IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

A study of self-similarity in vehicular arrival pattern

To cite this article: Hooi Ling Khoo 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **512** 012016

A study of self-similarity in vehicular arrival pattern

Hooi Ling Khoo

Department of Civil Engineering, Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Bandar Sungai Long Campus, 43000, Kajang, Selangor, Malaysia.

Abstract. Traditionally, the vehicle arrival pattern is modeled based on Poisson assumptions in the road traffic studies. However, recent research has shown that the vehicle arrival patterns exhibit characteristics of self-similarity. This finding will affect delay calculation, queue lengths and other variables that are used in traffic engineering which were based on Poisson assumptions. In this paper, empirical studies are carried out to investigate the existence of self-similarity characteristics in vehicle arrival pattern for traffic flow on highways and isolated signalized intersections. Traffic flow that is recorded from site is processed and analysed. A software is adopted to analyse the possible existence of self-similarity by computing the Hurst parameter from the data set. Hypothesis testing approach is then adopted to evaluate whether the findings are statistically significant. The results confirmed that the vehicle arrival pattern for arterial road exhibit self-similar pattern, and its corresponding headway is heavy-tailed. This finding is important especially in traffic signal design on arterial road.

1. Introduction

Vehicle arrival pattern is one of the fundamental studies in traffic engineering. Most of the traffic applications developed later is based on the knowledge of the arrival pattern, for example the vehicle generation process in traffic simulation model [1], the analysis of the capacity and performance of an unsignalized intersection [2], the determination of the level of service on highways [3] and the optimization of traffic signal setting. This shows that a well understanding of the vehicle arrival pattern is an important issue in the traffic flow theory.

The vehicle arrival pattern is defined as a sequence of vehicles arrive on a highway passing through a specific spot during a specific time period of a day, measured at successive times, spaced at a uniform time interval. In short, vehicle arrival is the number of vehicles counted/observed continuously at a certain fixed point on the highway, during some specific time interval (say every minute) over a time period (say 24 hours). From the statistical analysis point of view, the vehicle arrival pattern can be regarded as a time series or a stochastic process. The identification of the vehicle arrival pattern can expose the time headway distribution of the vehicles, in which the time headway is referred to as the separation between two consecutive vehicles in time.

Poisson process has already been used to model the vehicle arrival pattern for decades. In addition to the road traffic, it has also been adopted to describe the arrival patterns of data packet. Nevertheless, started from a seminar paper by [4], the data network researchers have discovered that most data packet arrival patterns are self-similar process as opposed to the Poisson process [5]; a time series or a stochastic process is called the self-similar if its statistical properties remain invariant under time (or space) rescaling and it often implies the heavy tailed distributions. Leland et al. [4] demonstrated that Ethernet LAN traffic shows the self-similar behavior, while Paxson and Floyd [6] found that Poisson process fails to model the WAN arrivals. They found that the Poisson process model underestimates the burstiness of TCP traffic over a wide range of time scales seriously. Following this, the VBR traffic [7], Local Area Network (LAN) traffic [8] and World Wide Web (www) traffic [9] are found to show the self-similarity characteristics in the data packet arrival patterns. This shows that the conventional assumption of Poisson process is incorrect since it fails to model the data packet arrival patterns accurately.

In road traffic, few researches investigate the self-similarity characteristics of the vehicle arrival patterns. Nagatani [10] is the only work on the traffic self-similarity study. Using the simulated data in a laboratory, he found that a single vehicle passing through a sequence of traffic light with varying cycle time shows the self-similarity behavior. This study shows that there is a possibility of existence of self-similarity in road traffic. Meng and Khoo [11] investigated the arrival on highways and found that the highway traffic exhibit self-similarity pattern. Inspired by these new findings, we examine whether the arterial road vehicle arrival pattern can be described as a self-similar process too.

This objective of the study is to investigate the vehicle arrival pattern on arterial roads in Malaysia. Vehicle arrival pattern is collected from a 4-leg junction in Setapak area, Kuala Lumpur through video camera. Hypothesis testing is adopted to study the significant of the existence of self-similarity pattern. The self-similarity is evaluated by estimating its Hurst parameter using Selfis software. Results indicated that the vehicle arrival pattern at arterial road is self-similar distributed. This means that the conventional way of optimizing traffic signal timing might need to be revised since the arrival pattern is self-similar.

2. Self-similar Process

Self-similarity, in general, describes a phenomenon where a certain property of an object is preserved with respect to scaling in space or time.

Let $X = \{X_t : t = 1, 2, \ldots,\}$ be a discrete time covariance stationary stochastic process or a stationary time series with mean μ and variance σ^2. In terms of a highway vehicle arrival pattern, X_t can be interpreted as the number of vehicles passing through a specific spot on a highway during time period t on a day. Given the stochastic process X, we can define the *m*-aggregated stochastic process $X^{(m)} = \{X_i^{(m)}, i = 1, 2, \ldots,\}$,where m is a positive integer, by averaging the original stochastic process *X* over non-overlapping blocks of size *m*, namely,

$$X_i^{(m)} = \frac{1}{m} \sum_{t=(i-1)\times m+1}^{i\times m} X_t,\ i = 1, 2, \cdots, \tag{1}$$

We say *X* is self-similar with Hurst parameter H ($0.5 < H < 1$), if for all positive integer *m*, $X^{(m)}$ has the same distribution as *X* rescaled by factor m^{H-1}, i.e.,

$$X_t =_d m^{1-H} \sum_{i=(t-1)\times m+1}^{t\times m} X_i,\ \text{for all } m \geq 1 \tag{2}$$

where symbol $=_d$ stands for equality in distribution. Eqn. (2) means that any aggregated stochastic process is distributionally self-similar in the sense that the distribution of the aggregated stochastic process is the same (expect for the change in scale) as that of the original one. This means that X will perceive its pattern with respect to different aggregate scale. In the layman words, X is like a fern leave or a Koch curve in which the shape of the leave or curve remains the same if it is view in different zoom-in scale. For short, *X* is called the self-similar process throughout the rest of this paper.

If X is a self-similar process, according to eqns. (1)-(2), the mean and variance of each random variable in the *m*-aggregated stochastic process $X^{(m)}$ can be calculated by [see, page 54 of Beran (1994)]

$$E\left(X_i^{(m)}\right) = \mu,\ i = 1, 2, \cdots \tag{3}$$

$$\operatorname{var}\left(X_i^{(m)}\right) = \frac{\sigma^2}{m^{2-2H}},\ i = 1, 2, \cdots \tag{4}$$

More interestingly, for all $m \ge 1$, the autocorrelation functions for stochastic processes X and $X^{(m)}$ are identical with the following explicit expression [see, pages 51-52 of Beran (1994)]:

$$r^{(m)}(k) = r(k) = \frac{1}{2}\left[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right], \forall k \ge 1 \tag{5}$$

where the autocorrelation functions $r^{(m)}(k)$ and $r(k)$ are mathematically defined by

$$r^{m}(k) = \frac{E\left[\left(X_{i+k}^{(m)} - \mu\right)\left(X_i^{(m)} - \mu\right)\right]}{\left(\sigma^2 / m^{2-2H}\right)},\ k \ge 1 \tag{6}$$

$$r(k) = \frac{E\left[\left(X_{t+k} - \mu\right)\left(X_t - \mu\right)\right]}{\sigma^2},\ k \ge 1 \tag{7}$$

Recall that the variance of sample mean $\overline{z}$ of a random variable z satisfies $\operatorname{var}(\overline{z}) = \sigma_z^2 / m$, where m is the sample size and σ_z^2 is the variance of random variable z, provided that samples are drawn independently. Therefore, eqn. (4) clearly indicates the statistically dependence among random variables $X_t, t = 1, 2, \cdots$, in the self-similar process X because the Hurst parameter $1/2 < H < 1$. However, the autocorrelation functions shown in eqn. (5) reflect the long-range dependence of the self-similar process, that is,

$$\lim_{k \to \infty} r(k) = 0 \tag{8}$$

$$\sum_{k=1}^{\infty} r(k) = \infty \tag{9}$$

According to eqns. (8) and (9), it can be seen that the autocorrelation function decays to zero as k intends to infinity so slowly that it is not summable. The autocorrelation function $r(k)$ in terms of magnitude is equivalent to function $H(2H-1)k^{2H-2}$ for the large argument k, that is,

$$\lim_{k \to \infty} \frac{r(k)}{H(2H-1)k^{2H-2}} = 1 \tag{10}$$

Let parameter $\beta = 2 - 2H$, we then have $0 < \beta < 1$ as Hurst parameter $1/2 < H < 1$. Eqn. (10) sometimes is rewritten as

$$r(k) \sim H(2H-1)k^{-\beta}, k \to \infty \tag{11}$$

where symbol ~ stands for the equivalence in the sense of eqn. (10).

Eqn. (10) implies that the autocorrelation function of a self-similar process follows a power-law. The power decay is slower than exponential decay, and since $0 < \beta < 1$, the sum of the autocorrelation values of a self-similar process approaches infinity. Note that the Poisson process has infinite variance, stationary and independent increments therefore it is asymptotically self-similar with Hurst parameter equals exactly to 0.5. Hence, it does not possess the long-range dependence according to eqn. (5). If Hurst parameter of a stochastic process is close to 0.5, the process resembles a Poisson behavior [8] otherwise, the closer the Hurst parameter to one, the greater the degree of self-similarity. Perfect self-similar process has the Hurst parameter value equivalent to 1. Karagiannis and Faloutsos [12] further pointed out that the Hurst parameter cannot be calculated in a definitive way. Nevertheless, it can be estimated from real observations.

3. Methodology

The methodology involved in this research is divided into three categories, i.e. data collection, data processing, and data analysis. Data collection is carried out by taking the video images at the intersection for 24 hours for 2 weeks. Then, the data is processed to obtain the vehicle arrival volume for the specified interval for analysis. Selfis software is used to estimate the Hurst parameter to quantify whether the vehicle arrival pattern exhibits self-similar pattern or not. Hypothesis testing is then adopted to ensure that the analysis was statistically significant. The details of the methodology adopted in this research is presented below.

3.1 Data Collection and Processing

Two isolated junctions are chosen for data collection, namely (1) intersection between Jalan Tumbuhan and Jalan Genting Klang, and (2) intersection between Jalan Taman Ibu Kota and Jalan Langkawi. Both intersections are located in Setapak, Kuala Lumpur as shown in Figure 1. These intersections are ideal for this study because the traffic flow of these intersections are not low (i.e traffic flow of 300 veh/hr/lane). In this paper, the lanes in which the vehicles are moving straight across the inter sections are considered. For the first intersection, there is a total of 10 lanes from 4 approaches, and for the second intersection, there is a total of 8 lanes. Therefore, in total, the data are being collected from 18 lanes for both intersections. The data collection was carried out from 13 August 2015 to 26 August 2015.

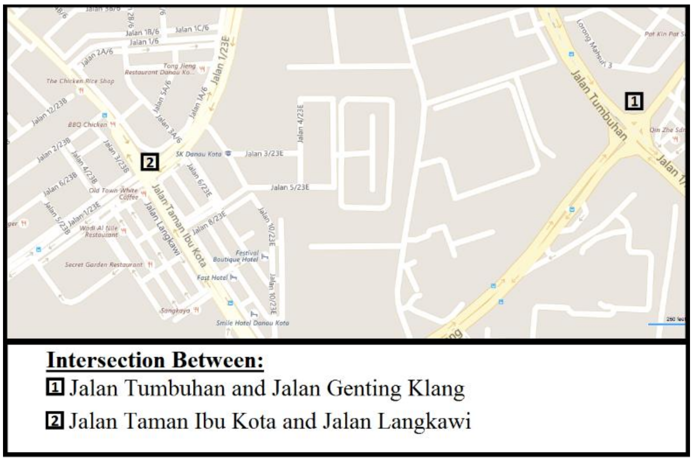

Figure 1. Location of Intersections

Video cameras are 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 obtained is continuous. The video cameras are set up on the light poles around 3 meters from the ground. This ensures the video cameras to capture incoming vehicles from approximately 500 m distance from the light poles. The statistical characteristics for both intersections are tabulated in Table 1. The first intersection is labeled as S1 and the second intersection is labeled as S2. Since there are four approaches for each intersection, each approach is named as C1, C2, C3 and C4 respectively. From Table 1, it can be observed that the maximum vehicle flow rate for the first intersection is approximately 350 vehicle/hour, whereas the second intersection has a lower rate of 270 vehicle/hour.

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 inaccurateness in measuring the traffic flow. Due to this limitation, the timeframe for the second intersection was limited from 7:00:00 AM to 7:00:00 PM, while for the first intersection, 24-hour data are investigated. However, this study ensures that 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 video 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 timestamps for each vehicle is tabulated and analyzed to obtain its vehicle arrival patterns and time headways.

Table 1. Statistic Summary for Traffic Data Collected

Intersection	Lane	Investigation Period	Average Flow Rate (veh/hr/lane)		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

3.2 Data Analysis

3.2.1 Poisson Distribution Test. The first part of the data analysis is focused on investigating whether the vehicle arrival pattern exhibits Poisson distribution pattern. The hypothesis testing for Poisson distribution is as followed:

H_0 : The vehicle arrival pattern follows Poisson distribution

H_1 : The vehicle arrival pattern does not follow Poisson distribution

11

4. Result and Discussion

Figure 2 shows the traffic volume plot vs time for different time scales, i.e. ranging from 1 minute, 2 minutes, 5 minutes, 10 minutes, 30 minutes to 60 minutes interval. It could be observed that the volume pattern is self-repeating for different time frame. This fulfils the definition of self-similarity which warrants the need to further investigate the potential existence of self-similarity in the vehicle arrival pattern.

The results of the hypothesis testing carried out to test the validity of self-similarity existence is shown in Table 2. It shows the average Hurst value for each lane of the intersections. It can be seen that for the first intersection, the average Hurst value estimated by Whittle estimator is more than 0.95, while the average Hurst value estimated by Whittle estimator is more than 0.75 for the second intersection. This shows that the first intersection exhibits a higher degreeness of self-similarity as compared to the second intersection. The results obtained are statistically significant at p=0.01. It is also worth mentioning that since the Hurst parameter is more than 0.5, the hypothesis testing of Poisson distribution of vehicle arrival pattern is rejected.

IOP Conf. Series: Materials Science and Engineering **512** (2019) 012016 doi:10.1088/1757-899X/512/1/012016

Table 2. Hurst value estimated

Intersection	Hurst value		
	Lane 1	Lane 2	Lane 3
Junction 1, Approach 1	0.9754	0.9728	0.9759
Junction 1, Approach 2	0.9592	0.96	-
Junction 1, Approach 3	0.9627	0.9622	0.9728
Junction 1, Approach 4	0.9851	0.9833	-
Junction 2, Approach 1	0.7293	0.7355	-
Junction 2, Approach 2	0.7982	0.7946	-
Junction 2, Approach 3	0.7982	0.7946	-
Junction 2, Approach 4	0.7533	0.7829	-

This study is important because vehicle arrival pattern is one of the assumptions in traffic signal design. The conventional assumption is that vehicle arrival follows Poisson distribution. However, this study shows that the vehicle arrival pattern on urban/arterial street exhibits self-similarity. As such, the traffic signal design using Webster Method might need a revision based on this finding.

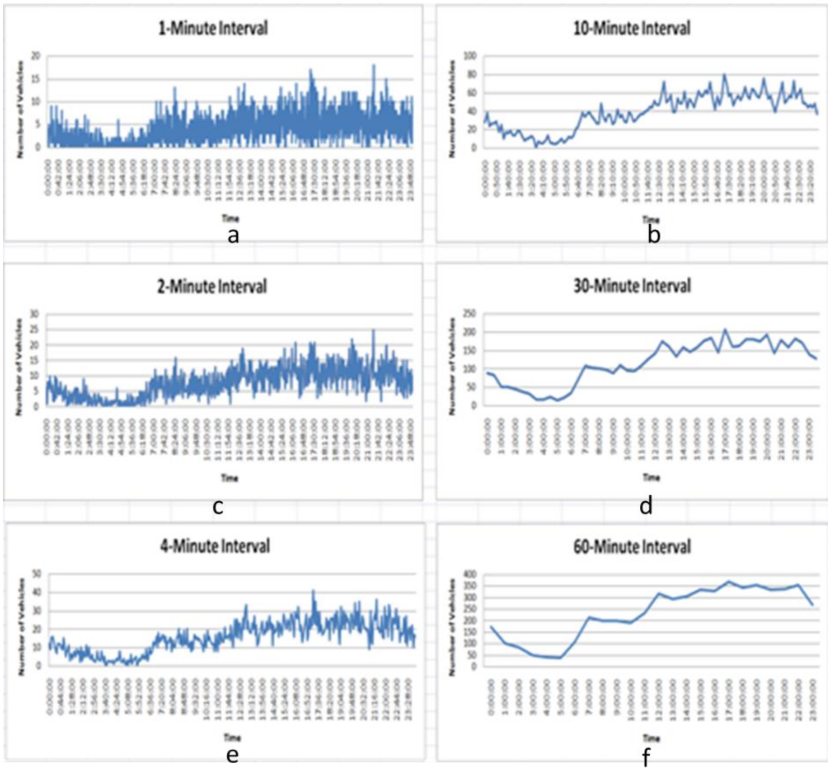

Figure 2. Traffic Volume Pattern At Different Time Frame

5. Conclusion

This study investigates the existence of self-similarity in vehicle arrival pattern. An empirical study was carried out in which vehicle arrival pattern at the isolated intersection was collected and analysed with Selfis software. The hypothesis testing on the Hurst value obtained shows that the vehicle arrival pattern exhibits self-similarity pattern, with statistical significant level of p=0.001. This finding is important as the existing traffic signal design is based on the assumption of Poisson distribution. A revisit of the traffic signal design based on self-similarity arrival pattern is needed.

6. References

[1] Lieberman E and Rath A K 2007 *Traffic Flow Theory: A State-of-Art Report* Available from www.tfhrc.gov/its/tft/tft.htm
[2] Sullivan D P and Troutbeck R J 1994 The use of Cowan's M3 headway distribution for modeling urban traffic flow *Traffic Engineering and Control* **35** 445-50
[3] HCM 2000 *Highway Capacity Manual* (Washington: Transportation Research Board)
[4] Leland W E, Taqqu M S, Willinger W and Wilson D V 1994 On the self-similar nature of ethernet traffic (extended version) *IEEE-ACM Transactions on Networking* **2** 1-15
[5] Park K and Willinger W 2000 *Self-similar Network Traffic and Performance Evaluation* (England: John Wiley &Sons)
[6] Paxson V and Floyd S 1995 Wide-area traffic: the failure of poisson modeling *IEEE/ACM Transactions on Networking* **3** 226-44
[7] Kyu B K and Elwalid A 1996 The importance of long-range dependence of VBR video traffic in ATM traffic engineering: myths and realities *Proceedings of the ACM Sigcomm'96* pp 3-14
[8] Willinger W, Taqqu M S, Sherman R and Wilson D V 1997 Self-similarity through high variability: statistical analysis of ethernet lan traffic at the source level *IEEE/ACM Transactions on Networking* **5** 71-86
[9] 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)* Philadelphia
[10] Nagatani T 2005 Self-similar behaviour of a single vehicle through periodic traffic lights *Physical A* **347** 673-82
[11] Meng Q and Khoo H 2009 Self-similar characteristics of vehicle arrival pattern on highways J. Transp. Eng. **135**(11) 864–72
[12] Karagiannis T and M Faloutsos 2002 *SELFIS: A Tool for Self-similarity and Long-range Dependence Analysis* The First Workshop on Fractals and Self-Similarity in Data Mining: Issues and Approaches (in KDD), Edmonton, Canada

Acknowledgement
This research is supported by Ministry of Education, Fundamental Research Grant Scheme, 2016.