# Statistics of Stop-and-go Traffic: Emergent properties of congestion behavior arising from collective vehicular dynamics in an urban environment

Abdul Majith N. and Sitabhra Sinha
The Institute of Mathematical Sciences,
CIT Campus, Taramani, Chennai 600113, India
Email: majithabdul@imsc.res.in, sitabhra@imsc.res.in

*Abstract*—**The movement of large numbers of vehicles along the complex network of roads in a city result in interactions between them that become stronger as the traffic density increases. The non-trivial behavior arising from the collective dynamics of vehicles include the occurrence of persistent congestion at different points of the transport network that typically reduce the efficiency of overall traffic flow. In order to understand the mechanisms responsible for the characteristic spatio-temporal patterns of urban traffic, we first need to identify statistically robust features from empirical observations, which one can then try to recreate in computational models of traffic dynamics. In this article, we have analyzed the GPS traces collected round the clock for more than a hundred taxis operating in a major Indian city over a period of 1 month. The available information allows us to precisely measure the periods during which the vehicle is static and when it is moving. We focus on the intermittent patterns of rest and motion that a car exhibits during its passage through city traffic, which provides a window into key aspects of collective dynamics resulting from congestion. We show that the distribution of waiting time, i.e., the period during which a car is static between two successive epochs of movement, has a highly skewed nature. The bulk of the probability distribution appears to follow power-law scaling with exponent value of 1.78. As city traffic has very different densities during peak hours and off-peak hours, we have also investigated this distribution at different times of the day. While the power-law scaling is found to be robust, the exact value of the exponent does change slightly. We have also considered the active time distribution, i.e., the period of movement between two epochs when the car is static, which does not exhibit a power-law signature but rather resembles a inverse Gaussian or a log-logistic distribution. We also look at the recurrence relation between the durations of successive waiting times, as well as, that between active time duration and the duration of the preceding waiting time. Our results can be used to help understand how the statistical properties of large-scale traffic movement over complex road networks which characterize cities deviate from that of other types of collective dynamics, e.g., the diffusion of random walkers.**

## I. Introduction

Vehicular traffic provides a fascinating perspective into large-scale social dynamics where the micro-dynamics of interactions between self-propelled agents results in macroscopic changes of phase, e.g., from free flow to jammed states [1], [2], [3]. In the recent decades the spatio-temporal behavior of highway traffic has been investigated in detail. By contrast, urban traffic, i.e., the large-scale movement of cars within a city, which is characterized by high densities and signals coordinating the movement of vehicles converging from several directions, has remained relatively unexplored. However, the increasing number of vehicles in our cities and the resulting higher probability of congestion and jams severely reduces the efficiency of the complex road networks spanning a city that severely affects the quality of life. A deeper understanding of the mechanisms leading to, e.g., stop-and-go traffic in typically urban settings, may help in devising better strategies for controlling the flow of vehicles along the road network, resulting in less likelihood of jams.

Several model studies have tried to understand various aspects of traffic dynamics that are usually seen in an urban environment, such as, the role of traffic light periodicities. However, in order to be able to construct traffic models that accurately describes real phenomena, we first need to identify statistically robust features in the empirical data which can then be reproduced in models. With this aim in view, we have analyzed a data-set containing the GPS traces of 127 taxis plying in the roads of a major Indian city for the period of one month. Use of such data is analogous to the use of tracer particles in fluids to obtain a global understanding of their flow patterns, as opposed to using local probes at specific regions (which would correspond to collecting data on traffic volume at specific locations of the road network) that yield highly localized information.

By focusing on the pattern of intermittent bursts of motion, that are preceded and followed by periods of rest, we try to gain an understanding of the collective dynamics resulting in congestion. In particular, we have analyzed the distribution of waiting times (i.e., the time during which a vehicle does not move at all), and seen power-law scaling whose exponents appear to depend on the overall traffic density. We have also seen the absence of power-law scaling in the distribution of the complementary quantity, active time (which corresponds to the period between two successive waiting times). We also observe that there are recurrence dynamical patterns between two successive waiting periods and between a waiting period and the subsequent active period. Our results should motivate further studies on the properties of waiting times in urban

978-1-4799-8439-8/15/$31.00 ©2015 IEEE

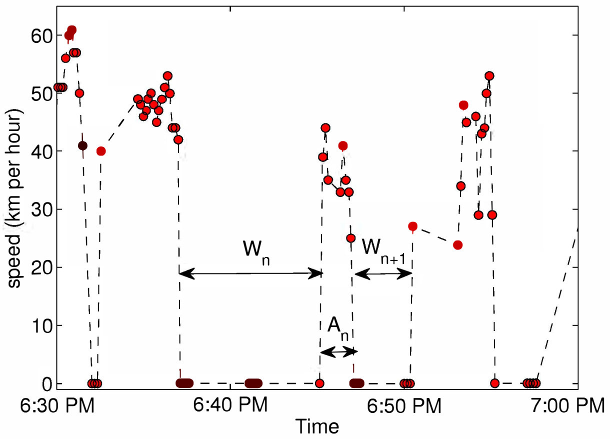


Fig. 1. Time-series of vehicular speed obtained from GPS trace of a single taxi showing typical pattern of intermittent periods of motion and stasis. The duration of the $n$-th period during which the vehicle is stopped is referred to as the $n$-th waiting period, $W_n$, while the duration of the period immediately following this when the car is in motion is called the $n$-th active period, $A_n$.

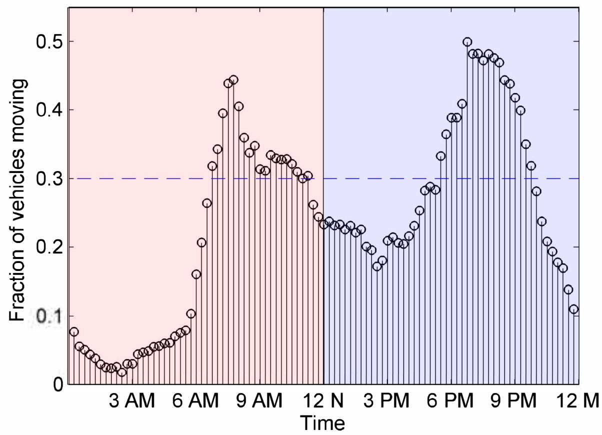


Fig. 2. The average fraction of taxis that are moving at different 15 minute intervals over the 24 hour cycle. The average is calculated over all the days whose data is available. The two different shaded regions correspond to the period before noon and that after noon respectively. An arbitrary threshold, set at the value 0.3, differentiates high-density traffic during peak hours in the morning and evening and low-density traffic during off-peak hours in the afternoon and night.

vehicular congestion.

## II. Data

The data that we have analyzed occurs as GPS traces of 127 taxis in a major Indian city over a period of 24 hours for 30 consecutive days in a month. Apart from the position, the instantaneous speed of the taxi is also mentioned in the data files that allow us to construct time-series, a typical example being shown in Fig. 1 (which also shows how waiting time and active time is defined). Note that, vehicular density is highly variable throughout the day (Fig. 2) - with very high values occurring during the morning and evening peak hours (which we have taken as corresponding to 7 AM-11 AM and 5 PM-10 PM respectively) and conversely lower values during much of the off-peak hours in the mid-day (11 AM-5 PM) and night (10 PM -7 AM).

## III. Results

We have looked at the statistical aspects of waiting times when a vehicle is stopped, either because of signals at intersections or because of congestion due to heavy traffic. The distribution of waiting times calculated over all taxis for the entire period under consideration is shown in Fig. 3. The bulk of the complementary cumulative distribution appears to be fit well by a power-law scaling form, so that the probability distribution of waiting times should also be described by a power law $P(W) \sim W^{-\alpha}$, while the tail is an exponential cut-off that presumably is appearing because of finite-size effects. Note that this deviation from power-law scaling at the tail is appearing for periods larger than $10^4$ seconds; such long durations of waiting periods are unlikely to be because of the vehicle being stopped in traffic. The power-law exponent for the bulk of the distribution is estimated by maximum likelihood methods [4] yielding $\alpha \simeq 1.78$ with the fitting being carried out over the range $W > W_{min} = 239$ secs. Statistical tests of significance have been carried out to ensure that one cannot reject the possibility that the distribution form is described by a power law.

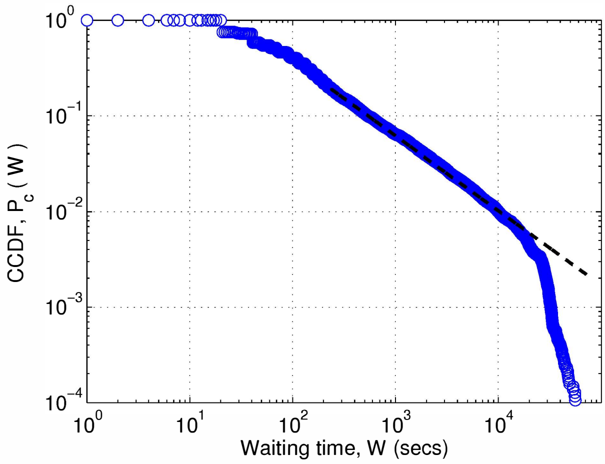


Fig. 3. The complementary cumulative probability distribution of waiting periods $W$ calculated over the weekdays of the entire period for which data is available, showing the fit obtained (broken line) with power-law scaling using maximum likelihood estimation technique (broken line).

As the traffic density varies over the course of a day, being highest during peak hours, it is natural to ask whether the waiting times distribution will be affected by it. Fig. 4 shows the distribution of waiting times at four different intervals of the day, corresponding to the morning peak hours (maximum likelihood estimate of the power-law exponent being $\alpha \simeq 1.75$), mid-day off-peak hours ($\alpha \simeq 1.60$), the evening peak hours ($\alpha \simeq 2.11$) and night off-peak hours ($\alpha \simeq 1.77$). While the exponent value indeed appears to depend on the overall traffic density, being higher during the evening peak hours (corresponding to a sharper decay of the distribution for higher values of the waiting period), the power-law scaling nature of the bulk of the distribution appears to be invariant. The observation of this statistically robust property naturally leads one to ask if a simple theoretical model can be used to generate

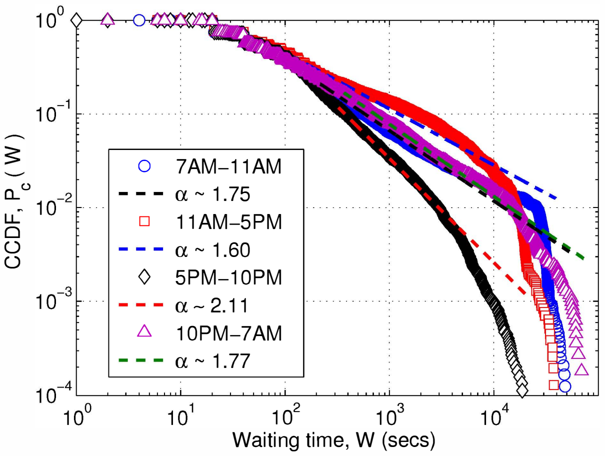


Fig. 4. The complementary cumulative probability distributions of the waiting periods at different times of a weekday, differentiating between peak hours in the morning and evening, and off-peak hours in the afternoon and night. The broken lines indicate the maximum likelihood estimates of the power law exponents during different periods.

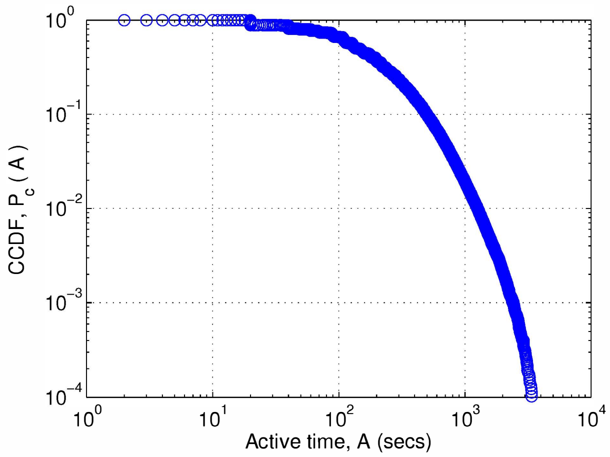


Fig. 5. The complementary cumulative probability distribution of active periods $A$ calculated over the weekdays in the entire period for which data is available. Note that it does not exhibit power-law scaling.

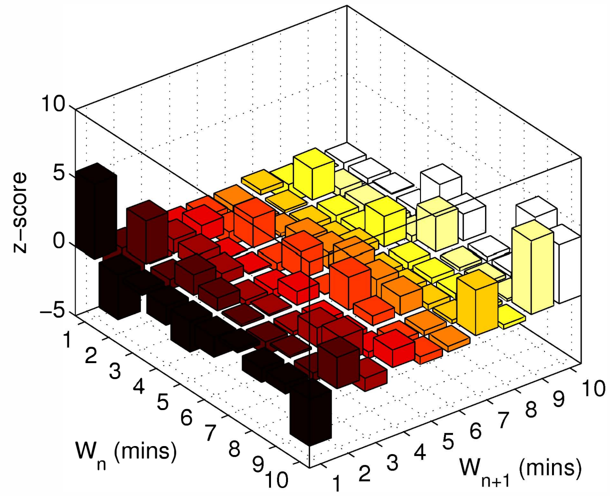


Fig. 6. Statistically significant elements (measured by z-score) for the joint distribution of the durations of two successive waiting periods $W_n$ and $W_{n+1}$ in the data corresponding to weekdays.

the exponent values that have been observed. Queuing models suggested earlier in different contexts are known to give rise to power laws with exponent $3/2$ (see e.g. Ref. [5] and references therein); however, the occurrence of exponent values closer to 2 in the empirical data raises the possibility that, for the present system, one may need to explicitly consider heterogeneity in the system [6] or interactions between congestions occurring at different sections of the network.

One can also ask whether the complementary quantity of active time (i.e., the period during which a car travels with a non-zero speed) also has a distribution characterized by power-law scaling. Fig. 5 shows that the active time distribution does not follow a power law, and is in fact better described by inverse Gaussian or log-logistic forms. Note that these distributions are often associated with times between two successive crossings of a threshold in a random walk. It thus appears that while the active time behavior may be described through a process akin to random diffusion, the waiting time may require a more complex explanation.

We have also looked at the recurrence relation between the durations of two successive waiting periods, $W_n$ and $W_{n+1}$ (Fig. 6), and between the duration of a waiting period $W_n$ and that of the active period $A_n$ immediately following it (Fig. 7). For both of these cases, the statistically significant elements of the joint distribution of the two periods are determined by measuring a $z$-score for the joint probability of each pair of periods, calculated as the difference between the empirical value and the mean value of random surrogates (corresponding to an ensemble of 500 randomized permutations of each sequence of the waiting and active periods), scaled by the standard deviation of the surrogates. Our results appear to suggest that very short waiting times ($\sim 1$ min) are likely to be followed by equally short waiting times, while long waiting times are more likely to be followed by long waiting times. Thus, there seems to be evidence for statistically significant long-range autocorrelation of waiting time durations. On the other hand, short waiting periods ($\sim 2$ mins) are likely to be followed by long active periods (($> 8$ mins) - although, long waiting periods ($\sim 8 - 9$ mins) are also likely to be followed by long active periods (($> 7$ mins). However, one would need to explore in detail the conditional probability distributions of these quantities separately for peak hours and off-peak hours before more substantive conclusions can be drawn.

## IV. Conclusion

In this paper we have considered waiting and active time statistics in data corresponding to GPS traces of more than a hundred taxis operating in a major Indian city for a period of 1 month. By focusing on the intermittent patterns of rest and motion that a car exhibits during its passage through city traffic, we hope to uncover statistically robust features of the collective dynamics. We believe that the use of data from a few mobile elements to understand the properties of the system-wide transport patterns is analogous to using tracer particles in fluids to obtain a global understanding of the flow patterns as opposed to using local probes at specific regions. We observe

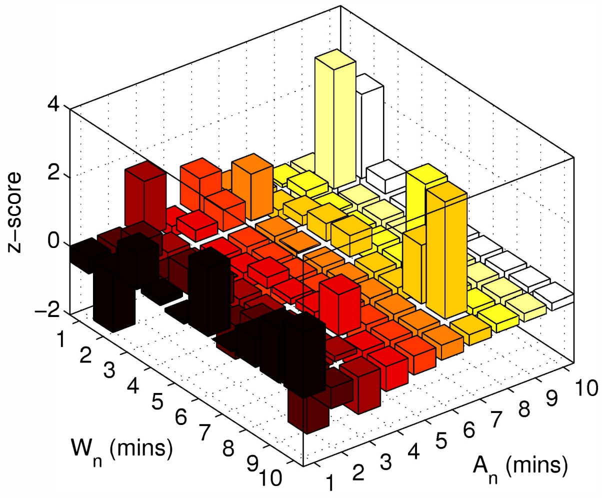


Fig. 7. Statistically significant elements (measured by z-score) for the joint distribution of the duration of a waiting period $W_n$ and the duration of the following active period $A_n$ in the data corresponding to weekdays.

that the distribution of waiting times exhibits a invariant power-law scaling regime in the bulk, whose exponent varies with the time of the day (and hence, the volume of traffic). By contrast, the active time distribution does not exhibit a power-law signature but rather resembles a inverse Gaussian or a log-logistic distribution. The recurrence relations between successive waiting times and between a waiting time and the following active time shows statistically significant elements suggesting evidence of long-range autocorrelation for waiting times. Our results can help in framing theoretical enquiries into the mechanisms that give rise to the large-scale patterns in urban traffic.

## Acknowledgment

The authors would like to thank D. Manjunath and Krishna Jagannathan for help in accessing the GPS trace data analyzed here, Soumya Easwaran and Rishu Kumar Singh for assistance in data analysis, Chandrashekar Kuyyamudi for helpful discussions and IMSc for providing access to the high-performance computing facility. This research was supported in part by the ITRA Media Lab Asia project "De-congesting India's Transportation Network using mobile devices" and IMSc Complex Systems Project.

## References

[1] P. Ball, "The physical modelling of human social systems," *Complexus*, vol. 1, pp. 190206, 2003.

[2] D. Helbing, "Traffic and related self-driven many-particle systems," *Rev. Mod. Phys.*, vol. 73, pp. 1067-1141, 2001.

[3] B. K. Chakrabarti, A. Chakraborti and A. Chatterjee, Eds. *Econophysics and Sociophysics: Trends and perspectives.* Weinheim: Wiley-VCH, 2007.

[4] A. Clauset, C. R. Shalizi and M. E. J. Newman, "Power-law distributions in empirical data," *SIAM Review*, vol. 51, pp. 661-703, 2009.

[5] J. G. Oliveira and A.-L. Barab'asi, "Darwin and Einstein correspondence patterns," *Nature (Lond.)*, vol. 437, p. 1251, 2005.

[6] C. A. Hidalgo, "Conditions for the emergence of scaling in the inter-event time of uncorrelated and seasonal systems," *Physica A*, vol. 369, pp. 877-883, 2006.