defined as the average duration of one vehicle arriving at a specific spot of the road. In this study, the vehicle arrival pattern is defined as a series of incoming vehicles crossing over a specific reference point, over a fixed period of time. It is the pattern constructed by aggregating the successive arrival of traffic over a fixed time period. This pattern might follow some distributions and many researchers have been modeling the vehicle arrival pattern in different types of distributions.

Initially, Poisson process is used in traffic modeling. In data network, the number of data packet arriving at a specific time period is following a random process, and its corresponding interarrival time follow exponential distribution. A random process, or commonly known as stochastic process, is defined as a set of random variables ordered by time (Knill, 1994). In road traffic, vehicle arrival pattern also follows a random process. This assumption is deduced because the traffic arrival pattern is assumed to be IID, and the distribution of interarrival time is memoryless. This indicates that each arrival is independent and is not influenced by precedent events. In vehicle arrival pattern, the number of incoming vehicles represents the random variables. It is noted that the successive arrival of traffic during a cycle is independent of the successive arrival of traffic in previous cycle (Teply, 1993). In short, by adopting the Poisson assumptions, the traffic flow is assumed to be memoryless, has independent arrivals, and is not fluctuating (Fricker and Whitford, 2005).

Adams (1936) is one of the pioneers who adopted Poisson process in analyzing the vehicle arrival pattern. By using the number of incoming vehicles per 10 seconds period, he observed that the vehicle arrival pattern follows Poisson distribution. It is notable that the arrival rate of the incoming vehicles is approximately at 220 vehicles per hour, and hence his study might not be applicable to traffic condition that is saturated and over saturated. However, his findings are significant, leading other studies in road traffic to be related to Poisson process.

Greenshields et al. (1947) also adopted Poisson process in vehicle counting. This is because the number of vehicles is a discrete variable. Their study has generated results that are sufficiently accurate to prove that Poisson process is appropriate to model the traffic flow. Similar to Adams' study, their study is also limited to traffic flow that is saturated or over saturated.

Kisi and Hiyoshi (1962) investigated the vehicle arrival pattern in Tokyo and Osaka. They have concluded that even if the traffic flow does not follow Poisson distribution at the beginning of the observation period, the traffic flow will have a tendency to follow Poisson distribution upon reaching the bottleneck of the road. They also found out that the impact of the distance of traffic light on vehicle flow. Their results show that the vehicle flow converges to Poisson distribution when the distance of traffic light is closer to the vehicles.

Later, Breiman (1968) modeled the flow of single phase traffic by using Poisson process. In his study, he assumed that the vehicle at time t=0 is distributed according to Poisson process, and the speed of each vehicle is IID. As the time converges to infinity, the traffic flow tends to converge to Poisson process. In reality, most of the roads in urban cities consist of more than single phase of traffic flow. Hence, it is essential to extend the study of vehicle arrival pattern on roads that have more than single phase traffic, for example a four-phase signalized intersection.

2.1.1 Inaccuracy of Poisson Process

As mentioned earlier, in data network, Poisson process is ideally used in modeling the network traffic because of its interesting statistical properties. However, it is later discovered that Poisson process is unable to fully reflect the actual situation of network traffic. A lot of researchers have discovered other distributions and stochastic processes that can fit the network traffic model in a more appropriate manner.

Paxson and Floyd (1995) discovered the inaccuracy in using Poisson process to model arrival traffic for WAN. Poisson process implied in Transmission Control Protocol (TCP) traffic showed a large deviation and is inappropriate to be used. However, they discovered the self-similarity characteristics within these network traffics. The inaccuracy of using Poisson process to model data network is due to the traffic burstiness. Traffic burstiness is a characteristic where the total units of data packets arriving at the node increases in a sudden manner. This will cause a “spike” in the network traffic distribution. This phenomenon is one of the properties that are possessed by self-similarity. Despite Poisson process is able to explain the random arrival of data packets, it is unable to capture the traffic burstiness.

Besides that, Leland et al. (1994) explained the inability of Poisson process in modeling network traffic. In reality, the network traffic is not smooth and consistent. When there is a sudden increase in the total number of data packets arriving at the node, the “spike” in the network traffic distribution will increase. If the traffic model is based on Poisson process, it always tends to underestimate the number of data packets at certain period of time.

Polaganga and Liang (2017) explained that despite majority of telecommunication systems were suitable to be modeled by Poisson distribution, if the system was involved with multiple service capacity, Poisson distribution might not be suitable. Their research focused in understanding the modeling of Long Term Evolution (LTE) and Long Term Evolution-Advanced (LTE-A) network traffics. The result showed that both LTE and LTE-A network traffics is indeed possessed self-similarity characteristics.

These findings have led the modeling of data network towards the implementation of self-similarity characteristics in traffic model, in which the accuracy in estimating the average delays of network traffic to improve drastically. Motivated by this, researchers in road traffic are eager to find out whether the situation is applicable for the modeling of road traffic. Clement et al. (2005) and Shalaiket et al. (2012) have demonstrated the similarities between data and road traffic networks in term of how the movable units (i.e. data packets in network traffic and vehicle arrival patterns in road traffic) are being transferred from its origin to destination. Network data are divided into smaller packets, which are transfer individually and is similar to traffic vehicles that are served individually at intersections. Hence, it is no surprise that Poisson process is also not appropriate in modeling the road traffic.

In road traffic, due to an increasing volume in traffic flow, the Poisson process can no longer be used to model traffic flow that are saturated and oversaturated. Although a real-time algorithm is developed to improve the performance of traffic signal based on Poisson process, the results affirmed that the Poisson assumptions have limit the algorithm to be applicable on non-congested period only (Zheng and Recker, 2013). Even though various studies have shown that Poisson assumption no longer holds for traffic flow with heavier density, Poisson assumptions are still commonly used to construct traffic signal (Comert, 2016). This is because Poisson model is relatively easy to construct as compared to other statistical models. For instance, Barua et al. (2015) stated that even though ARIMA model provided more accurate results in predicting vehicle arrival pattern, ARIMA model is a complex model while Poisson model is a simple model to be used in predicting vehicle arrival pattern.

Nowadays, the study of vehicle arrival pattern is divided into two main perspectives, namely the theoretical perspectives and the analytical perspectives. For theoretical perspective, the vehicle arrival pattern is investigated via scientific approaches such as simulations. It provides a specific and focus understanding on vehicle arrival pattern. For analytical perspective, the study of vehicle arrival pattern is methodologically driven. It provides a realistic and natural understanding on vehicle arrival pattern. In this thesis, the analytical perspective is preferred because there is very limited study on vehicle arrival pattern from actual data.

2.1.2 Theoretical Perspectives

To study the vehicle arrival pattern from the theoretical perspectives, a few scientific approaches are being used, for instance Queuing Theory and Markov chain. Simulation plays an important role in researches that are heavily based on theoretical perspectives because the data collected from simulation is used as the evidence that the traffic model is accurate in estimating the vehicle arrival pattern.

By using traffic simulation, Yan Xing et al. (2016) constructed a delay model for signalized intersection by analyzing the vehicle arrival pattern. For simplicity, the assumptions made in this study were (1) the rate of arrival was constant and (2) low saturated traffic flow and over saturated traffic flow are not considered. Since our research is aimed to construct a better alternative to reduce the average delay per vehicle, over saturated traffic flow must be considered, as analyzing it helps us to understand the root cause for this issue.

For time series analysis, Barua et al. (2015) constructed an autoregressive integrated moving average (ARIMA) model for signalized intersection to estimate the vehicle arrival pattern. Time series model is used in this study because it is a more appropriate candidate to model a stochastic process. In this study, it is shown that ARIMA model provides a better fit as compared to Poisson model. However due to the complexity in generating results, it is less practical to construct a traffic model based on ARIMA time series.

By using Markov chain model, Zuylen (1985) investigated the arriving vehicles at isolated intersections. He assumed that the rate of arrival follows Poisson process. His study developed a series of formulae related to Markov chain to estimate the average queue length and average delay and theoretically, the results are valid. Later, Viti and Zuylen (2004) improved this work to model overflow queues at signalized arterial corridors by using Markov chain model. The authors have addressed the limitations of their study. The Markov chain model is sufficient to find out the distribution of vehicle queues but it might not be applicable for other types of roads, due to the difference in road capacity and duration of green period of traffic light.

2.1.3 Analytical Perspectives

In analytical perspectives, the real data collected from actual field is used. It provides a realistic overview on the road traffic. In reality, the traffic flow is influenced by a lot of confounding factors, such as driver's behavior, the condition of the road, the types of vehicles that are using the road, and etc. These factors are not easy to be included in theoretical perspectives, as it is challenging to make assumptions based on these factors.

Pasagic et al. (1998) analyzed the vehicle arrival patterns in 2 intersections in Zagreb, Croatia. The study analyzed the vehicle arrival pattern in different time intervals. The results have shown that in non-peak period, Normal distribution can be used to model the vehicle arrival pattern and in peak period, Gauss distribution is more appropriate in modeling the vehicle arrival pattern.

Thakur et al. (2013) analyzed the vehicle arrival pattern in 6 main urban cities in the world. Their research focuses on traffic density, which is defined as the total number of vehicles on 1 km^2 of the road. In vehicle arrival pattern, the unit of measurement is vehicle per hour, whereas in traffic density, the unit of measurement is vehicle per km^2. The results have shown that the traffic density distribution is not well modeled by using Poisson distribution but different heavy-tailed distributions such as Lognormal distribution and Gamma distribution provide a better fit in modeling the traffic density distribution.

2.2 Self-Similarity Characteristics

Self-similarity characteristics are firstly discovered in hydrology. Hurst (1951) was finding the optimal storage size for the dams by using the historical water level data in Nile River. He observed that the water level follows a series that is persistent. If the water level is over-flowed in that particular year, the water level for that consecutive year tends to follow a higher average of over-flowed water level. On the other hand, if the water level is lowered in that particular year, the water level for that consecutive year tends to follow a lower average of water level. Hurst exponent, which is the parameter to measure the degree of self-similarity, is introduced from his work.

Later, Mandelbrot (1975) introduced the term "fractal" when he discovered the scale invariance in fractional Brownian motion (FBM). In layman's term, he defined fractal as a shape or pattern that can be separated into smaller fractions, in which the smaller fractions are similar in different dimension. In this section, the development of self-similarity in data network and road traffic is discussed.

In data network, there are various sources of researches that are related to self-similarity. Self-similarity has proved to be able to capture the traffic burstiness existed in network traffic. Hence, by incorporating the concept of self-similarity in traffic modeling, the traffic model can give a better estimation in the measurement of interarrival time of data packets.

Contributed by findings of Leland et al. (1994), there are various studies of self-similarity in data network. Leland's discovery of self-similarity in Ethernet has provided a fundamental method to investigate self-similarity in data network. By applying his methods, Crovella and Bestavors (1996) stated that self-similarity characteristics are found in World Wide Web traffic. Elagha and AlShafee (2007) also discovered the self-similarity characteristics in asynchronous transfer mode (ATM) network. Besides analyzing the self-similarity properties, they also managed to generate pseudo self-similar traffic by using chaotic map model.

However in road traffic, the discovery of self-similarity characteristics is a relatively new area as compared to data network. Due to the established assumptions in Poisson process and the convenience in adopting it in traffic modeling, adapting self-similarity properties in road traffic modeling are less attractive due to its complexity.

In road traffic, the most prevailing traffic model is built by Webster (1958). The formulation of queue length and cycle length is useful in optimizing the design of traffic signals. Webster (1958) formulated the average delay per vehicle based on Poisson assumptions. He discovered that optimum traffic condition could be archived if the total delay is minimized. His formulations are widely used in traffic signal design.

However, Webster's delay formula is inaccurate when there is overflow traffic condition. Cheng et al. (2003) modified Webster's minimum delay cycle length equation to give a more accurate optimal cycle length for overflow traffic condition. Van Zuylen and Viti (2006) formulated the queue length within a cycle time and random delay function based on Markov chain model.

Nagatani (2005) is the pioneer to analyze the self-similarity characteristics in road traffic. Nagatani discovered that a single vehicle passing through a sequence of traffic lights under different cycle time possesses self-similar behavior by using simulated data. This has suggested that the existence of self-similar is possible in road traffic.

Meng and Khoo (2009) discovered that self-similar behavior exists in the vehicle arrival pattern of highways. They also discovered the time headway in highways are heavy-tailed. However, due to the fact that the data for vehicle arrival pattern was collected from Texas, USA and the data for time headway is collected from Kuala Lumpur, Malaysia, there was no direct relationship between vehicle arrival pattern and its corresponding time headway.

Wong (2010) and Peratiet al. (2012) also discovered the self-similar behavior of vehicle arrival pattern on arterial street roads and highways respectively. Whereas Thakur et al. (2013) performed a large scale of analysis on traffic density in six cities around the world. The result showed that vehicle arrival patterns should be modeled by heavy-tailed distribution, which indicated a self-similar process.

To close up the gap from the previous studies mentioned above, our thesis investigates the vehicle arrival pattern and its corresponding time headway in isolated intersections. Besides that, since the time headway of traffic in isolated intersections is greatly dependant on the cycle time of traffic light, we incorporate the Hurst parameter in calculating the optimum cycle time and least average delay of vehicles.

2.3 Time Headway

There is a close relationship between vehicle arrival pattern and time headway. Time headway is defined as the time gap between two successive arriving vehicles at a specific spot of the road. Luttinen (1996) defined time headway as the duration of time consumed by a vehicle to travel across a specific point where its previous vehicle has passed through. It is measured between the front wheels of two vehicles. Hence, time headway is derived from the vehicle arrival pattern, as the vehicle arrival is measured by the time where the front wheels of the vehicle travel across a specific point. For instance, we assume that the front wheels of first vehicle travel across a specific point at time t_1=5 s and the front wheels of second vehicle travel across the same point at time t_2=8 s. The time headway for second vehicle is just the difference of t_1 and t_2, which is $t_2 - t_1 = 3$ s.

Time headway in road traffic is similar to the interarrival time of data packets in data network. In data network, the interarrival time is defined as the time interval between two units of arriving data packets. In data network, the interarrival time plays as an indicator of the performance of data transfer. The bottleneck effect that occurs within the node of network can be easily detected by analyzing the interarrival time of data packets (Varga and Kun, 2005). Besides that, the estimation of interarrival time of data packets is important in fair bandwidth sharing algorithm. It can be used to measure the quality of service (QOS) of network (Phit and Abe, 2006).

In road traffic, time headway distribution can either be modeled by using common distributions such as Lognormal distribution, Normal distribution and Pereto distributions, or a mixture of two or more distributions such as hyper-Erlang, semi-Poisson distribution and Cowan M3 distrubution (Zhang et al., 2007). The existence of various time headway distributions are due to the complex nature of time headway. Factors such as road capacity, traffic volume and types of intersections can influence the driving behavior, thus affecting the time headway distributions.

Initially, the vehicle arrival pattern of traffic model is modeled by Poisson process. The interarrival time of Poisson process follows an exponential distribution. This contributes to the memoryless characteristics in the time headway distribution. Gerlough and Huber (1975) stated that exponential distribution is not appropriate to be used to model time headway. This is because exponential distribution is light-tailed. The probability density function of exponential distribution will produce time headway that has very short duration. Buckley (1968) also stated that exponential distribution fails to model the time headway distribution because most of the time headway generated by exponential distribution tends to have very low values.

Cowan (1975) proposed Cowan M3 model to estimate the time headway in unsignalized intersection such as roundabout. He proposed that the time headway consists of mixture of following vehicles and non-following vehicles. The following vehicles are modeled by single headway Δ, which represents the minimum headway of these vehicles, and the non-following vehicles are modeled by shifted exponential distribution. Cowan M3 model is proven to be practical to be used in traffic modeling for unsignalized intersection because of its simplicity in calculation.

Branston (1976) proposed Generalized Queueing Model (GQM) to estimate the time headway in single lane. In GQM, initially the time of arriving vehicles are determined by Poisson process. Then, this arrival time is modified such that the time headways are greater than the time headway of following vehicles. For non-following vehicles, the time headway consists of a mixture of following vehicles and negative exponential distribution. Rossi et al. (2014) stated that the GQM model is proven to be accurate in estimating the time headway distribution for different traffic flow.

However, the estimation of time headway in modern road traffic is not accurate by using the traditional models mentioned above. This is contributed by the increase in traffic volume, causing a longer queue length that influences the stability of saturation flow. Lin (2003) observed that the traditional models tend to underestimate the saturation flow by approximately 400 vehicles per hour.

Cong et.al (2005) investigated the time headway distribution of motorcycle traffic in Hanoi, Vietnam. This is because motorcycles are the main vehicle types used in Vietnam. Their results showed that normal distribution can describe the time headway distribution well. The time headway distribution for motorcycle deviates from the time headway distribution for other vehicle types. Hence, attention must be given when defining the subject of investigation before modeling its time headway distribution.

Jang et al. (2011) investigated the time headway distribution in arterial roads in Gyeonggi, South Korea. Their results show that Johnson S_U model and Johnson S_B model were the appropriate model to describe the time headway distribution in arterial road. Johnson S_U model is a four-parameter model transformed from the normal distribution and Johnson S_B model is transformed from Johnson S_U model. They mentioned that the deviation in searching the appropriate models for each study is due to the characteristics of the road. Road capacity and traffic light designs are the factors that influence the time headway distribution. Hence, it is essential to include these factors in the study of time headway distributions.

2.4 Traffic Model

Vehicle arrival pattern and its corresponding time headway is the building block to construct the traffic model. An accurate traffic model that can describe the actual situation on road traffic is essential for traffic signal designs.

In road traffic, traffic models can be categorized into two types, namely the deterministic models, and the time dependant delay models. For deterministic models, the most prevailing traffic model is constructed based on Webster's method. On the other hand, for time dependant delay models, the H.C.M delay model is commonly used.

Wardrop (1952) developed a formula to calculate the average delay per vehicle by assuming that the vehicle arrival follows a uniform distribution. He expressed the delay formula as,

$$d = \frac{(r - 0.5s)^2}{2C(1 - y)} \quad (1)$$

where

d = average delay per vehicles

r = red duration

s = saturation flow rate

C= cycle length

y = vehicle flow ratio

Later, Webster (1958) has proposed a traffic model that assumes that the vehicle arrival is random and follows Poisson process. Webster's assumption on vehicle arrival becomes the fundamental assumptions in traffic signal designs and is used until now. Since this paper compares the average delay obtained from computation with the average delay from Webster's formula, the terminology and the theory behind Webster model is discussed in Section 3.3.

However the shortcoming of Webster's model is the inaccuracy in estimating the average delay for traffic conditions that are saturated and over saturated. Webster's model works well in traffic condition that is under saturated only. Hence, Miller (1963) encountered the overflow condition by introducing a upper boundary value (u) to estimate the average overflow queue, and this is represented by

$$u = \frac{\exp\left[\frac{(-1.33)\sqrt{gs}(1-x)}{x}\right]}{2(1-x)} \tag{2}$$

where

g= effective green duration

x=degree of saturation

Wormleighton (1965) discovered the randomness in the arrival rate by analyzing the collected data from signalized intersection. His results for average queue in overflow condition matched with the estimated results from Miller (Akcelik,1980).

As for delay model, Miller's proposed a delay formula for under saturated condition that produced similar results when compared with Webster's formula, which is

$$d = \frac{(1-\lambda)}{2(1-\lambda x)}(C(1-\lambda) + \frac{(2x-1)I}{q(1-x)} + \frac{I+\lambda x - 1}{s}) \tag{3}$$

where

I= ratio of variance to mean

λ = fraction of effectively green time in one cycle time

As for oversaturated condition, Miller incorporates the upper boundary value into his delay formula, which then becomes

$$d = \frac{(1-\lambda)}{2(1-\lambda x)}[C(1-\lambda) + \frac{u}{q(1-x)}] \tag{4}$$

Miller claimed that the delay formula for oversaturated condition is the solution to overcome the shortcoming for Webster's formula. Since then, many researchers have modified the delay model designed by either Webster or Miller. McNeil (1968) assumed that the vehicle arrival distribution to follow compound Poisson distribution. He incorporated Miller's upper boundary value in his formula, and is represented by

$$d = \frac{(c-g)}{2c\left(1-\frac{q}{s}\right)}((c-g) + \frac{2}{q}\left[1 + \frac{\left(1-\frac{q}{s}\right)(1-B^2)}{2s}\right]u + \frac{1}{s}\left(1 + \frac{I + \frac{B^2 q}{s}}{1-\frac{q}{s}}\right)) \tag{5}$$

where

B=index of dispersion for vehicle departure

However, McNeil's formula is complex and Ohno (1978) proposed another modified Miller's formula, in which he assumed the vehicle arrival distribution to be following a Poisson process. His delay equation is

$$d = \frac{c(1-\lambda)^2}{2(1-y)} + \left(\frac{1-\lambda}{1-y}\right)\left(\frac{u}{q}\right) + \left(\frac{1-u}{2s(1-y)}\right) + \frac{1-u}{2s(1-y)^2} \tag{6}$$

Ohno's equation is in a simpler form, but it failed to estimate the traffic condition that has a relatively low degree of saturation (Akcelik,1980).

On the other hand, Cheng et al. (2013) modified Webster's formula to overcome the shortcoming in estimating the delay in oversaturated condition. They modified Webster's formula based on HCM 2000, and is represented by

$$d = d_1 PF + d_2 + d_3 \tag{7}$$

where

d_1 = average delay of vehicle by assuming vehicle arrival is uniform

PF = progression factor (the adjustment factor for the effect of traffic signal)

d_2 = average delay of vehicle by assuming random arrival

d_3=initial queue delay

The first term of the delay equation is following the first term of Webster's formula. whereby

$$d_1 = \frac{0.5c\left(1 - \frac{g}{c}\right)^2}{1 - \left[\min(1, x)\frac{g}{c}\right]} \quad (8)$$

The biggest difference of this formula as compared to Webster's formula is that the denominator of the first term will not be 0, as to encounter the shortcoming of Webster's formula in handling saturated and oversaturated traffic flow.

However, the 2nd term of this formula is different as compared to Webster's formula, as it contains the duration of analysis period to increase the accuracy of the estimation. The 2nd term is represented by

$$d_2 = 900T[(x - 1) + \sqrt{(x - 1)^2 + \frac{8kIx}{CT}} \quad (9)$$

where

T=duration of analysis period

k=incremental delay factor

I=upbound traffic adjustment factor

From their results, HCM 2000 model outperforms Webster's formula, especially for traffic condition that are nearly saturated and over saturated. However, due to the limitations of their research, in which they limit the analysis period to 15 minutes, they would want to further their study to create a generalized model.

2.5 Conclusion

Although the existence of self-similar characteristics in network traffic are well studied, studies regarding the existence of self-similar behavior in road traffic are remained as an open-ended question. This thesis is significant in providing a different perspectives for self-similarity characteristics in road traffic, as most of the researches are using simulations to model the road traffic, which do not necessarily represent the actual scenarios on the road. Furthermore, most of the studies emphasized in analysis of the properties of self-similarity characteristics on road traffic. Therefore, this thesis is significant in incorporating the self-similarity characteristics in traffic signal construction and acts as a milestone in optimizing the cycle time in traffic and reducing the average delay of vehicles.

CHAPTER 3

BACKGROUND OF STUDY

3.1 Characteristics of Self-Similarity

Generally, self-similarity is a process with statistical invariance over a period of timescales. The property of a self-similar object is bounded with respect of rescaling of time or space. A second-order self-similar process is defined as below.

Let $X = \{X_t, t = 1,2,\dots\}$ be a covariance stationary or wide-sense stationary stochastic process with mean $\mu = E[X_t]$ and variance $\sigma^2 = E[(X_t - \mu)^2]$. The autocorrelation of X, $r(k)$, is defined as:

$$r(k) = \frac{E[(X_t-\mu)(X_{t+k}-\mu)]}{E[(X_t-\mu)^2]}, k \geq 0 \qquad (10)$$

Let $X^{(m)} = \left(X_k^{(m)}, k = 1,2,\dots\right)$ be a new stochastic process by averaging the original stochastic process X over the non-overlapping block of size m such that:

$$X_k^{(m)} = \frac{1}{m}\sum_{t=[m\times(k-1)]+1}^{k\times m} X_t\,, k = 1.2,\dots \qquad (11)$$

The stochastic process X is an exact second-order self-similar process with Hurst parameter $H = 1 - \frac{1}{2}\beta$ where $0 < \beta < 1$, if the autocorrelation of the function $r(k) = r_k^{(m)}$ and according to equation (10):

$$r(k) = r_k^{(m)} = \frac{1}{2}[(k+1)^{2H} - 2K^{2H} + (k-1)^{2H}], \forall k \geq 1 \qquad (12)$$

A notable characteristic of self-similar process is it possesses long-range dependency (LRD). Self-similar process with LRD has an autocorrelation function $r(k) \sim k^{-\beta}$as $k \to \infty$. This shows that the autocorrelation function is decaying slowly and hyperbolically. (Crovella and Bestvros, 1995). This implies that self-similar process with LRD has an autocorrelation function that is non-summable, i.e$\sum_{k=1}^{\infty} r(k) \to \infty$. Hence, self-similar process with LRD can be easily characterized by a single parameter which is the Hurst parameter and the parameter is bounded in $\frac{1}{2} < H < 1$. The degree of self-similarity increases as $H \to 1$.

As the degree of self-similarity can be measured by estimating the Hurst parameter of the empirical data from a few perspectives, for example if Hurst parameter is estimated via time-domain analysis, R/S-statistic is used; if Hurst parameter is estimated via frequency-domain analysis, the periodogram is used; other methods such as absolute value method, the Whittle estimator, the Abry-Veitch method, variance-time plot and variance of residuals method can also be used in estimating the Hurst parameter.

However, R/S-statistic and variance-time plot are not adequate to estimate the Hurst parameter for refined data analysis due to the lack of limit law (Leland et.al, 1994). On the other hand, the Whittle estimator is the most robust estimator to estimate the Hurst parameter (Karagiannis et.al, 2003).

3.2 Heavy-tailed Distribution

In data network, the network traffic is observed to follow heavy-tailed distribution. Paxson and Floyd (1995) have discovered that the FTP data follows heavy-tailed distribution. Crovella and Bestavros (1997) have discovered that WWW traffic also follows heavy-tailed distribution. Supported by these findings in the data network, it is observed that heavy-tailed distribution is more appropriate in modeling the road traffic. In road traffic, the traffic density in some metropolitan cities follows heavy-tailed distributions such as Lognormal, Weibull and Log-gamma (Thakur et.al, 2013). In Malaysia highways, the time headways are observed to follow heavy-tailed distribution as well (Meng and Khoo, 2009).

It is observed that if a distribution is heavy-tailed, its tailed distribution can be written as:

$$P[X > x] \sim x^{-\alpha}, 0 < \alpha < 2 \qquad (13)$$

when $x \to \infty$. Equation 13 represents that the distribution is heavy-tailed if its asymptotic shape is hyperbolic. It is known that if $\alpha \leq 2$, the distribution possesses infinite variance and if $\alpha \leq 1$, then the distribution possesses infinite mean.

An approach to investigate whether a given distribution is indeed a heavy-tailed distribution is by using the log-log complementary distribution (LLCD) plot. LLCD plot is the plot of the tailed-distribution $\bar{F}(x)$ on log-log axes, in which $\bar{F}(x)$ is estimated from the data. From equation 14, a heavy-tailed distribution has the characteristic that:

$$\frac{d \log \bar{F}(x)}{d \log x} = -\alpha, x > \theta \tag{14}$$

for some θ. To inspect the distribution is heavy-tailed, firstly the value of α is estimated by choosing the value of θ which makes the LLCD plot to appear linear. Then, equally space points are selected from the points of LLCD plot which is bigger than θ. Lastly, the slope α is estimated by using least square regression method. As described earlier, heavy-tailed distribution possess infinite variance and the weight of their tails can be described by the parameter $\alpha < 2$ (Rezaul and Grout, 2007).

The time headway is then fitted to selected heavy-tailed distributions such as Weibull, Pareto and Lognormal distribution. Exponential distribution, which is a light-tailed distribution, is also included in the distribution fitting process. After fitting these distributions to the time headway, different goodness-of-fit statistics is used to select the appropriate distribution. In this paper, the package *fitdistrplus* from R is used to generate goodness-of-fit statistics. Since the time headway is a continuous distribution, Anderson-Darling (AD), Kolmogorov-Smirnov (KS) and Cramer-von Mises (CM) statistics are used. Aikake's Information Criterion (AIC) and Bayesian Information Criterion (BIC) are also used in justifying the results obtained from the goodness-of-fit statistics.

3.3 Webster's Average Delay

Traditionally, Webster has developed a series of formulae to aid the traffic signal design. These formulae are being modified over years but are still being used in modern traffic signal design. However, due to the facts that his formulae are being developed under the assumption that vehicle arrivals follow Poisson distribution, this contradicts the fact that modern vehicle arrival patterns possess self-similarity characteristics and are not Poisson distributed.

Webster has deduced a formula for average delay per vehicle, which is

$$d = \frac{c(1-\lambda^2)}{2(1-\lambda x)} + \frac{x^2}{2q(1-x)} - [0.65(\frac{c}{q^2})^{\frac{1}{3}} x^{(2+5\lambda)}] \quad (15)$$

where

d = average delay per vehicle

c = cycle time in seconds

λ = fraction of effectively green time in one cycle time

q = vehicle flow rate

s = vehicle saturation flow rate

x = degree of saturation, i.e. the ratio of real vehicle flow rate to the saturation flow rate, which is given by $x = \frac{q}{\lambda s}$

Besides that, Webster has deduced the optimal cycle length formula to provide the most optimal cycle time for least delay of the traffic, which is

$$c_0 = \frac{1.5L+5}{1-\sum Y} \quad (16)$$

where

c_0 = optimal cycle time

L = total lost time in a cycle

$\sum Y$ = summation of the ratio of actual vehicle flow rate over saturation flow for all arms of the intersection

From equation (15), it can be seen that vehicle flow rate is one of the factors that influences the average delay. However, since the vehicle arrival is random, and the queuing of the vehicle in an intersection is greatly dependent on the green and red phase of traffic light, it is more insightful to use vehicle arrival pattern that possesses self-similarity characteristics in computing the average delay. It is notable that higher value of self-similarity will lead to longer queuing delays (Sikdar et.al, 2002). From this perspective, it might be insightful to use vehicle arrival patterns that possess self-similarity characteristics in computing the average delay.

CHAPTER 4

METHODOLOGY

4.1 Data Description

This research emphasizes on the properties of the traffic flow in Kuala Lumpur City Centre. Two signalized isolated intersections are chosen, which are the (1) intersection between Jalan Tumbuhan and Jalan Genting Klang, and (2) intersection between Jalan Taman Ibu Kota and Jalan Langkawi. Both of the intersections consist of 4 approaches. In this research, the lanes in which the vehicles are moving straight across the intersections are considered. For the 1st intersection, there are a total of 10 lanes from 4 approaches, and for the 2nd intersection, there are a total of 8 lanes. Therefore, in total, the data are being collected from 18 lanes for both intersections. The data collection was performed from 13 August 2015 to 26 August 2015. The locations of the intersections are shown in Figure 4.1.

Figure 4.1 Locations of the intersections for data collection

The video cameras were set up on the light poles around 3 m from the ground. This ensures the video cameras to be able to capture incoming vehicles from approximately 500 m distance from the light poles.

Video cameras were set up on 4 arms of the intersections and the traffic flow is recorded continuously for 24 hours for two-week period. This ensures that the data is collected from various sources and the data obtained is continuous. The statistical characteristics for both intersections are tabulated in Table 4.1. From Table 4.1, it can be observed that the maximum vehicle flow rate for the 1st intersection is approximately 350 vehicle/hour, whereas the 2nd intersection has a lower rate of 270 vehicle/hour.

Table 4.1 Statistical properties of intersections

Intersection	Lane	Investigation Period	Average Flow Rate (veh/hr)		Total Number of Vehicles Per Investigation Period (veh)
			Min	Max	
S1C1	1	24 Hours	37	356	5467
	2	24 Hours	81	425	7321
	3	24 Hours	27	364	5356
S1C2	1	24 Hours	22	311	4117
	2	24 Hours	31	398	5531
S1C3	1	24 Hours	17	270	3804
	2	24 Hours	42	381	5805
	3	24 Hours	18	348	4868
S1C4	1	24 Hours	17	356	4328
	2	24 Hours	21	355	4481
S2C1	1	7AM - 7PM	157	272	2489
	2	7AM - 7PM	200	355	3381
S2C2	1	7AM - 7PM	94	186	1697
	2	7AM - 7PM	146	310	2675
S2C3	1	7AM - 7PM	105	206	1785
	2	7AM - 7PM	163	302	2749
S2C4	1	7AM - 7PM	113	224	1960
	2	7AM - 7PM	175	324	3000

For video recordings that are affected by factors such as road constructions, stationary vehicles by the roadsides and blurry video recordings, these video recordings are not used due to its inaccuracy. Due to this limitation, the timeframe for second intersection was limited from 7:00:00 AM to 7:00:00 PM. While for the first intersection, 24-hour data is investigated. However, this study ensures a total data set for 10 days was used. Therefore, there are 18 lanes that are being investigated and the total datasets used are 180.

For the video recordings collected from the sites, a reference line is added. As mentioned earlier, the reference line is important because it is the point whereby the timestamp of the vehicle is recorded. When the front wheels of the vehicle passed through the reference line, its timestamp is recorded. The screenshot of the video recordings is shown in Figure 4.2.

Figure 4.2 Screenshot of the video recordings with reference line

The timestamps for each vehicle is tabulated and analyzed to obtain its vehicle arrival pattern. A sample of the tabulation for vehicle arrival pattern is shown in Table 4.2.

Table 4.2 Vehicle arrival pattern at different time scale

	Total Number of Vehicles (One Lane)					
Timestamp	veh/1 min	veh/2 min	veh/4 min	veh/16 min	veh/32 min	veh/64 min
1:00:00 PM	4	9	17	73	147	286
1:01:00 PM	5					
1:02:00 PM	5	8				
1:03:00 PM	3					
1:04:00 PM	4	10	22			
1:05:00 PM	6					
1:06:00 PM	5	12				
1:07:00 PM	7					
1:08:00 PM	4	5	16			
1:09:00 PM	1					
1:10:00 PM	3	11				
1:11:00 PM	8					

1:12:00 PM	2	9	18			
1:13:00 PM	7					
1:14:00 PM	5	9				
1:15:00 PM	4					
1:16:00 PM	4	7	14	74		
1:17:00 PM	3					
1:18:00 PM	4	7				
1:19:00 PM	3					
1:20:00 PM	7	12	24			
1:21:00 PM	5					
1:22:00 PM	6	12				
1:23:00 PM	6					
1:24:00 PM	2	8	15			
1:25:00 PM	6					
1:26:00 PM	3	7				
1:27:00 PM	4					
1:28:00 PM	6	15	21			
1:29:00 PM	9					
1:30:00 PM	1	6				
1:31:00 PM	5					
1:32:00 PM	4	4	15	59	139	
1:33:00 PM	0					
1:34:00 PM	5	11				
1:35:00 PM	6					
1:36:00 PM	3	10	13			
1:37:00 PM	7					
1:38:00 PM	3	3				
1:39:00 PM	0					
1:40:00 PM	3	11	17			
1:41:00 PM	8					
1:42:00 PM	4	6				
1:43:00 PM	2					
1:44:00 PM	3	7	14			
1:45:00 PM	4					
1:46:00 PM	1	7				
1:47:00 PM	6					
1:48:00 PM	3	8	15	80		
1:49:00 PM	5					
1:50:00 PM	4	7				
1:51:00 PM	3					
1:52:00 PM	9	11	23			
1:53:00 PM	2					
1:54:00 PM	5	12				
1:55:00 PM	7					

1:56:00 PM	6	14	23			
1:57:00 PM	8					
1:58:00 PM	3	9				
1:59:00 PM	6					
2:00:00 PM	2	8	19			
2:01:00 PM	6					
2:02:00 PM	7	11				
2:03:00 PM	4					
2:04:00 PM	6	9	20	56	120	282
*continue	*continue	*continue	*continue	*continue	*continue	*continue

The aggregation of the number of incoming vehicles is crucial in graph plotting to show the vehicle arrival pattern in different time scales.

4.2 Analysis of Properties of Isolated Intersection

4.2.1 Vehicle Arrival Pattern

The first part of the data analysis is focused on investigation of Poisson distribution on the vehicle arrival pattern, as Poisson assumptions are initially used to model the vehicle arrival pattern. The hypothesis testing for Poisson process is as followed:

H_0: The vehicle arrival pattern follows Poisson distribution

H_1: The vehicle arrival pattern does not follow Poisson distribution

By using R, a statistical computing software, the p-value of the hypothesis testing is obtained. The p-value serves as a measure on whether the null hypothesis should be accepted or not. The null hypothesis will be accepted if the p-value is larger or equal to the value of level of significance. For this research, since the confidence interval is set to be within 95%, the value of level of significance is 0.05.

Then, SELFIS tool is used to investigate the existence of self-similarity characteristics in the empirical data. However, in this research, the graphical method is also implied to fully illustrate the effect of self-similarity characteristics in the vehicle arrival pattern. It is considered as the intuitive method to investigate the existence of self-similarity characteristics. Time series graphs are plotted with respect to different timescales and the patterns of the graphs are observed. Figure 3 shows the overview of the graphs at time scales of 1 minute, 2 minutes, 4 minutes, 10 minutes, 30 minutes and 60 minutes time interval per investigation period.

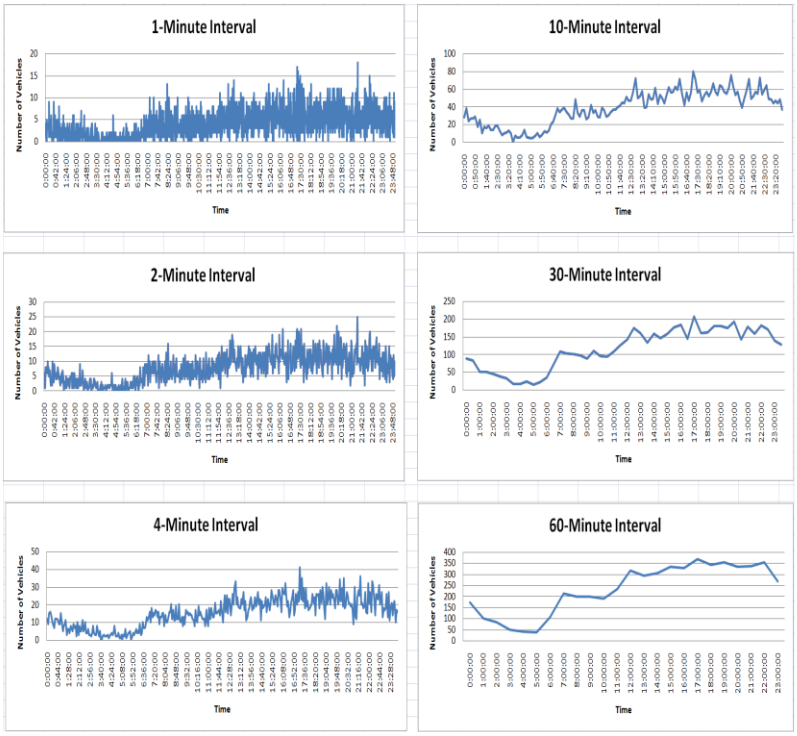

Figure 4.3 Vehicle arrival patterns of six different time scales for S1C1 Lane 1 on 19 August 2015

4.2.2 SELFIS (Self-similarity Analyis) Tools

SELFIS (self-similarity analysis) tools, which is developed by Karagiannis et al. (2003) is an open source Java software to investigate the degree of self-similarity. Hurst exponent is estimated from different methods and the user can choose to use any types of Hurst estimator. It is notable that a small time series sample that contain data lesser than 64 values are not applicable in SELFIS. Hence, this research is estimating the Hurst parameter for each lane per day, in which the total number of values is always more than 64 values.

Long range dependence (LRD) is closely related to the autocorrelation function (ACF). In Section 3.1, the LRD possesses the characteristics of ACF that decays hyperbolically. This indicates an ACF that will only decay to zero after a long period. As for short range dependence, its ACF is decaying exponentially. It is observed in certain case, the short range dependence decreases to zero at certain time lag (Leland et al., 1994). However, it is complex and time consuming to estimate the ACF.

A simpler method to determine the LRD is via Hurst estimator. A Hurst estimator that is close to 0.5 indicates that the distribution is not LRD. A Hurst estimator that is close to 1 indicates that LRD is present. From the introduction of Hurst exponent by Hurst in 1951, various types of estimators are introduced in estimating the Hurst exponent. In SELFIS, there is a total of 7 types of Hurst estimator that can be used. The list of estimators implemented in SELFIS is as follow.

1. **R/S (Rescaled Range) Method**

   R/S method was firstly introduced by Hurst (1951) to measure LRD. It provides consistent accurate results because of its robustness against the fluctuation in marginal distribution (Leland et al., 1994). However, there is a few limitations. Firstly, R/S method is not suitable for small samples. This is because R/S method employs heuristic methods, in which different time lag at different points are used in plotting. A limited sample will cause the estimator to be inaccurate. Secondly, R/S method is only accurate for linear processes. For non-liear processes, a modified R/S method is used (Cajueiro and Tabuk, 2005).

2. **Variance Method**

Since the variance of decays is slower than the reciprocal of experiment size, variance method can be used to estimate the Hurst parameter. Slope β is obtained via plotting log-log plot. Then by using the relationship between the Hurst exponent, H and β such that $H = 1 - \frac{\beta}{2}$, we can obtain the Hurst estimator.

3. **Absolute Value Method**

The non-overlapping aggregated series $X^{(m)}$ where m is the block size is used in this method. The log of the aggregated series is plotted against the 1st moment of the aggregated series. If LRD is presence, the line plotted is a straight line with a slope of $\beta = H - 1$.

4. **Periodogram Method**

Periodogram method estimates the Hurst exponent via the pattern of power spectral density of a time series. The Hurst parameter is obtained via the slope β of the straight line of the power spectral density, which is $\beta = 2(1 - H)$.

5. **Whittle Estimator**

This method is a parametric method by minimizing the likelihood function and applying it on Periodogram. The Hurst parameter is then obtained via the slope β of the straight line of the power spectral density, which is $\beta = 2(1 - H)$.

6. **Abry-Veitch**

Abry-Veitch is a method to estimate the Hurst parameter via the coefficient of wavelet decomposition. The cooeficient is obtained by using the orthogonal basis function. The energy of the wavelet decomposition in different scales are used in Hurst estimation. However, it is greatly affected by noise that existed in LRD and tends to over-estimate the Hurst exponent (Karagiannis, 2002).

7. **Variance of Residuals**

The idea of variance of residuals method is similar to the variance method. However, in variance of residuals, the aggregated series is plotted against the average variance residuals of the aggregated series in log-log plot. By using the relationship between the Hurst exponent, H and β such that $\beta = \frac{H}{2}$, the Hurst parameter can be obtained.

In this paper, Whittle estimator is preferred. This is because Whittle estimator has the ability to provide consistent and robust results as compared to the other types of estimators.

The dataset is being input into SELFIS tool to investigate the value of Hurst parameter of the vehicle arrival pattern. A total of seven types of estimators are used to estimate the Hurst parameter, as shown in Figure 4.4.

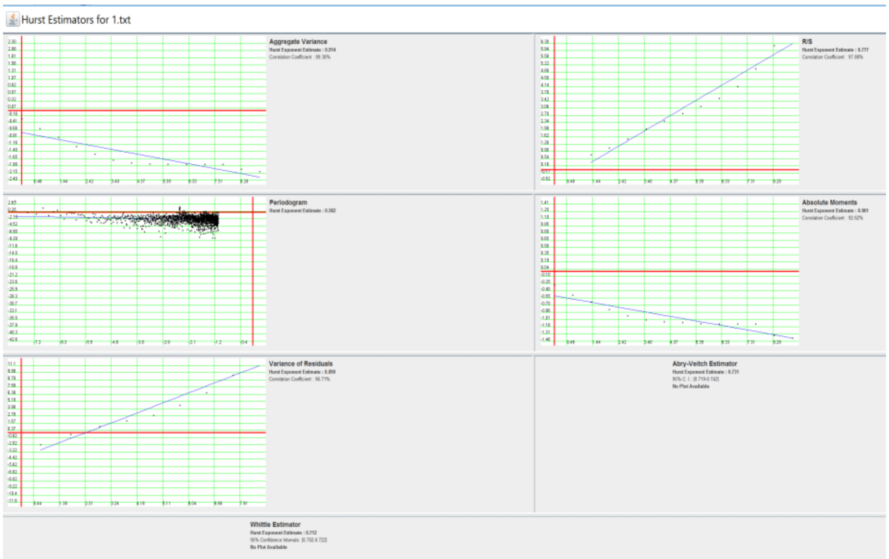

Figure 4.4 Screenshot of the SELFIS tool

4.2.3 Time Headway

The second part of the analysis is focused in further investigation in the time headway of the traffic. It is known that if the inter-arrival times (time headways) follows an exponential distribution, then the number of arrivals (vehicle arrival patterns) is a Poisson process. The hypothesis test for exponential distribution is as followed:

H_0: The time headway follows Exponential distribution

H_1: The time headway does not follow Exponential distribution

The goodness-of-fit test used is Anderson-Darling test. Anderson-Darling test is useful in testing whether a dataset follows a specific distribution. By using Minitab, a statistical software, the p-value of the hypothesis test is obtained. For this study, since the confidence interval is also set to be within 95%, the value of level of significance is 0.05.

Afterwards, this research employs LLCD plots to investigate whether the time headway distribution is heavy-tailed or not. For example, the time headway for S1C1 of Lane 1 on 20 August is being investigated. By using the LLCD plot method as discussed in Section 3.2, and by using the least square regression method, it is observed that the distribution has the slope of -1.62 with R^2= 0.9442, which estimates the α value to be 1.62. Figure 4.5 illustrates the LLCD plot obtained from the time headway.

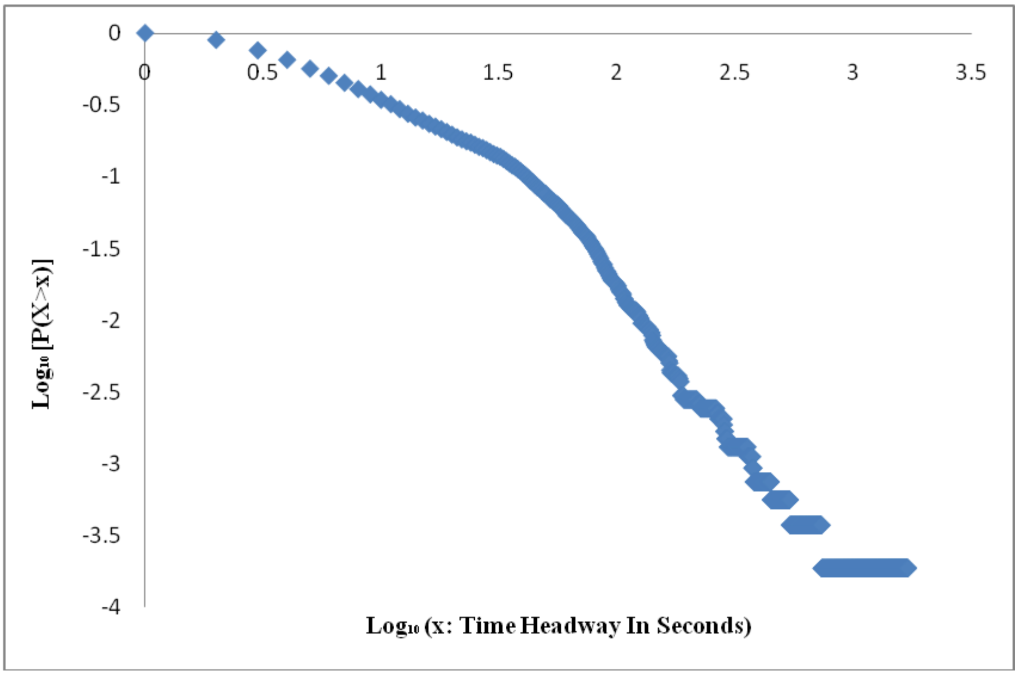

Figure 4.5 LLCD plot for time headway for S1C1 Lane 1 on 20 August.

Note that the least square is fitted to evenly spaced data points that are greater than 1 second.

Then, the same time headway distribution is fitted to different distributions such as Weibull, Exponential, Pareto and Lognormal distribution. Goodness-of-fit statistics such as Anderson-Darling (AD), Kolmogorov-Smirnov (KS) and Cramer-von Mises (CM) statistics are used. In this research

, Aikake's Information Criterion (AIC) and Bayesian Information Criterion (BIC) are also used to justify the results obtained from the goodness-of-fit statistics.

4.3 Incorporation of Hurst Parameter in Average Delay

To compute the average delay per vehicle that is based on self-similarity, some underlying assumptions are deduced in this research:

1. The computation of average delay is based on first in, first out (FIFO) method
2. One approach is considered for this research. However, even if we have only one approach, we still have to make assumption for green duration and red duration. This is because if we lack of this assumption, the least delay case for 1 approach will always be the case when green duration is maximum and red phase is minimum(i.e. 0 seconds)
3. To construct the optimal duration of green and red phase, a two-phase intersection scenario is illustrated.
4. The flow for the intersection is the same, so that it is appropriate to assume that the effective green duration is the same as the red duration, because if the flow for both phases is the same, the duration of green and red will be approximately the same as well, in order for both phases to clear vehicle queue
5. Each vehicle will take 2 seconds to drive across the intersection
6. One hour traffic is considered for each case
7. The cycle time is calculated by

$$c = r + g + a + l \tag{17}$$

where

c= cycle time

r= red light duration

a= amber light duration

g= green light duration

l= start-up lost time

8. The amber duration for each cycle is fixed at 3 seconds
9. The minimum duration for red light or green light is 2 seconds.
10. No start-up loss time and no all red-period (i.e. l= 0 s)

Since we own the actual timestamp for vehicle arrival, the vehicle arrival pattern obtained from actual data is used to compute the average delay for each vehicle. In this computation, a total of 894 sets of one-hour data obtained from the vehicle arrival pattern is used. These data are categorized based on traffic flow.

Based on the assumptions mentioned above, the hourly vehicle arrival pattern is analyzed. Since the timestamp of each vehicle is obtained from the data, each vehicle can be assigned to be in red or green duration based on the allocation of red and green light duration from equation (8). Different combination of red and green light duration is used in the computation and the amber light duration is fixed at 3 seconds. For example, a cycle time of 8 seconds can have 2 combinations, which is (1) 2 seconds of red light duration and 3 seconds of green light duration, and (2) 3 seconds of green light duration and 2 seconds of red light duration.

The total delay experienced by each vehicle is computed based on the assumptions above. For vehicles that did not queue up to go across the intersection experienced no delay when they were assigned in green duration, while for vehicle experienced delay when they were assigned in red duration. Vehicles that required to queuing up in order to go across the intersection experienced more delay, as their delays were affected by the queue up process. The delay experienced by each vehicle is computed. Then, the average delay per vehicle is obtained by averaging the total delay with the total number of vehicles.

It is observed that different combination of green and red duration yields different average delay. However, the combination of green and red duration that yields the least delay would be the optimal cycle length, and its average delay would be the minimum delay. The minimum average delay obtained from the vehicle arrival pattern is called self-similar delay, as it is the delay obtained from the vehicle arrival pattern that possesses self-similarity characteristics. The whole process for the computation is explained in following.

4.3.1 Computation of Average Delay

The computation of average delay per vehicle by inputting the timestamp of each vehicle requires certain sets of rules to be followed.

Since we are investigating the hourly time headway, 3600 timestamps are generated. Each timestamp will be assigned to either red phase, amber phase or green phase, depending on the allocation of green duration and red duration. One full cycle is consisted as 1 green phase, 1 amber phase and 1 red phase. Hence, each timestamp will be assigned to its corresponding cycle. In this stage, we have obtained the initial phase and cycle of each timestamp. If the vehicle is in red phase under the same cycle, the number of vehicles in that phase is sum up to obtain the total number of vehicles in queue for that cycle. Afterwards, the initial discharge timestamp for the vehicles are computed, based on the number of vehicles in queue. Since we have set the discharge time to be 2 seconds for each vehicle and we assume there is no start-up loss time, the discharge time for the vehicle is an increment of 2 seconds from its previous vehicle. Therefore, the initial discharge timestamp is the exact time where the vehicle starts to discharge. For green phase, the arrival timestamp of the vehicle in green phase is considered as the initial discharge timestamp, since these vehicles can travel across the intersection without queuing in front of the traffic signal.

At this stage, all the initial discharge timestamp to discharge for each vehicle are obtained. The initial discharge timestamp is analyzed to find out if the discharge time for each vehicle is reasonable. The possible scenario to adjust the initial discharge time is as followed.

1. **The difference of initial discharge timestamp of two vehicles is lesser than 2 seconds.**

   In this scenario, this is possible because the initial discharge timestamp for vehicle in red phase is conflicting with the initial discharge timestamp for vehicle in green phase under the same cycle. 2 seconds will be added to the second vehicle. The adjusted initial discharge timestamp will be analyzed again. If this scenario persisted, the adjustment process is repeated.

2. **The initial discharge timestamp of second vehicle is earlier than the initial discharge timestamp of first vehicle.**

   This scenario indicates that the vehicle in green phase is not travelling across the intersection without stopping. In fact, this vehicle has to queue up to wait for the previous vehicle to travel across the intersection. Hence, this vehicle is added into the queue and the initial discharge timestamp will be analyzed again. If this scenario persisted, the adjustment process is repeated.

After the adjustment process is completed and the initial discharge timestamp for each vehicle is free of error, the initial discharge time is compared with the generated timestamp to determine its corresponding phase and cycle. If the initial discharge timestamp is in green phase, it indicates that the vehicle can travel across the intersection. If the initial discharge timestamp is in red phase, it indicates that the vehicle is queuing up and only can discharge in the green phase for next cycle. This process is repeated until all the vehicles are successfully discharged.

The waiting duration for the vehicle to discharge is the delay (d^w) of the vehicle and the total delay of the vehicle (D) is the summation of the delay for each vehicle, and is represented by

$$D = \sum_{n=1}^{n} d_n^w \tag{18}$$

The average delay of the vehicle (d) is the total delay of the vehicle (D) over the total arrival of the vehicles (a) in the investigation period, and is represented by

$$d = \frac{D}{a} \tag{19}$$

At this stage, the average delay is obtained for 1 phase of the cycle. To complete the computation, it is required to have a minimum of 2 phases. If the 1st phase starts with a green duration, then the 2nd phase must starts with a red duration. In other word, green duration in 1st phase is equal to the red duration in 2nd phase, and vice versa. The computation of average delay is repeated for the 2nd phase. For simplicity, a two-phase intersection is considered in this paper.

4.3.2 Computation of Optimal Cycle Length and Least Average Delay

In the previous section, the computation to find the average delay is explained. However, with the fixed duration of red phase and green phase, it is impossible to deduce that this fixed duration is the optimal duration. Hence, the process needs to be carried out for different combination of green and red duration.

The amber phase is fixed at 3 seconds and the duration for vehicle to discharge is fixed at 2 seconds. Since it is also assumed that the vehicle is able to travel across the intersection during the amber phase, the minimum duration of green phase is 5 seconds. However, since the vehicle count is in discrete value, it is better to set the minimum duration of green phase to be 6 seconds, as it allows 3 vehicles to travel across the intersection. As mentioned earlier, the green duration for 1st phase is the red duration for 2nd phase, and vice versa. Hence, the minimum duration of red phase is set to be 6 seconds.

Different combination of red duration and green duration are used to compute the average delay. The combination of red duration and green duration that provides the least average delay is the optimal red duration and green duration. The summation of optimal green duration and optimal red duration is the optimal cycle length.

Table 4.3 Computation of average delay under different combination of red duration and green duration

		Green Duration												
		6 Sec	7 Sec	8 Sec	9 Sec	10 Sec	11 Sec	12 Sec	13 Sec	14 Sec	15 Sec	16 Sec	17 Sec	18 Sec
Red Duration	6 Sec	6.085	14.252	5.637	18.720	63.003	7.704	64.373	9.427	11.310	25.196	67.780	67.833	82.262
	7 Sec	14.498	**4.722**	5.676	5.830	6.165	61.579	7.524	8.432	8.730	64.998	65.351	11.152	24.740
	8 Sec	5.734	5.974	5.325	5.955	60.942	5.964	7.205	7.537	7.442	63.810	9.550	9.940	66.616
	9 Sec	62.086	5.322	6.048	5.371	60.696	6.054	6.345	5.915	61.959	7.983	8.013	9.215	9.927
	10 Sec	61.925	6.360	5.846	60.594	5.350	5.647	5.411	5.915	6.508	6.777	7.524	9.276	8.654
	11 Sec	8.191	6.721	61.502	5.331	5.379	5.613	5.736	6.120	5.999	6.663	6.654	7.452	7.930
	12 Sec	9.822	62.614	6.145	5.639	5.759	5.892	5.939	5.694	6.268	7.003	6.782	7.059	7.634
	13 Sec	76.092	7.570	6.478	6.286	6.028	6.033	5.853	6.084	6.078	6.484	6.422	6.894	6.916
	14 Sec	9.998	63.552	19.015	6.687	6.417	6.129	6.207	6.478	6.052	6.492	6.709	6.764	7.238
	15 Sec	66.376	21.557	20.319	7.393	6.712	6.633	7.123	6.282	6.477	6.750	6.500	6.838	6.867
	16 Sec	79.516	22.804	10.193	7.820	7.406	7.647	6.829	6.715	6.884	6.605	6.666	6.767	7.474
	17 Sec	25.701	11.945	64.857	9.014	9.466	7.775	7.249	7.182	6.698	6.627	6.993	7.096	7.494
	18	70.	66.	11.	10.	8.6	8.2	7.9	7.1	6.5	7.1	7.1	7.6	7.2

	Sec	154	887	377	817	36	95	31	66	50	32	33	58	58
	19 Sec	70.490	14.261	13.974	10.438	9.837	8.508	23.454	7.118	7.434	7.183	7.623	7.318	7.797
	20 Sec	71.560	15.065	12.745	11.446	9.846	24.919	7.908	7.998	7.758	7.823	7.392	7.752	7.445

For instance, from Table 4.3, it can be seen that for this hourly time headway, the optimal green duration and red duration is 7 seconds. This is because the average delay for this combination, which is 4.722 seconds per vehicle, is the least average delay per vehicle obtained. Hence, the optimal cycle length for this hourly time headway is 14 seconds.

4.3.3 Analysis of Computation Results

For the average delay computation, Hurst parameter is used as the variable. Hourly traffic for each lane of the intersection is analyzed by using SELFIS to obtain the Hurst parameter. In this research, the Hurst parameter obtained from Whittle estimator is used.

In order to prove that the computed delay is similar to the actual delay, a 2-sample t-test is performed on 30 sets of hourly time headway data. The actual average delay per vehicle is obtained by aggregating the duration used by each vehicle when going across the intersection. This actual average delay per vehicle is then compared with the average delay obtained from the vehicle arrival pattern. The hypothesis testing for 2-sample t-test is as followed:

$H_0: \mu_0 = \mu_1$

$H_1:: \mu_0 \neq \mu_1$

where μ_0 is the mean for actual delay and μ_1 is the mean for computed delay. By using Minitab, a statistical software, the p-value of the hypothesis testing is obtained. For this research, since the confidence interval is also set to be within 95%, the value of level of significance is 0.05.

The optimal cycle length obtained from the computation is assumed to be the least delay obtained from combination of red and green light duration from equation (17). This optimal cycle length is compared with the Webster's optimal cycle length in equation (15). On the other hand, the least delay obtained from the combination of red and green light duration is compared with the Webster's average delay per vehicle formula in equation (16). Webster's average delay per vehicle is compared with the computed delay by the Hurst parameter found in each dataset. Scatter plot test is also performed on the computed delay based on the Hurst parameter to study their correlation.

CHAPTER 5

RESULTS AND DISCUSSION

5.1 Properties of Isolated Intersection

By inputting the entire vehicle arrival pattern into Minitab, hypothesis testing on whether the vehicle arrival pattern follows Poisson distribution or not is found out. From the results of the data, it can be observed that for all days, all lanes for both of the intersections have p-value of 0.0000, which are less than 0.05. Hence, it is concluded that the null hypothesis is not accepted. The alternative hypothesis is accepted, which proved that the vehicle arrival pattern does not follow Poisson distribution.

Since it is clear that Poisson distribution failed to model the vehicle arrival pattern of the empirical data, SELFIS tool is employed to investigate the existence of self-similarity characteristics in the empirical data. By employing the graphical method, time series graphs are plotted with respect to different timescales. Figure 5.1 to Figure 5.6 illustrates the times series graphs with timescales ranging from 1 minute, 2 minutes, 5 minutes, 10 minutes, 30 minutes to 60 minutes interval for S1C1 Lane 1 on 19 August 2015. Vehicle volumes are being aggregated according to the timescales. From Figure 6 to Figure 11, it can be seen clearly that the distribution of vehicle arrival for different timescales have very similar patterns.

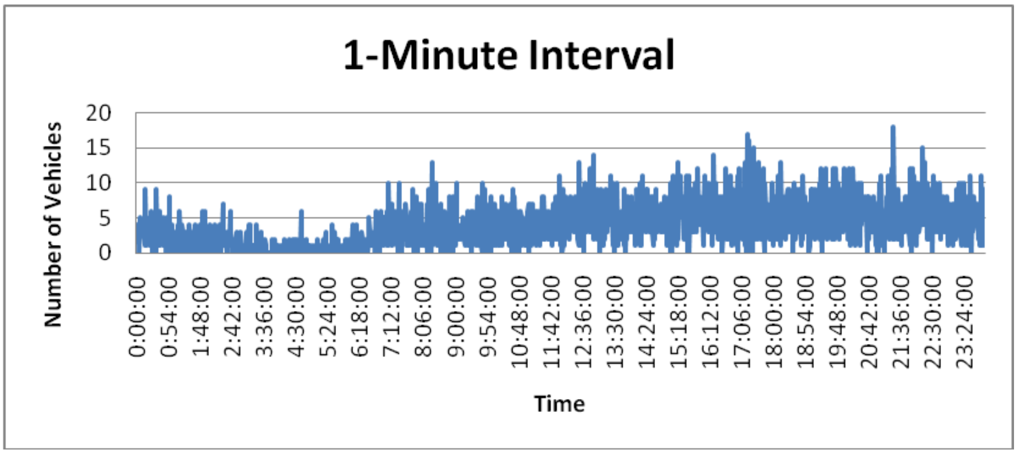

Figure 5.1 The vehicle arrival pattern for 1-minute interval

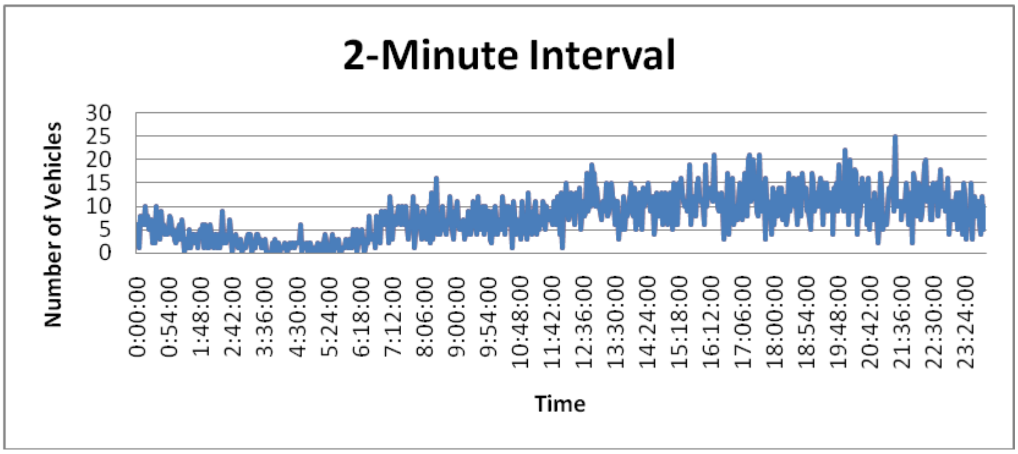

Figure 5.2 The vehicle arrival pattern for 2-minute interval

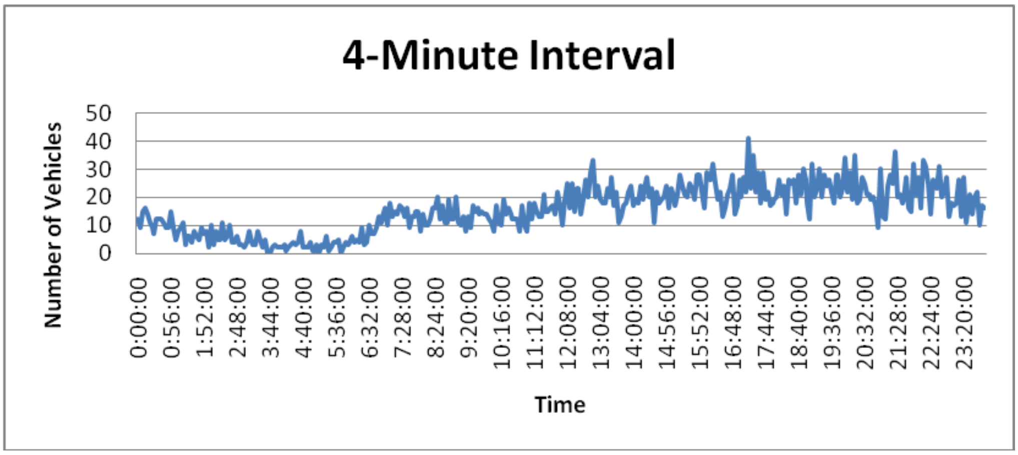

Figure 5.3 The vehicle arrival pattern for 4-minute interval

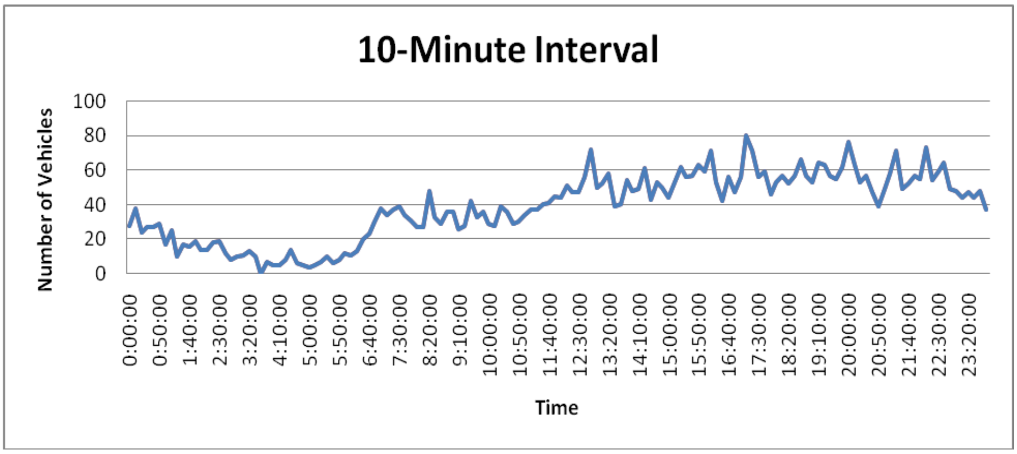

Figure 5.4 The vehicle arrival pattern for 10-minute interval

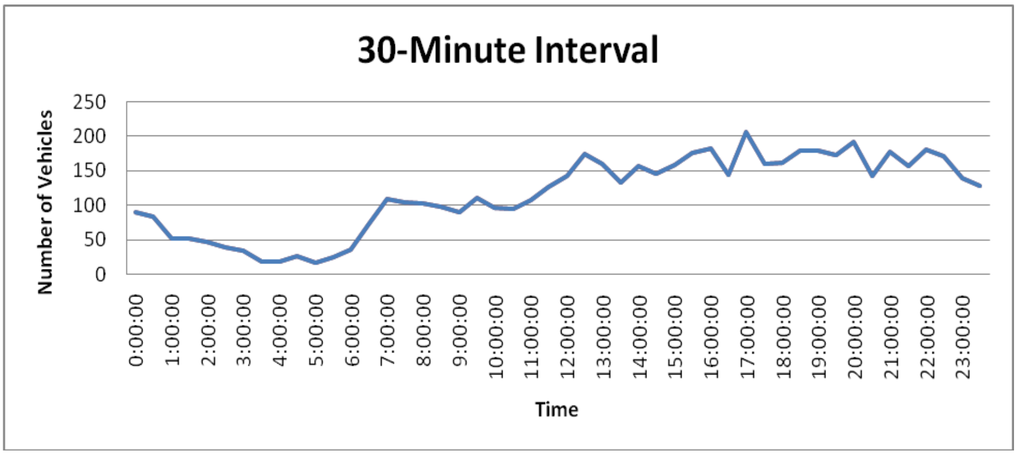

Figure 5.5 The vehicle arrival pattern for 30-minute interval

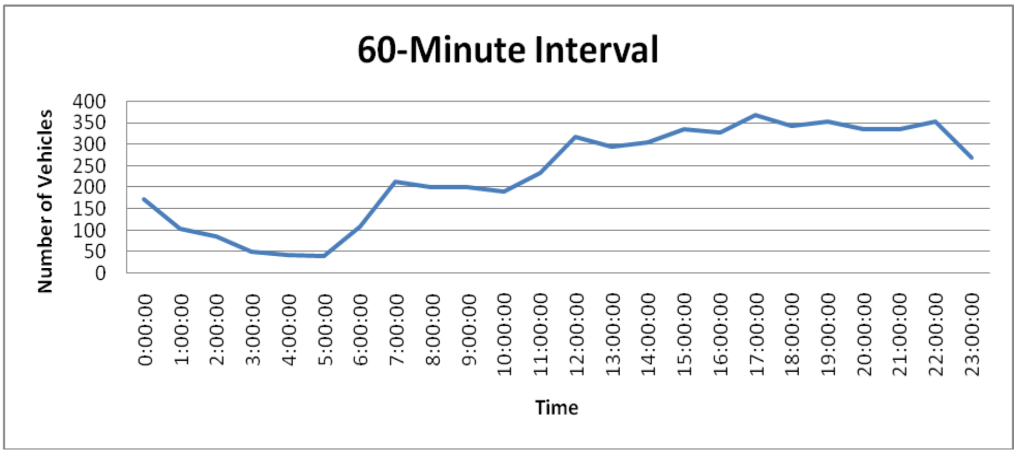

Figure 5.6 The vehicle arrival pattern for 60-minute interval

Then, the same dataset is being input into SELFIS tool to investigate the value of Hurst parameter of the vehicle arrival pattern. Table 5.1 shows the average Hurst value for each lane of the intersections. From Table 5.1, it can be seen that for the first intersection, the average Hurst value estimated by Whittle estimator is more than 0.95. From Table 5.1, the average Hurst value estimated by Whittle estimator for the second intersection is more than 0.75. This shows that the first intersection exhibit a higher degreeness of self-similarity as compared to the second intersection, which results in higher burstiness for the vehicle arrival pattern in first intersection.

Table 5.1 Whittle estimator obtained from 1st intersection

S1C1			S1C2			S1C3			S1C4		
Date	Lane	Whittle	Date	Lane	Whittle	Date	Lane	Whittle	Date	Lane	Whittle
	1	0.985		1	0.968		1	0.986		1	0.989
	2	0.966	13	2	0.962		2	0.980	13	2	0.99
14	3	0.986		1	0.962	13	3	0.982		1	0.989
	1	0.948	14	2	0.979		1	0.984	14	2	0.99
	2	0.896		1	0.971		2	0.966		1	0.978
15	3	0.968	15	2	0.971	14	3	0.968	15	2	0.982
	1	0.985		1	0.951		1	0.986		1	0.99
	2	0.986	16	2	0.959		2	0.980	17	2	0.991
16	3	0.978		1	0.96	15	3	0.982		1	0.986
	1	0.981	17	2	0.959		1	0.919	18	2	0.988
	2	0.985		1	0.979		2	0.913		1	0.979
17	3	0.983	18	2	0.975	16	3	0.954	19	2	0.979
	1	0.974		1	0.968		1	0.937		1	0.985
	2	0.982	19	2	0.943		2	0.946	20	2	0.965
18	3	0.985		1	0.955	17	3	0.966		1	0.987
	1	0.983	20	2	0.952		1	0.961	21	2	0.976
	2	0.990		1	0.936		2	0.973		1	0.987
19	3	0.986	21	2	0.938	18	3	0.980	22	2	0.988
	1	0.984		1	0.942		1	0.963		1	0.981
	2	0.977	22	2	0.962		2	0.974	23	2	0.984
20	3	0.978				19	3	0.981			
	1	0.986					1	0.978			
	2	0.991					2	0.982			
21	3	0.986				20	3	0.986			
	1	0.980					1	0.969			
	2	0.982					2	0.967			
22	3	0.983				21	3	0.986			
	1	0.948					1	0.944			
	2	0.973					2	0.941			
23	3	0.926				22	3	0.943			
	Average	0.975		Average	0.9596		Average	0.966		Average	0.9842

Table 5.2 Whittle estimator obtained from 2nd intersection

S2C1			S2C2			S2C3			S2C4		
Date	Lane	Value	Date	Lane	Value	Date	Lane	Value	Date	Lane	Value
13	1	0.654	13	1	0.845	13	1	0.633	13	1	0.731
	2	0.535		2	0.804		2	0.675		2	0.669
14	1	0.757	14	1	0.764	14	1	0.858	14	1	0.825
	2	0.773		2	0.786		2	0.877		2	0.843
15	1	0.787	15	1	0.779	15	1	0.765	15	1	0.704
	2	0.912		2	0.762		2	0.772		2	0.747
17	1	0.714	16	1	0.726	16	1	0.942	16	1	0.783
	2	0.802		2	0.804		2	0.861		2	0.897
18	1	0.625	18	1	0.94	17	1	0.699	17	1	0.648
	2	0.715		2	0.939		2	0.768		2	0.699
19	1	0.657	19	1	0.681	18	1	0.785	18	1	0.661
	2	0.702		2	0.7		2	0.795		2	0.704
21	1	0.723	20	1	0.8	19	1	0.678	19	1	0.778
	2	0.634		2	0.776		2	0.682		2	0.87
22	1	0.892	22	1	0.856	20	1	0.748	20	1	0.718
	2	0.741		2	0.82		2	0.528		2	0.744
25	1	0.774	23	1	0.817	21	1	0.68	21	1	0.842
	2	0.721		2	0.783		2	0.794		2	0.747
26	1	0.71	24	1	0.774	22	1	0.664	22	1	0.843
	2	0.82		2	0.772		2	0.725		2	0.909
	Average	0.7324		Average	0.7964		Average	0.74645		Average	0.7681

On the other hand, hypothesis test is also performed on the time headway of the vehicles. The time headway datasets are inputted in Minitab to perform the hypothesis test. The result showed that for all days, all lanes for both of the intersections have p-value less than 0.003, which is less than 0.05.

Hence, it is concluded that the null hypothesis is not accepted. The alternative hypothesis is accepted, which proved that the time headway for both intersections does not follow exponential distribution.

Since it is proved that the time headway distributions do not follow exponential distribution, this thesis employs LLCD plot to investigate whether the time headway distribution is heavy-tailed or not. This method is applied onto all lanes of the intersections for each day. From the observation, the average α value obtained for the first intersection is 1.9976. Since α is less than 2, it can be concluded that the time headway for the first intersection possesses infinite variance and is indeed heavy-tailed. However, for the second intersection, the average α value is 2.2146. This results show that the time headway for second intersection has a much lighter tail weight as compared to the first intersection. This is due to the facts that the degreeness of self-similarity for second intersection is much lower than the first intersection.

Motivated by the finding that the time headway distributions are indeed heavy-tailed, a few types of heavy-tailed distributions are fitted into the time headway to find out the more appropriate distribution that can be used in describing the distribution of time headway. For all days on all intersections, it is observed that Lognormal distribution has the lowest goodness-of-fit statistics for AD, KS and CM statistics. Table 5.3 shows the results obtained from one approach of the intersection for all the investigation period. It is also observed that Lognormal distribution produces the lowest AIC and BIC when compared with other distribution.

Table 5.3 Statistical results for different types of distributions

14	KS Test	CVM Test	AD Test	Chi square test	AIC	BIC
Weibull	0.123956	35.57175	214.8604	2610.418	44089.58	44103.02
Exponential	0.182732	75.96937	375.4905	4496.017	44584.49	44591.21
Lognormal	**0.082871**	**13.12171**	**78.73287**	**917.8424**	**42170.99**	**42184.43**
Pareto	0.122534	19.2566	142.4021	1616.452	43229.99	43243.42
15	KS Test	CVM Test	AD Test	Chi square test	AIC	BIC
Weibull	0.141396	17.66726	115.3212	7.12E+13	34141.95	34154.67
Exponential	0.134972	22.3125	131.9844	7.42E+13	34170.11	34176.47
Lognormal	**0.121499**	**9.919451**	**63.24403**	**5.40E+13**	**33112.36**	**33125.07**
Pareto	0.156104	12.03046	94.27237	6.35E+13	33909.71	33922.42
16	KS Test	CVM Test	AD Test	Chi square test	AIC	BIC
Weibull	0.132484	40.82056	236.8884	2814.634	41020.06	41033.31
Exponential	0.222303	103.6483	506.1961	4225.157	41758.76	41765.39
Lognormal	**0.088619**	**16.8283**	**102.9051**	**1283.119**	**39230.34**	**39243.59**
Pareto	0.125533	20.61744	147.4143	1817.11	40118.45	40131.7
17	KS Test	CVM Test	AD Test	Chi square test	AIC	BIC
Weibull	0.128853	31.97371	194.5481	2421.191	41698.3	41711.58
Exponential	0.181088	68.8044	341.1681	3432.563	42149.37	42156.01
Lognormal	**0.072386**	**11.29141**	**70.39718**	**959.3304**	**39944.41**	**39957.69**
Pareto	0.126196	16.96806	128.0455	1596.208	40917.93	40931.22
18	KS Test	CVM Test	AD Test	Chi square test	AIC	BIC
Weibull	0.133477	31.52583	189.7272	4928.829	39552.69	39565.84
Exponential	0.200536	81.8259	-	6577.78	40204.78	40211.36
Lognormal	**0.082494**	**12.91894**	**78.88396**	**2810.405**	**37892.09**	**37905.24**
Pareto	0.126257	15.98654	119.9192	3643.356	38731.65	38744.8
19	KS Test	CVM Test	AD Test	Chi square test	AIC	BIC
Weibull	0.126587	32.44908	193.7703	5229.233	41140.25	41153.5
Exponential	0.20007	79.91687	-	6975.516	41716.4	41723.03
Lognormal	**0.083762**	**12.61217**	**76.5925**	**3119.019**	**39477.76**	**39491.01**
Pareto	0.121567	16.27262	121.7189	3923.729	40359.19	40372.44
20	KS Test	CVM Test	AD Test	Chi square test	AIC	BIC
Weibull	0.136649	31.4386	-	3814.221	39927.56	39940.74
Exponential	0.201874	83.47444	-	6108.845	40654.7	40661.29
Lognormal	**0.077584**	**11.25223**	**70.1397**	**2009.499**	**38163.97**	**38177.16**
Pareto	0.126885	15.5909	118.6155	2695.316	39024.1	39037.28
21	KS Test	CVM Test	AD Test	Chi square test	AIC	BIC
Weibull	0.129717	35.07509	211.9083	2522.603	42696.93	42710.28

Exponential	0.192408	83.69241	412.9262	4034.717	43302.03	43308.71
Lognormal	**0.078974**	**12.9748**	**80.98145**	**997.5239**	**40855.11**	**40868.47**
Pareto	0.125738	17.82836	134.4598	1613.633	41821.5	41834.86
22	KS Test	CVM Test	AD Test	Chi square test	AIC	BIC
Weibull	0.119998	32.7815	-	2267.688	41967.04	41980.34
Exponential	0.191217	75.97065	-	3145.461	42504.47	42511.12
Lognormal	**0.077766**	**11.56241**	**70.93074**	**896.7342**	**40238.56**	**40251.86**
Pareto	0.114784	16.84987	125.3126	1491.856	41174.19	41187.5
23	KS Test	CVM Test	AD Test	Chi square test	AIC	BIC
Weibull	0.129145	33.71718	197.15	2422.717	38925.32	38938.42
Exponential	0.209074	84.52605	415.0632	3591.065	39500.57	39507.12
Lognormal	**0.079592**	**12.89401**	**80.36597**	**1119.165**	**37409.31**	**37422.42**
Pareto	0.117851	16.30726	120.1037	1623.014	38204.18	38217.28

This result is significant for time headway modeling, as lognormal distribution has proven to be applicable for time headway modeling. Since this thesis aims at identifying the appropriate distribution that can be used for time headway modeling, more study should be conducted to build a time headway model based on lognormal distribution.

5.2 Computation of Average Delay

By employing hypothesis test, it is observed that for all the 30 sets of hourly time headway data, the p-value is more than the significant level, which is 0.05. The result is showed in Table 5.4. Therefore, the alternative hypothesis is rejected and it can be concluded that there is no significant difference between the mean for actual delay and the mean for computed self-similar delay. This is significant in affirming that the average delay obtained from our calculation is similar to the actual average delay.

Table 5.4 P-value for 30 sets of time headway data

No	Intersection	Day	Time	DF	P-Value	Conclusion
1	S1C3	14	2:00:00 AM	36	0.0566	Accept H_0
2	S1C1	18	4:00:00 AM	38	0.05068	Accept H_0
3	S1C1	19	3:00:00 AM	32	0.05762	Accept H_0
4	S1C1	19	6:00:00 AM	37	0.07182	Accept H_0
5	S1C2	15	4:00:00 AM	34	0.05532	Accept H_0
6	S1C2	19	6:00:00 AM	43	0.06331	Accept H_0
7	S1C2	20	7:00:00 AM	49	0.05152	Accept H_0
8	S1C3	18	5:00:00 AM	22	0.06378	Accept H_0
9	S1C3	21	2:00:00 AM	21	0.0612	Accept H_0
10	S1C3	22	3:00:00 AM	23	0.06299	Accept H_0
11	S1C3	13	2:00:00 AM	35	0.05153	Accept H_0
12	S1C4	20	1:00:00 AM	39	0.05444	Accept H_0
13	S2C1	21	5:00:00 AM	40	0.05133	Accept H_0
14	S2C1	22	3:00:00 AM	32	0.05372	Accept H_0
15	S2C1	25	5:00:00 AM	56	0.05934	Accept H_0
16	S2C1	26	1:00:00 AM	57	0.05017	Accept H_0
17	S2C2	18	2:00:00 AM	45	0.05332	Accept H_0
18	S2C2	20	3:00:00 AM	23	0.05723	Accept H_0
19	S2C2	22	4:00:00 AM	34	0.0542	Accept H_0
20	S2C2	13	6:00:00 AM	30	0.06821	Accept H_0
21	S2C2	14	5:00:00 AM	12	0.06782	Accept H_0
22	S2C3	17	3:00:00 AM	27	0.06911	Accept H_0
23	S2C3	18	2:00:00 AM	19	0.05921	Accept H_0
24	S2C3	21	3:00:00 AM	32	0.05988	Accept H_0
25	S2C3	22	5:00:00 AM	41	0.00551	Accept H_0
26	S2C4	14	4:00:00 AM	32	0.05017	Accept H_0

27	S2C4	15	6:00:00 AM	41	0.05762	Accept H_0
28	S2C4	18	3:00:00 AM	22	0.053	Accept H_0
29	S2C4	19	2:00:00 AM	23	0.05488	Accept H_0
30	S2C4	22	6:00:00 AM	33	0.05671	Accept H_0

The computation of average delay and optimal cycle length is performed. Then by using the same set of data, the average delay and optimal cycle length is calculated by using Webster's formulae. The results can be found in Appedix I. The results are then analyzed and compared.

A scatter plot of average delay obtained from computation per vehicle and Hurst parameter is plotted in Figure 5.7

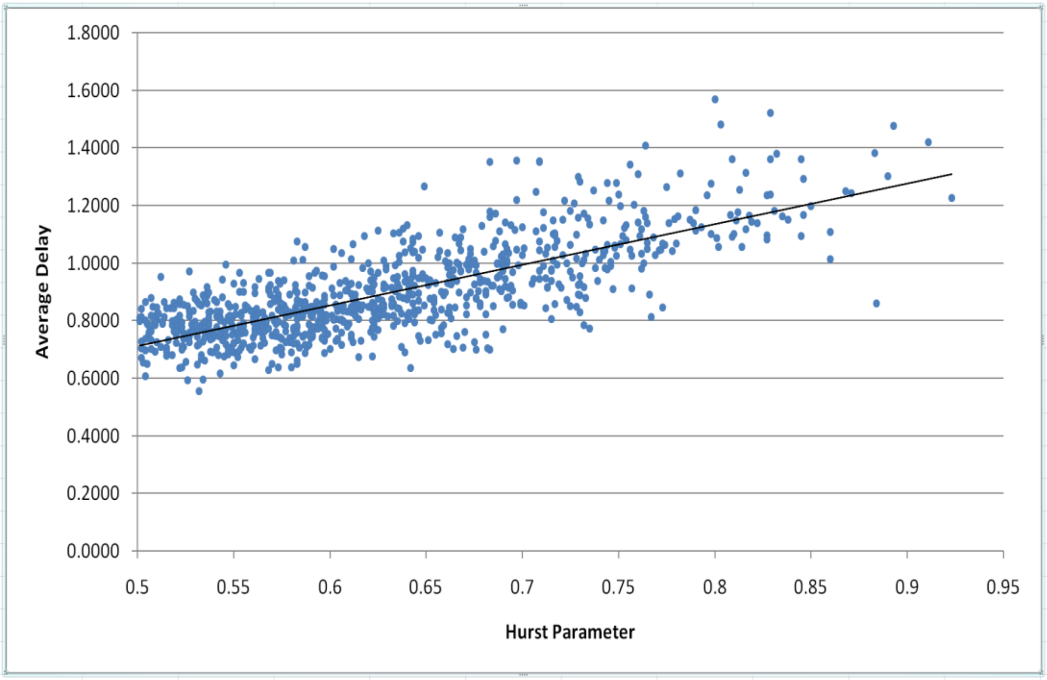

Figure 5.7 Scatter plot of the computed average delay and Hurst parameter

From Figure 5.7, it can be seen that there is a positive correlation between the average delay per vehicle and the Hurst parameter. The correlation coefficient obtained from this scatter plot is 0.76. This indicates that the relationship between the average delay and Hurst parameter has a strong positive correlation. Hence, higher value of Hurst parameter indicates a higher average delay per vehicle.

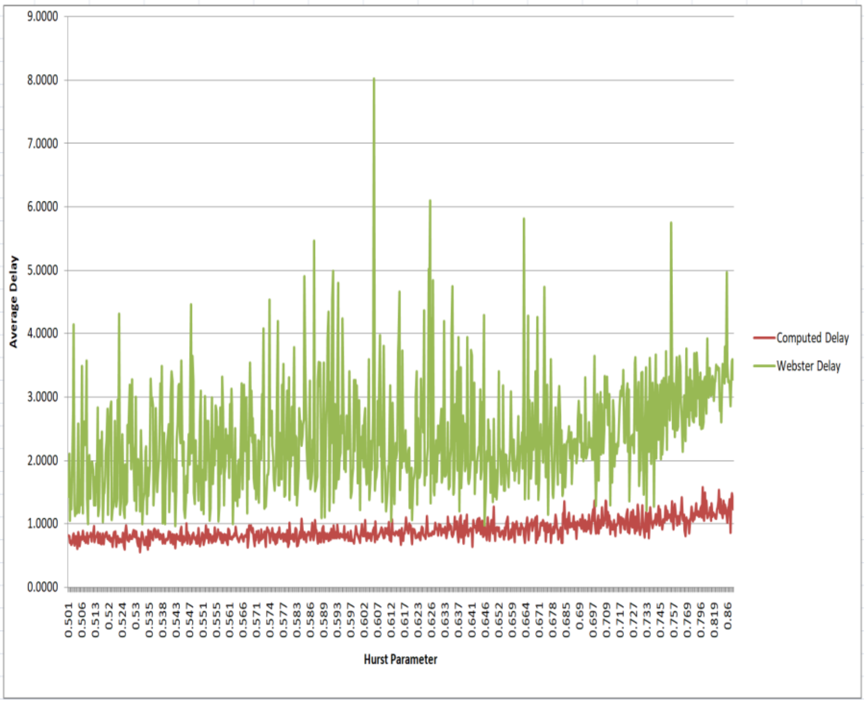

Figure 5.8 Comparison between Webster's delay and computed delay

Figure 5.8 shows a line graph of average delay obtained from Webster's formula in equation (15) and average delay obtained from computation, which are plotted against Hurst parameter of each dataset. The average Webster delay was calculated and the result was categorized based on the Hurst parameter of the dataset.

Hurst parameter is assigned to the Webster's delay as well, since the same dataset is being used in the computation of average delay and Webster's delay formula. A scatter plot for average delay obtained from Webster's delay formula and Hurst parameter is plotted in Figure 5.9

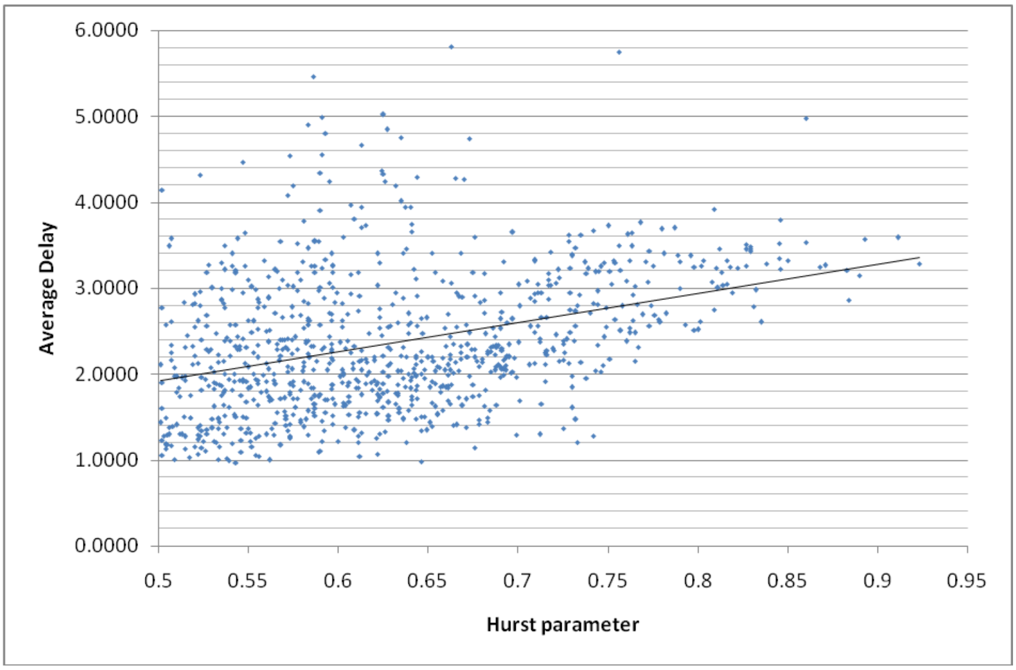

Figure 5.9 Scatter plot of the Webster's average delay and Hurst parameter

From Figure 5.9, it can be seen that there is no relationship between the Webster's average delay and the Hurst parameter. The correlation coefficient obtained from this scatter plot is 0.13. This indicates that there is no close relationship between the average delay obtained by Webster's delay formula and Hurst parameter. This also explains the high fluctuation of average delay for Webster's method in Figure 5.8

After observing the comparison of average delay, it should be noted that the optimal cycle length should be observed, as it provides the duration setting of red phase and green phase in traffic signal design. Figure 5.10 shows a line graph of the computed cycle length and the Webster's optimal cycle length, which are plotted against Hurst parameter of each datasets.

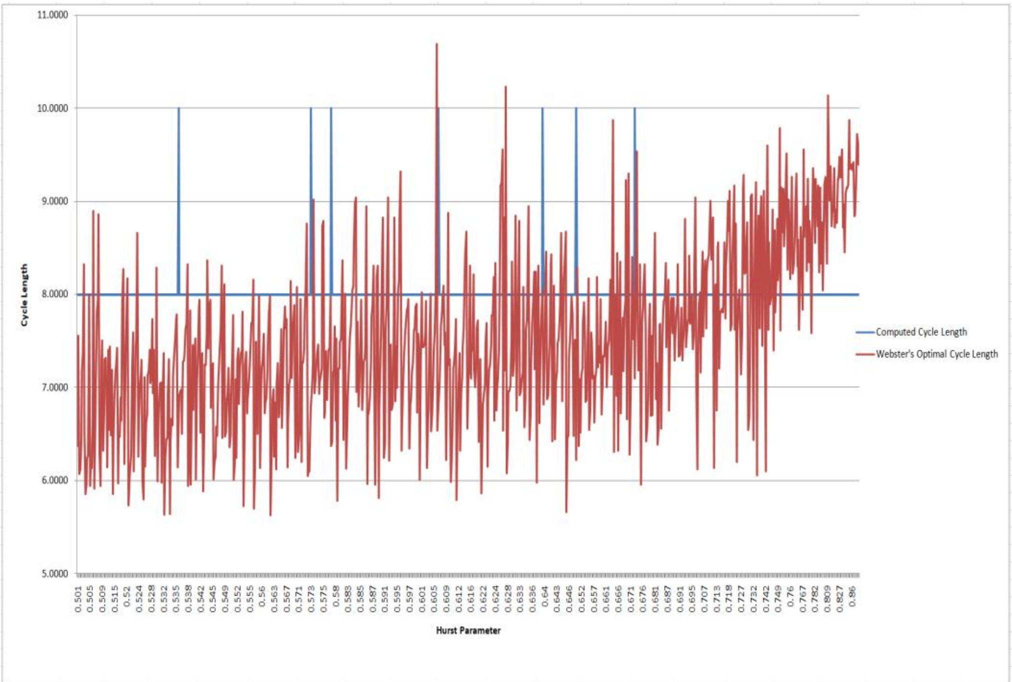

Figure 5.10 Comparison of average delay between Webster's delay and computed delay

There are a few observations that can be deduced from Figure 5.8 and Figure 5.10. First, from Figure 5.8, it can be seen that for all sets of one-hour data used, the minimum average delay of self-similar computation is always lower than the Webster's delay. Next, from Figure 5.10, the optimal cycle length that yields the self-similar delay stays at 8 seconds for most of the datasets, while the optimal cycle length that yields the Webster's delay is ranging from 5 seconds to 10 seconds based on different traffic flow.

From these observations, two major implications are drawn. Firstly, since the minimum average delay of self-similar computation yields a lower results as compared to Webster's delay under all types of traffic flow observed in this research, it is shown that the incorporation of the Hurst parameter in finding the optimum cycle length and the least average delay per vehicle provides a more effective method in reducing the vehicle delay.

Secondly, the optimum cycle length obtained from combination of green and red duration that yield the least delay is more consistent than the optimum cycle length obtained from Webster's method. This is because when we are looking in hourly traffic, Webster's method is only emphasized on the flow rate and not considering the vehicle arrival behavior, which exhibit self-similar characteristics. Hence, the optimum cycle length and the average delay per vehicle for Webster's method tend to provide a larger value, which might not be that effectively in optimizing the delay of the vehicles.

Since the vehicle arrival pattern of the datasets exhibit self-similar characteristics, which showed patterns that are similar across scales, the average delay obtained from computation provides more consistent results as compared to Webster's method. It is also notable that the average delay per vehicle obtained from Webster's method is not consistent. The computed results shows a relatively consistent average delay, which indicates that it is more appropriate to incorporate self-similarity in the calculation of average delay.

CHAPTER 6

CONCLUSION

For this research, it is observed that the vehicle arrival pattern of the isolated signalized intersections in Kuala Lumpur exhibit the characteristics of self-similarity. The convention method of using Poisson assumption to model vehicle arrival pattern no longer holds. Whereas, the corresponding time headways follow heavy-tailed distribution, in which Lognormal is appropriate in modeling such headways.

In traffic engineering, delay is an important factor in traffic signal design. Many researches are contributed in finding the optimal cycle length that minimizes the average delay of the vehicle. However, the actual average delay is very difficult to obtain. Webster has deduced formulation of optimal cycle length and minimum average delay based on Poisson assumption. Since this research has shown that Poisson assumption is not appropriate in modeling the vehicle arrival pattern, Webster's formulation might not be the best method to be used in traffic signal design.

From the results, it can be shown that the computation of optimal cycle length and minimum average delay by using vehicle arrival pattern of different Hurst parameter gives a lower average delay and a consistent duration of optimal cycle length. Furthermore, it is shown that higher value of self-similarity will lead to longer queuing delays. Hence, maintaining the traffic flow in a lower Hurst value might be a solution to reduce the delay of traffic.

However, this computation is a preliminary investigation in self-similarity. There are also a few limitations that need to overcome in this research. Firstly, this paper only investigates the traffic at isolated intersections while the results might be different in other types of road traffic. Then, the traffic flow is manipulated by traffic light, which limits the maximum vehicles that can pass through the intersection during a phase. Regardless of these limitations, this paper shall provide a step further for the incorporation of self-similarity in traffic signal design. Even though it is sufficient to prove the feasibility of using Hurst parameter to compute optimal cycle length and minimum delay, more extensive thoughts and ideas should be considered in traffic signal design.

References

Adams, W. (1936). “Road Traffic Considered As A Random Series.” *Journal of Institution of Civil Engineering*, Volume 4, November (1936), pp. 121.

Akcelik, R. (1980). “Time-Dependant Expressions For Delay, Stop Rate and Queue Length at Traffic Signals.” *Australian Road Research Board*, Vermoth South, Australia.

Barau, S. et al. (2015). “Estimation of Traffic Arrival Pattern at Signalized Intersection using ARIMA Model.” *International Journal of Computer Application*, Volume 128 (1), pp. 28-32.

Breiman, L. (1963). “The Poisson Tendancy in Traffic Distrubution.” *The Annals of Mathematical Statistics*, Volume 34, Number 1 (1963), pp. 308-311.

Buckley, D. (1968). “A Semi-poisson Model of Traffic Flow.” *Transportation Science*, 2, pp.107-133.

Cajueiro, D. and Tabuk, B. (2005). “The Rescaled Variance Statistic and The Determination of The Hurst Exponent.” *Mathematics and Computers in Simulation*, 70, pp. 172-179.

Cheng, D. et al. (2013). “Modification of Webster’s Minimum Delay Cycle

Length Equation Based on HCM 2000." *Transportation Research Board*, Annual meeting in Washington, D.C.

Clement, S. and Taylor, M. (2006). "The Simple Platoon Advancement Model of Its Technologies Applied to Vehicle Control At Signalized Intersections." *Journal of Advanced Transportation*.

Clement, S. et al. (2005). "Road Networks of the Future: Can Data Networks Help?" *28th Australasian Transport Research Forum*, pp. 12.

Comert, G. (2013). "Simple analytical models for estimating the queue lengths from probe vehicles at traffic signals." *Transportation Research Part B*, 55, pp. 59-74.

Comert, G. (2016). "Queue Length Estimation From Probe Vehicles At Isolated Intersections: Estimators For Primary Parameters." *European Journal of Operational Research*, Volume 252, Issue 2, pp. 502-521.

Cong, C. et al. (2005). "The Speed, Flow And Headway Analyses of Motorcycle Traffic." *Journal of the Eastern Asia Society for Transportation Studies*, Volume 6, pp. 1496-1508.

Cowen, R. (1975). "Useful Headway Models." *Transportation Research*, 9(6), pp. 371-375.

Crovella, M. and Bestavors, A. (1996). “Self-similarity in World Wide Web Traffic: Evidence and Possible Causes.” *International Conference on Measurement and Modeling of Computer System (ACM SIGMETRICS)*. ACM, Philadelphia.

Dion, F., Rakha, H., & Kang, Y.S. (2004). “Comparison of delay estimates at under-saturated and over-saturated pre-timed signalized intersection.” *Transportation Research Part B*, Volume 38 (2), pp. 100-122.

Dong, L. and Chen, W. (2010). “Real-time traffic signal timing for urban road multi-intersection.”*Intelligent Information Management*, 2.p.483-486.

Fricker, J. and Whitford, R. (2005). *Fundamentals of Transportation Engineering.* 1st ed., Prentice Hall, pp. 100-114.

Elagha, H. and AlShafee, M. (2007). “On The Self-Similar Nature of ATM Network Traffic.” *Issues In Information Science and Information Technology*, Volume 4, 2007.

Frost, V.S. and Melamed, B. (1994). “Traffic Modeling for Telecommunications Network.” *IEEE Communications Magazine*, 32(3), pp. 70-81.

Gerlough, D. and Huber, M. (1975). “Traffic Flow Theory – A Monograph.” *Special Report 165*, Transportation Research Board, National Research Council, Washington D.C, pp. 56-75.

Greenshields, B. et al. (1947). “Traffic Performance at Urban Street Intersections.” *Technical Report No.1*, Yale Bureau of Highway Traffic, pp. 60-85.

Hossain, M. (2006). “Integrated Management of Malaysian Road Network Operations Through ITS Intiatives: Issues, Potentials and Challenges.” *International Seminar On Intelligent Transport Systems (ITS) In Road Network Operations,* Kuala Lumpur, pp. 3.

Hurst, H. (1951). “Long-term Storage Capacity of Reservoirs.” *Transactions of American Society of Civil Engineers*, Volume 116, Issue 1, pp. 770-799.

Jang, J. et al. (2011). “Modeling of Time Headway Distribution on Suburban Arterial: Case Study from South Korea.” *6th International Symposium on Highway Capacity and Quality of Service, Procedia-Social and Behavioral Sciences,* 16, pp. 240-247.

Karagiannis, T., Faloutsos, M., and Molle, M. (2003). “A user-friendly self-similarity analysis tool.” *Special Section on Tools and Technologies for Networking Research and Education,* ACM SIGCOMM Computer Communication Review.

Karagiannis, T., Faloutsos, M., and Riedi, R. (2002). "Long-Range Dependence: Now you see it, now you don't!" *Global Internet Symposium*, pp. 4-9.

Kim, Y. et al. (2004). "Prediction-Based Routing Through Least Cost Delay Constraint." *Proceedings of the 18th International Parallel and Distributed Processing Symposium*, pp. 56.

Kisi, T. and Hiyoshi, K. (1962). "The Effect of A Traffic Light On A Poisson Flow." *The Operations Research Society of Japan*, pp. 67-73.

Knill, O. (1994). "Probability and Stochastic Processes With Applications." *Havard Web-Based*, pp. 11-12.

Leland, W., Taqqu, M., Willinger, W., and Wilson, D. (1994). "On the self-similar nature of Ethernet traffic (extended version)." *IEEE-ACM Transactions on Networking*, 2, pp. 1-15.

Lin, F. (2003). "Variations In Queue Discharge Patterns and Their Implications In Analysis of Signalized Intersections." *TRB 2004 Annual Meeting*, Washington D.C.

Mandelbrot, B. (1975). "Limit Theorems of The Self-Normalized Range For Weekly And Strongly Dependant Processes." *Z.Wahrscheinlichkeitstheorie verw.*, Gebeite (31), pp. 271-285.

McNeil, D. (1968). “A Solution to Fixed-Cycle Traffic Light Problem For Compound Poisson Arrivals.” *J.Appl. Prob 5*, pp. 624-635.

Meng, Q. and Khoo, H. (2009). “Self-similar Characteristics of Vehicle Arrival Pattern on Highways.” *J. Transp. Eng*., 135(11), pp. 864–872.

Miller, A. (1963). “Settings For Fixed-Cycle Traffic Signals.” *Operational Research Quarterly*, Volume 14, pp. 373-386.

Nagatani, T. (2005).Self-similar behaviour of a single vehicle through periodic traffic lights. *Physical A*, 347, pp.673-682.

Park, K. and Willinger, W. (2000). “Self-similar Network Traffic and Performance Evaluation.” *John Wiley & Sons*, Inc.

Pasagic, H. et al. (1998). “Analysis of Vehicle Arrivals At A Selected Intersection In The City of Zagreb.” *Promet-Traffic-Traffico*, Volume 10, No. 5-6, pp. 215-231.

Paxson, V. and Floyd, S. (1995). “Wide-area traffic: The Failure of Poisson Modeling.” *IEEE/ACM Transactions on Networking*, 3, pp. 226-244.

Perati, M.R. et al. (2012). “Self-similar Behavior of Highway Road Traffic and Performance Analysis at Toll Plaza.” *Journal of Transportation*

Engineering, 138(10).p.1233-1238.

Phit, T. and Abe, K. (2006). “Packet Inter-arrival Time Estimation Using Neural Network Models.” *Internet Conference*, Tokyo.

Polaganga, R.K. and Qilian, L. (2015). “Self- Similarity and modeling of LTE/LTE-A data traffic.” *Measurement*, 75, pp.218-229.

Rezaul K.M., Grout V. (2007). “A Comparison of Methods for Estimating the Tail Index of Heavy-tailed Internet Traffic.” *Proceedings of the 2nd international conference on Performance evaluation methodologies and tool*, pp. 29.

Rossi, R., Gastaldi, M. and Pascucci, F. (2014). “Gamma-GQM Time Headway Model: Endogenous Effects In Rural Two-Lane Two-Way Roads.” *Procedia-Social and Behavioral Sciences*, 111, pp. 859-868.

Shalaik, B., Jacob, R. and Winstanley, A.(2012). “Modeling passenger-flow in real-time bus tracking system.” *International Journal of Modeling and Optimization*, 2(4), pp. 432-435.

Sikdar, B. et al. (2002). “Queue Management Algorithms and Network Traffic Self-Similarity.” *Conference: High Performance Switching and Routing*, 2002.

Tanner, J. (1051). “The Delay to Pedestrians Crossing A Road.” *Biometrika 38*, pp. 383-342.

Teply, Stan. (1993). “Probability of Overload at Signalized Intersections.” *Transportation Research Record No.1398, Traffic Flow and Highway Capacity*, pp. 101-110.

Varga, P. and Kun, G. (2005). “Utilizing Higher Order Statistics of Packet Interarrival Times For Bottleneck Detection.” *Proceedings of IFIP/IEEE*, End-to-End Monitoring Techniques and Services, pp. 152-163.

Wardrop, J. (1952). “Some Theoretical Aspects of Road Traffic Research.” *Road Engineering Division Meeting*, pp. 325.

Webster, F. (1958). “Traffic Signal Settings.” *Road Research Laboratory Technical Paper*, No. 39, HMSO, London.

Willinger, W. et al. (1997). “Self-similarity through high variability: Statistical analysis of Ethernet LAN traffic at the source level.” *IEEE/ACM Transactions on Networking*, 5, pp.71-86.

Wormleighton, R. (1965). “Queues At A Fixed Time Traffic Signal With Periodic Random Input.” *Canadian Opnl. Reseach Society*, pp. 100-141.

Yan Xing et al. (2016). “Study On Vehicle Delay Based On The Vehicle

Arriving Distribution At Entrance Lanes Of Intersection." *Procedia Engineering*, Volume 137, pp. 599-608.

Zhang G. and Wang Y. (2011). "Optimizing minimum and maximum green time settings for traffic actuated control at isolated intersections." *IEEE Transactions on Intelligent Transportation System*, 12(1).p.164-173.

Zhang, G. et al. (2007). "Examining Headway Distribution Models Using Urban Freeway Loop Event Data." *TRB 2007 Annual Meeting*, pp. 1-16.

Zheng, X. and Recker, W. (2013). "An adaptive control algorithm for traffic-actuated signals." *Transportation Research Part C Emerging Technologies*, 30, pp. 93–11.