Stochastic Models of Traffic Flow Interrupted by Incidents

M. Baykal-Gürsoy*. Z. Duan, and H. Xu *****

Industrial and Systems Engineering Department, Rutgers University
96 Frelinghuysen Road, Piscataway, NJ 08854-8018
**gursoy@rci.rutgers.edu, (732)445-5465, fax (732)445-5465*
*** zheduan@eden.rutgers.edu*
**** xvhao1985@gmail.com*

Abstract: We consider congestion that is caused by irregular occurrences such as traffic accidents, disabled vehicles, adverse weather conditions, spilled loads and hazardous materials. Due to these unexpected events, travel times on the roadways are uncertain. In this paper, we present stochastic models for traffic flow that incorporates uncertain conditions. These models include queueing systems in which customers experience service interruptions from time to time. When a traffic incident happens, either all lanes or part of a lane is closed to the traffic. As such, we model these interruptions either as complete service disruptions where none of the servers work or partial failures where servers work at a reduced service rate. Additionally, the affect of congestion on the traffic flow is also considered. These models are then utilized in estimating the travel times. We present traffic simulation results to show the validity of stochastic models in travel time estimation.

Keywords: queue, Markov modulated, stochastic decomposition

1. INTRODUCTION

Due to increased population, economic growth, changes in the lifestyles (employees living far from their workplace), etc., the demand for transportation has increased exponentially. However, improvements in infrastructure have not kept up with this trend, because prohibitive investment costs and environmental concerns make expansion of current highway systems difficult. Increasing traffic flow on existing roadways inevitably results in a rise in congestion. Congestion leads to delays, decreasing flow, higher fuel consumption and has negative environmental effects. The cost of total delay in rural and urban areas is estimated by the USDOT to be around $1 trillion per year (National Conference 2002). Although congestion during peak hours is expected, congestions at other times are caused by irregular occurrences. This nonrecurrent congestion may be due to traffic accidents, disabled vehicles, natural causes such as adverse weather conditions, and spilled loads and hazardous materials. Well over half of nonrecurring traffic delay in urban areas and almost 100% in rural areas are attributed to such incidents. An "incident" is defined here as any occurrence that affects capacity of the roadway (Skabardonis et al. (1998)). These incidents might also cause other incidents when response to the initial incident is not fast enough or if the traffic flow is not managed well. USDOT estimates that the crashes that result from other incidents make up 14-18% of all accidents. There is close to $200 billion per year in economic loss due to accidents and fatalities (National Conference 2002). In addition, supply chain disruption as a result of incidents decrease overall economic productivity.

Travel time on the roadways is uncertain as a result of these unexpected events. We have shown that even a short lived traffic interruption would have significant effect on travel times; in addition, travel time variability also increases (Baykal-Gursoy and Xiao (2004), Baykal-Gursoy and Duan (2006)). On the other hand, this high variability feeds congestion since drivers cannot make informed decisions on their route selection. Providing reliable travel time estimates, by monitoring the impact of incidents continuously, in combination with effective incident management can decrease congestion, secondary crashes, improve roadway safety and decrease traffic delays. Travel time estimation is also crucial in the field of evacuation, considering the issues such as how to evacuate and move people, and how to move traffic in major cities and on interstate highways, during an emergency.

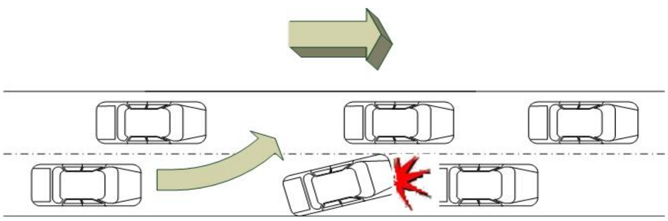

Fig.1. Traffic Flow Interrupted by an Incident

We survey the traffic flow modeling literature in the next section. In section 3, we introduce the stochastic queueing model of traffic flow interrupted by incidents. We present a new approach in section 4 that combines both effects of recurrent and nonrecurrent congestion on traffic flow. Finally, we discuss future research in section 5.

2. TRAFFIC FLOW MODELING

2.1 Classical Theory

Researchers from widely varying disciplines have been paying more and more attention to modeling vehicular travel in order to improve the efficiency of current highway systems. Classical traffic models are mostly based on the treatment of interacting vehicles, their statistical distribution, or their average velocity and density as a function of time and space. Main modeling approaches can be classified as microscopic (particle-based), mesoscopic (gas-kinetic), and macroscopic (fluid-dynamic, deterministic queue) (see Helbing (2001)).

Microscopic approach was developed based on driver's acceleration and deceleration behaviors due to the interaction of vehicles nearby, and called as the car-following model (Richards (1956), Gazis et al. (1959, 1961). Newell (1961) introduced an optimal velocity model by considering a distance-dependent velocity to reflect the safety restriction. Recently, Helbing et al. (1999) proposed an intelligent driver model, taking all the aspects into account for microscopic traffic modeling such as relative velocity and safe driving distance.

Macroscopic traffic theory explains the traffic behavior in terms of average parameters, such as average velocity and average traffic density, based on continuity flow equation, in contrast to microscopic traffic modeling. Lighthill and Whitham (1955) developed the so-called L-W model on the assumption that there is no interruption to the traffic system and they obtained the fundamental diagram. Based on the continuity equation, Whitham (1974) derived the nonlinear wave equation for the propagation kinetic wave and developed the basis of shock wave theory. Whitham (1974) presented the Burgers equation by introducing a diffusion term into wave equation based on the relationship between velocity and density. Kuhne (1987) introduced a viscosity term in the Burgers equation for the negative drivers' reaction to the gradient of traffic flow and a Navier-Stokes velocity equation was obtained. Payne (1971) transformed microscopic variables to macroscopic scale and obtained the Payne's velocity equation, which described the reaction of individual vehicles to the surroundings and adaptation of individual velocity to the equilibrium velocity.

In the mesoscopic approach, driving vehicles are treated as the interacting particles in gas environment. By the continuity equation in phase space, Prigogine and Andrews (1960) modeled the acceleration and overtaking behaviors and obtained the critical density of the phase transition from free flow to congestion (see also Prigogine and Herman (1971)). Paveri-Fontana (1975) improved this model by the introduction of diversity of driver types. Helbing (1995) included a term for adaptation to the road condition and his approximation managed to explain the increase in velocity variance before a phase transition.

Some other authors considered congestion around planned road work and incidents. Gas-kinetic models are introduced to describe the behaviors at bottleneck areas by Shvetsov and Helbing (1999) and Kerner (2004). Redner et al. (1994) introduced the ballistic agglomeration to model one-lane traffic flow and clustering. Lia et al. (2008) developed a traffic control plan based on empirical data, the Dynamic Late Lane Merge System (DLLMS), to improve traffic flow volume and solve potential traffic congestion problems close to work zones.

Kuhne et al. (2002) and Mahnke et al. (2005) developed a stochastic model to describe the traffic behavior and the general master equation was constructed. Combining Markov process and optimal velocity model, they concluded that the formation of traffic congestions was due to stochastic perturbation and dissolution of cluster depended on the cluster size. In analogy to nucleation mechanism, they developed a multi-cluster model on one-lane circular road.

Helbing (2003) proposed a deterministic queueing model for traffic network by dividing the road into free road and congested sections. He estimated the average travelling time and congestion pattern, assuming a fundamental diagram with linear free and congestion branches. Lammer et al (2008a) introduced a model to anticipate the queueing process at the traffic lights and estimate the waiting time. Based on different evolutions of queue length at green, yellow and red lights, they derived the hybrid dynamical equations to obtain the required green time to clear the queue. Lammer and Helbing (2008b) proposed a self-organized traffic-light control at intersections with live data input that minimizes the total waiting time.

2.2 Stochastic Queueing Models

The arrival process in roadway traffic is modeled as a singly arriving Poisson process (Darroch et al. (1964), Tanner (1953)), and as platoons to represent the behavior of the vehicles moving between traffic signals (Alfa and Neuts (1995), Daganzo (1994), Dunne (1967), Lehoczky (1972)). Daganzo (1994) presented a cell transmission model, representing traffic on a highway with a single entrance and exit, which can be used to predict changes in the traffic pattern over time and space. Initially, queuing analysis has been mainly utilized for performance evaluation using *deterministic* (fluid-dynamic) models (May and Keller (1967a,b), Newell (1971)), and synchronization of traffic-lights (Newell (1965)). Stochastic queues were studied by Cheah and Smith (1994) that explored the generality and usefulness of finite server queuing models with state dependent service rate (traveling speed) for modeling pedestrian traffic flows. As an extension, Jain and Smith (1997) used such queues for modeling and analyzing vehicular traffic flow on a roadway segment that can accommodate a finite number of vehicles. In the Jain and Smith model, arrivals are assumed to follow Poisson process (M), travel times are assumed to be generally (G) distributed random variables, and if the link is full, new arrivals should leave and find alternate paths. Consider vehicles traveling on a link as shown in Fig. 2.

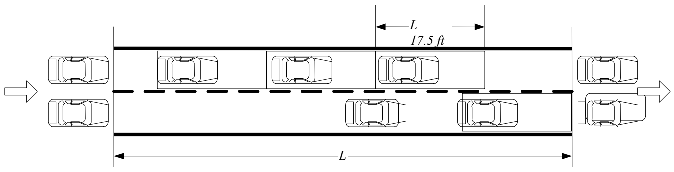

Fig.2**:** A Two-Lane Roadway Link

The space occupied by an individual vehicle on the road segment can be considered as one queuing "server", which starts service as soon as a vehicle joins the link and carries the "service" (the act of traveling) until the end of the link is reached. A "server" in this context is the moving albeit virtual vehicle-space including the safe distance to the vehicle in front. Thus, the maximum number of vehicles that can be accommodated on the link provides the number of servers in the model. Although there are several different types of vehicles utilizing the roadway, in Jain and Smith they are all assumed to be identical and considered as a passenger car equivalent. In practice, the service rate (traveling speed) is assumed to be a decreasing function of the number of the vehicles on the link to represent the congestion caused by the traffic volume. Such a queue is called an M/G/C/C model with the first M denoting the Poisson arrival process, G representing the general service times and finally C denoting both the number of servers and roadway capacity.

Heideman (1996) used M/M/1 where the second M represents the exponentially distributed service times, and M/G/1 queues to model uninterrupted traffic flows. Note that in all the queuing models, both deterministic and stochastic, the link is considered as a **point queue** (or vertical queue, see Daganzo (1997)). Rakha and Zhang (2005) show the consistency of the total delay and total travel time estimates in the gas-kinetic and deterministic queuing models. While in the multi-server case, a link is separated into cells, contrary to the cell transmission model; there is no interdependence between the service times. However, we emphasize that these models have been shown to be effective in representing traffic flow. Van Woensel and Vandaele (2006a) and Van Woensel et al. (2006b) validate the use of queueing models via empirical data and simulation, respectively. They conclude that M/G/1 queueing models are the best models to describe normal traffic flow on a highway, while state-dependent GI/G/m queues were more realistic for the congested traffic. Heidemann (2001) studied the transient behavior of M/M/1 queues to analyze non-stationary traffic flow. Vandaele et al. (2000) also used M/M/1 and M/G/1 queues to model traffic flow. **Although some of these queuing models consider congestion, they all ignore the impact of random incidents on traffic flow.**

3. MODELING TRAFFIC FLOW INTERRUPTED BY INCIDENTS

Consider vehicles traveling on a roadway link, as shown in Fig. 2, which is subject to traffic incidents. During an incident, traffic deteriorates such that both number of servers and the service rate of all servers decrease. Once an incident occurs, the incident management system sends a traffic restoration unit to clear the incident. The number of servers and the service rate of all servers are restored to their normal level when the incident is resolved. The negative impact of incident involves both *congestion* and *reduction of road capacity*. In this study, a lower service rate $\mu' \geq 0$, affecting every server will be used to represent the impact of congestion caused by incidents. The type of service system with batch interruptions is also considered as a Markov modulated service mechanism. Note that, the concepts in this paper also cover, the M/M/1 queuing model considered in Heidemann (1996, 2001) and Vandaele et al. (2000).

Consider a queue with *C* servers working at free speed service rate μ, subject to random interruptions of exponentially distributed durations. During interruptions, the free speed service rates of these *C* servers drop from μ to $\mu' \geq 0$. At the clearance of the interruption, the service rates are restored to μ. We assume that interruptions arrive according to a Poisson process with rate f, and the repair time is exponentially distributed with rate r. The customer arrivals are in accordance with a homogeneous Poisson process with intensity λ. The service times are assumed to be independent and identical exponentially distributed. The interruption and customer arrival processes, and the service and repair times are all assumed to be mutually independent. We would like to emphasize that the Poisson assumption for vehicle arrivals (Van Woensel et al. (2006b), Van Woensel and Vandaele (2006a)) and exponential interarrival times for the incidents (Skabardonis et al. (1997)), are shown to be reasonable. Although the exponential service times may not seem realistic, in our setting, the total time to traverse a link is not going to be exponential. Thus, our model may be considered as having a generally distributed service time.

3.1 Queues with Service Interruptions

The study of queuing systems with service interruptions has received significant attention by researchers in the field. One type of service interruption has already been considered in the context of "vacation" queues where interruptions only happen as soon as the queue becomes empty, or a service is completed. **In general, queues with server vacations are used to model non-preemptive priority systems where customers receive service according to their priority level**. The server continuously serves low priority customers until higher priority customers arrive. When a high priority customer arrives, the server starts serving the new customer upon completion of the service of one, a number of, or all of the low priority customers. Thus, in these models only *complete service breakdowns that happen at the instant of service completion are considered*. These vacation models in steady-state are shown to exhibit the stochastic decomposition property. This fundamental result establishes the relationship between a performance measure (system size distribution, waiting time distribution, sojourn time distribution, etc.) for the queuing system with vacations, and the same performance measure for the same queuing system without vacations (Yadin and Naor (1963), Cooper (1970), Levy and Yechiali (1975), Fuhrmann and Cooper (1985), Shanthikumar (1986), Altiok (1987), Doshi (1990), Chao and Zhao (1998)).

3.2 Queues with Random Service Interruptions

In the traffic flow models that we consider in this research, incidents happen randomly, independent of service completions. The literature on queues with this type of interruptions is relatively scarce. White and Christe (1958) studied a single server queue with preemptive resume discipline, and related such queues to queues with random service interruptions. Gaver (1962) and Keilson (1962) also studied a single server queue with random interruptions. Gaver (1962) obtained the generating functions for the stationary waiting time and the number in the system in an M/G/1 queue. Avi-Itzhak and Naor (1963) derived the expected queue length for M/G/1 queue with server breakdown, also see, Halfin (1972), Fischer (1977), and. Federgruen and Green (1986, 1988). Mitrany and Avi-Itzhak (1968) analyzed M/M/C queue where each server may be down independently of the others for an exponential amount of time. They obtained an explicit form of the moment generating function of the queue size for one-server and two-server systems, and gave a computational procedure for cases with more than two servers. In the above models, servers fail independently of each other and failures are **complete service breakdowns**.

M/G/∞ queue with alternating renewal breakdowns was studied in Jayawardene and Kella (1996); who show that the decomposition property, a well known property of vacation type queues, holds for such queues: the stationary number of customers in the system can be interpreted as the sum of the state of the corresponding system with no interruptions and another nonnegative discrete random variable.

Considering the case of partial failure, the M/M/1 system in a two-state Markovian environment where the arrival as well as the service process are affected, is analyzed via generating functions first by Eisen and Tainiter (1963), then by Yechiali and Naor (1971), and by Purdue (1973). Such queues, in general, in n-state Markovian environments are said to have Markovian arrival processes (MAP) and Markovian service process (MSP), and might be represented in Kendall notation as MAP/MSP/1. Yechiali (1973) considered the general MAP/MSP/1 queue. Neuts (1981) studied M/M/1 and briefly M/M/C queues in a random environment using matrix-geometric computational methods. O'Cinneide and Purdue (1986) analyzed the n-state MAP/MSP/∞ queue, where ∞ represents the number of servers as infinite. In infinite server queues, customers need not wait because a server is always available. For all these queuing models no closed form solution was given. Keilson and Servi (1993) studied a matrix M/M/∞ system in which both the arrival and service processes are Markov modulated. They obtained the generating function of the stationary number of customers in the system in terms of Kummer functions. A single server queue with Markov-modulated arrival and service processes is analyzed in (Adan and Kulkarni (2003)) and asymptotic results are presented. For the special case of M/M/∞ queue with two-state Markov modulated arrival process, they showed that the decomposition property holds, and provided the explicit solution.

Baykal-Gürsoy and Xiao (2004) considered the M/M/∞ system with the two-state Markov modulated service process, e.g., M/MSP/∞ queue. Using the method introduced in Keilson and Servi (1993), they proved that this model also exhibits a stochastic decomposition property, and gave the stationary distribution in closed form as in the following theorem.

Stochastic Decomposition Theorem: The number of vehicles on a link, *X*, in equilibrium has the form

$$X = X_{\varphi} + Y$$

where X_{φ} represents the stationary number of vehicles on a link in uninterrupted traffic and is Poisson distributed. Y represents the additional vehicles on the link as a result of traffic incidents. Y is distributed as Poisson random variable with truncated beta distributed parameter and it is independent of X_{φ}.

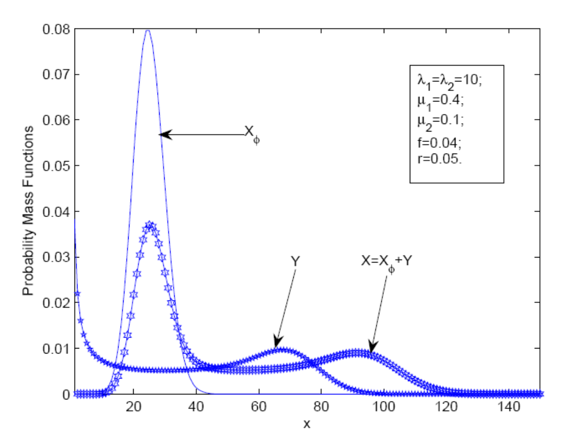

Figure 3: Probability mass function of the stationary number in the system.
Vertical axis gives the probability of having k vehicles on a link.

In Baykal-Gürsoy and Xiao (2004) and Baykal-Gürsoy and Duan (2007), the expected value and variance of X is obtained, it is shown that $Var(X) > E(X)$ in both the complete and partial failure cases, while for uninterrupted traffic, variance is equal to its expected value. Thus, we see that variability of traffic flow increases under incidents. We should also note that the Poisson process where its parameter is randomized by truncated beta is more clustered than a regular Poisson process, which is completely random. The probability mass function of X is given in Figure 3 for a particular set of parameters, together with the Poisson distributed X_{φ}. This figure clearly shows the thick tail characteristic of the stationary number of vehicles on a link. Although the probability of having 100 vehicles in the uninterrupted traffic is almost zero and on the average there are 25 vehicles on the link, under incidents this probability increases to considerable level.

For the infinite server queue with a two-state service mechanism, Jayawardene and Kella (1996) in the case of complete breakdown, and Baykal-Gürsoy and Xiao (2004) also in the case of partial failure, are the first papers showing the validity of the decomposition property. The latter paper has created a renewed interest in infinite server queues in Markovian environment (D'Auria (2005, 2007), Yechiali (2007), Whitt and Pang (2008)).

Baykal-Gürsoy and Duan (2006) consider the queue with multiple classes of interruptions classified according to their severity level, level n being the highest level. Service rate of levels are given as $\mu_1, \ldots, \mu_n$, and they are ordered as the n-th level service rate the lowest, while 1-st level service rate the highest. They show the following result for the M/M/C queue in a general n-failure state service environment.

Proposition 1: For the n-state M/MSP/C queue, the stability condition is,

Stability Condition

$$\lambda < \frac{\sum_{j=1}^{n} C\mu_j \left(\prod_{i=1}^{j-1} f_i \prod_{i=j+1}^{n} r_i \right)}{\sum_{k=1}^{n} \left(\prod_{i=1}^{k-1} f_i \prod_{i=k+1}^{n} r_i \right)}$$

On the other hand, closed form solutions can be obtained for the case of complete service breakdowns such as stopping for traffic lights. Because breakdowns sometimes happen during the service time of customers, the service completion time, i.e., dwell time on a link, may not remain exponential. So, we can say that the system we are solving is an M/G/C queue with a special service structure. Since there is very little known about M/G/C queues, the closed form solutions obtained in Baykal-Gürsoy, Xiao and Ozbay (2009) and Baykal-Gürsoy and Duan (2006) provide a good basis for M/G/C queueing analysis.

3.3 Completion Times

Completion time represents the total travel time on a link. When there is no server failure, completion time is exactly equal to the service time or the travel time without interruptions. But in the case of partial failures or breakdowns, completion time expands to include extra service needed to complete the travel following a failure. Using Little's formula we can obtain the expected travel times from the expected number of vehicles in the system. Furthermore, Baykal-Gürsoy and Duan (2008) have obtained generating function of the completion time for M/M/∞ system with partial failures (incidents). Thus, whole distributional information of completion time, including its mean and variance, can be computed.

4. COMBINED TRAFFIC FLOW MODELING UNDER RECURRENT AND NON-RECURRENT CONGESTION

A modified M/MSP/C/C queuing model is used to model traffic flow subject to both incidents and congestion. The following diagram is the state transition diagram for such a queuing system.

The two dimensional stochastic process $\{X(t), U(t)\}$ describes the state of the system at time t, where $X(t)$ is the number of customers in the system, and $U(t)$ is the status of the system at time t. If at time t, the system is experiencing an interruption, then $U(t)$ is equal to F (*failure*); otherwise, $U(t)$ is N (*normal*). The system is said to be in state (i, F), if there are i customers in the system which is experiencing an interruption, while the system is said to be in state (i, N), if there are i customers in the system which is functioning normally.

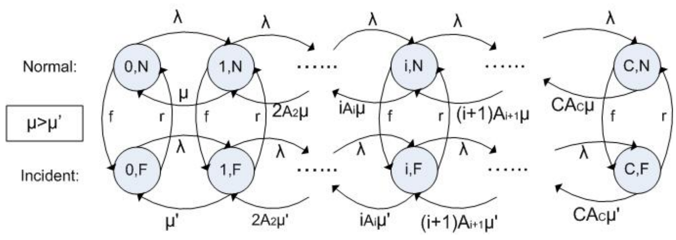

Figure 3: State Transition Diagram for modified M/MSP/C/C Model

In this system, we use the exponential congestion factor A_i to modify the M/MSP/C/C queuing model with

$$A_i = \frac{V_n}{V_{free}}.$$

Here, V_n is the vehicle speed based on free speed $V_{free}$when there are totally n vehicles on the road link given by,

$$V_n = V_{free} \cdot \exp\left[-\left(\frac{n-1}{\beta} \right)^{\gamma} \right]$$

For detailed explanation about exponential congestion model, please refer to Jain and Smith (1997).

C	λ	μ	f	r	Simulation	Incident only Model (M/MSP/C)	Combined Model
400	0.3	0.015	0.0002	0.005	74.93	74.57 (-0.48%)	85.81 (14.52%)
200	0.3	0.03	0.0002	0.005	39.94	39.19 (-1.88%)	47.40 (18.68%)
100	0.3	0.06	0.0002	0.005	23.38	20.84 (-10.86%)	28.85 (23.4%)
50	0.3	0.12	0.0002	0.005	13.08	11.08 (-15.29%)	13.57 (3.75%)

Table 1: Comparison of analytical models with simulation

Table 1 presents the expected travel times for the incident only model of (Baykal-Gürsoy et al. 2009) and the combined model described in this section, together with the traffic simulation results. The relative errors between the simulation and analytic models are given inside the parenthesis in the last two columns. We use INTEGRATION^{TM} to simulate a link with travel speed at full capacity v = 57.5 *km/hr* (65 *miles/hr*). Note that INTEGRATION^{TM} treats the arrival process as fluid, thus generating λ vehicles per hour deterministically. We consider various arrival rates and link lengths on a two-lane roadway where minor incidents happen. Minor incidents take on the average of 7min. to clear (Skabardonis et al. (1997)). They report 0.5 incidents per hour for a one kilometer roadway. The values for f and r are chosen accordingly. For each setting, we run the simulation to obtain average travel times for 100 replications and each replication simulates a 12000-second period. Under congestion, each replication takes 5 minutes, thus each scenario takes more than four hours. This is very time-consuming compared to the analytical model. We would like to emphasize that in the simulation model as in real life, the service times are also **neither independent nor exponential**. The incident process is the only random process in the simulations. Also, here we take μ'=μ/14.

From this set of results, we can conclude that, the incident only model will work well under light traffic situation such as most highway traffic, and combined model will usually be effective under heavy traffic situation, such as city traffic.

5 CONCLUSIONS

In this paper, we have introduced stochastic queueing models for traffic flow interrupted by incidents. We are currently investigating the use of more general service time distributions. We plan to validate our models with empirical data. Empirical validation of our models is currently not possible because real-time queuing and delay data due to accidents are not readily available. Since accidents are random events, it is almost impossible to predict their location and time for real-time data collection. However, we are investigating various approaches for validation.

REFERENCES

Adan I. J. B. F. and Kulkarni V. G., Single-server queue with Markov-dependent inter-arrival and service times, *Queuing Systems*, 45, 113-134, 2003.

Alfa A. S. and Neuts M. F., Modelling vehicular traffic using the discrete time Markovian arrival process, *Transportation Science,* 29:2, 109-117, 1995.

Altiok T., Queues with group arrivals and exhaustive service discipline, *Queuing Systems*, 2, 307-320, 1987.

Avi-Itzhak B. and Naor P., Some queuing problems with the service station subject to breakdown, *Operations Research*, 11, 303-320, 1963.

Baykal-Gürsoy M. and Duan Z., M/M/C queues with Markov modulated service processes, *ValueTools'06*, Oct. 14, 2006.

Baykal-Gürsoy M. and Duan Z., Stochastic decomposition of queues in a random environment, ISE Working Paper 07-022, 2007.

Baykal-Gürsoy M. and Duan Z.,, "Completion time analysis of queues with Markov modulated service", *in preparation*, 2008.

Baykal-Gürsoy M. and Xiao W., Stochastic decomposition in M/M/∞ queues with Markov-modulated service rates, *Queuing Systems,* 48, 75-88, 2004.

Baykal-Gürsoy M.., Xiao W. and Ozbay K., Modeling traffic flow interrupted by incidents, *European Journal of Operational Research,* 195, pp. 127-138, 2009.

Chao X. and Zhao Y. Q., Analysis of multi-server queues with station and server vacations, *European J. of Operational Research*, 110: 2, 392-406, October 1998.

Cheah J. Y. and Smith J. M., Generalized M/G/C/C state dependent queuing models and pedestrian traffic flows, *Queuing Systems*, 15, 365-385, 1994.

Cooper R. B., Queues served in cyclic order: Waiting times, *Bell Syst. Tech. J.*, 49, 399-413, 1970.

Daganzo C. F., The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research-B*, 4, 269-287, 1994.

Daganzo C. F., *Fundamentals of Transportation and Traffic Operations*, Pergamon-Elsevier, Oxford, U.K. (1997).

Darroch J. N., Newell G. F. and Morris R. W. J., Queues for vehicle-actuated traffic light, *Operations Research,* 12, 882-895, 1964.

D'Auria B., M/M/∞ queues with on-off service speeds, http://www.eurandom.nl/reports/2005/044BAreport.pdf

D'Auria B., Stochastic decomposition of the M/G/∞ queue in a random environment, *Operations Research Letters*, 35, p. 805-812, 2007.

Doshi B. T., Single Server Queues with Vacations, in *Stochastic Analysis of Computer and Communication Systems* edited by H. Takagi, 217-265, 1990.

Dunne M. C., Traffic delays at a signalized intersection with Binomial arrivals, *Transportation Science*, 1, 24-31, 1967.

Eisen M. and Tainiter M., Stochastic variations in queuing systems, *Operations Research,* 11:6, 922,927, 1963.

Federgruen A. and Green L., Queuing systems with service interruptions, *Operations Research,* 34: 5, 752-768, 1986.

Federgruen A. and Green L., Queuing systems with service interruptions II, *Naval Research Logistics,* 35, 345-358, 1988.

Fischer M. J., An Approximation to Queuing systems with interruptions, *Management Science,* 24: 3, 338-344, 1977.

Fuhrmann S. W., A Note on the M/G/1 queue with server vacations, *Operations Research*, 32, 1368-1373, 1984.

Fuhrmann S. W. and Cooper R. B., Stochastic decomposition in the M/G/1 queue with generalized vacations, *Operations Research*, 33, 1117-1129, 1985.

Halfin S., Steady-state distribution for the buffer content of an M/G/1 queue with varying service rate, *SIAM J. Appl. Math.,* 23: 3, 356-363, 1972.

Gaver D. P. Jr., A waiting line with interrupted service, including priorities, *J. Roy. Stat. Soc.* B24, 73-90, 1962.

Gazis D. C., Herman R. and Potts R. B., Car-following theory of steady-state traffic flow, *Operations Research*, 7, 499-505, 1959.

Gazis D. C., Herman R. and Rothery R. W., Nonlinear follow the leader models of traffic flow, *Operations Research*, 9, 545-567, 1961.

Heidemann D., A queuing theory approach to speed-flow-density relationships, *Proc. Of the 13th International Symposium on Transportation and Traffic Theory,* France, July 1996.

Heidemann D., A queuing theory model of nonstationary traffic flow. *Transportation Science*, 35:4, 405-412, 2001.

Helbing D., Theoretical Foundation of Macroscopic Traffic Models, *Physica A* 219, 375, 1995.

Helbing.d, Hennecke.A, and Treiber.M, Phase diagram of traffic states in the presence of inhomogeneities, *Phys. Rev. Lett.* 82, 4360, 1999.

Helbing D., Traffic and related self-driven many-particle systems, *Reviews of Modern Physics,* 73, 1067-1124, 2001.

Helbing D., A section-based queuing-theoretical traffic model for congestion and travel time analysis in networks, *J.Phys.A.* 36, L593, 2003.

Jain R. and Smith J. M., Modeling Vehicular traffic flow using M/G/C/C state dependent queuing models, *Transportation Science*, 31: 4, 324-336, 1997.

Jayawardene A. K. and Kella O., M/G/∞ with alternating renewal breakdowns, *Queuing Systems*, 22, 79-95, 1996.

Kaupuzs J., and Mahnke R., A stochastic multi-cluster model of freeway traffic, Eur. Phys. J. B 14, 793-800, 2000.

Keilson J., Queues subject to service interruptions, *Ann. Math. Statistics*, 33, 1314-1322, 1962.

Keilson J. and Servi L. D., The matrix M/M/∞ system: retrial models and Markov modulated sources, *Advances in Applied Probability*, 25, 453-471, 1993.

Kerner B.S., Three phase traffic theory and highway capacity, *Physica A* 333, 379, 2004.

Kuhne.R.D., *in Proceedings of the 10th International Symposium on Transportation and Traffic Theory*, p.119, 1987.

Kuhne R., Mahnke R., Lubashevsky I., and Kaupuzs J., Probabilistic description of traffic breakdowns, Phys. Rev. E, 65, 6125, 2002

Lammer S., Donner R., and Helbing D., Anticipative control of switched queueing systems, Euro. Physics.J.B 63, 341-37, 2008.

Lammer S. and Helbing D., Self-control of traffic lights and vehicle flows in urban road networks, J. Stat. Mech, 1742-5468, 2008

Lehoczky J. P., Traffic intersection control and zero-switch queues, *J. of Appl. Prob.,* 9, 382-395, 1972.

Levy Y. and Yechiali U., Utilization of idle time in an M/G/1 queuing system, *Management Sc*., 22, 202-211, 1975.

Lia G., Tapan D., Catherine H., Dynamic late lane merge system at freeway construction work zones, *Transp.Res.Rec.* 2055, 3, 2008.

Lighthill M. J. and Whitham, G. B., On kinematic waves: II. A theory of traffic on long crowded roads, *Proc. Roy. Soc. London Ser.* A 229, 317-345, 1955.

Mahnke R., Kaupuzs J., and Lubashevsky I., Probabilistic description of traffic flow, *Phys.Rep.* 408, 1, 2005.

May A. D. and Keller H. E. M., Non-integer car-following models, *Highway Res. Rec.,* 199, 19-32, 1967.

May A. D. and Keller H. E. M., A deterministic queuing model, *Transportation Res.,* Vol. 1:2, 117-128, 1967.

Mitrany I. L. and Avi-Itzhak B., A many-server queue with service interruptions, *Operations Research*, Vol. 16, No.3, 628-638, May-June 1968.

National Conference on Traffic Incident Management: A Road Map to The Future, Proceedings, 2-4, March 2002.

Neuts M. F., Further results on the M/M/1 queue with randomly varying rates, *OPSEARCH,* Vol. 15:4, 158-168, 1978.

Neuts M. F., *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach,* The John Hopkins University Press, 1981.

Neuts M. F., *Structured Stochastic Matrices in MG/M/1 Type and Their Applications*, Marcel Dekker, NY, 1989.

Newell.G.F., Nonlinear effects in the dynamics of car following, *Oper. Res*. 9, 209-229, 1961.

Newell G. F., Approximation methods for queues with application to the fixed-cycle traffic light, *SIAM Rev.,* 7:2, 223-240, 1965.

Newell G. F., *Applications of Queuing Theory,* Chapman and Hall, London, 1971.

O'Cinneide C. and Purdue P., The M/M/∞ queue in a random environment, *J. of Appl. Probability*, 23, 175-184, 1986.

Ozawa T., Analysis of queues with Markovian service processes, *Stoch. Models*, 20:4, 391-413, 2004.

Paveri-Fontana S.L, On Boltzmann-like treatments for traffic flow: a critical review of the basic model and an alternative proposal for dilute traffic analysis, *Transp.Res.* 9, 225, 1975.

Payne H.J., Models of freeway traffic and control in G.A.Bekey, *Mathematical Models of Public Systems,* 1, 51, 1971.

Prigogine I. and Andrews F.C., boltzman-like approach for traffic flow, Operations Research 8(6), 789-797, 1960.

Prigogine I. and Herman R., *Kinetic Theory of Vehicular Traffic,* Elsevier, NY, 1971.

Purdue P., The M/M/1 queue in a Markovian environment, *Operations Research*, 22, 562-569, 1973.

Rakha H. and Zhang W., Consistency of shock-wave and queuing theory procedures for analysis of roadway bottlenecks, *TRB Annual Meeting CDROM*, Paper # 05-1763, 2005.

Redner S., Ben-Naim E., and Krapivsky PL., Kinetics of clustering in traffic flows, *Phys.Rev. E* 50, 822, 1994.

Richards P. I., Shock waves on the highway, *Operations Research,* 4, 42-51, 1956.

Shanthikumar J. G., On stochastic decomposition in M/G/1 type queues with generalized server vacations, *Operations Research*, 36, 566-569, 1986.

Shvetsov V., and Helbing D., Macroscopic dynamics of multilane traffic, *Phys.Rev. E* 59, 6328, 1999.

Skabardonis A., Petty, K. F., Bertini R. L., Varaiya P. P., Noeimi H., Rydzewski D., I-880 Field experiment: analysis of incident data, TRB Volume 1603, Washington, D.C., 1997.

Skabardonis A., Petty, K., Varaiya P. and Bertini R., Evaluation of the freeway service patrol (FSP) in Los Angeles, *UCB-ITS-PRR-98-31, California PATH Research Report, Institute of Transportation Studies, University of California, Berkeley,* 1998.

Tanner J. C., A problem of interface between two queues, *Biometrica*, 40, 58-69, 1953.

Transportation Research Board, 2000. Highway Capacity Manual, fourth ed. National Research Council, Washington, DC.

Vandaele N., Van Woensel T. and Verbruggen N., A queuing based traffic flow model, *Transportation Research-D: Transportation and Environment,* 5:2, 121-135, 2000.

Van Woensel T. and Vandaele N., Emprical validation of a queueing approach to uninterrupted traffic flows. 4OR-A Quart. J. of Operations Research 4 (1) 59-72, 2006.

Van Woensel T., Wuyts B. and Vandaele N., Validating state-dependent queuing models for uninterrupted traffic flows using simulation, *4OR-A Quart. J. of Operations Research*, 4, 159-174, 2006.

White H. C. and Christe L. S., Queuing with preemptive priorities or with breakdown, *Operations Research*, 6, 79-95, 1958.

Whitham. G. B., *Linear and Nonlinear Waves,* Wiley, New York, 1974.

Whitt W. and Pang G., Heavy-traffic limits for many-server queues with service interruptions, unpublished report, 2008.

Yadin M. and Naor P., Queuing systems with a removable service station, *Operations Research Quart*., 14, 393-405, 1963.

Yechiali U., A queuing-type birth and death process defined on a continuous-time Markov chain, *Operations Research*, 21, 604-609, 1973.

Yechiali U. and Naor P., Queuing problems with heterogeneous arrivals and service, *Operations Research*, 19, 722-734, 1971.

Yechiali U., Queues with system disasters and impatient customers when system is down, *Queueing Systems,* Vol. 56:3/4, 195-202 (2007).