Queue spillovers in city street networks with signal-controlled intersections

Nikolas Geroliminis Urban Transport Systems Laboratory, EPFL

Conference paper STRC 2009
Queue spillovers in city street networks with signal-controlled intersections

Nikolas Geroliminis
School of Architecture, Civil and Environmental Engineering, EPFL
Lausanne

September 2009

ABSTRACT
Quick and reliable estimation of traffic conditions is of critical importance for advanced traffic management systems, especially when heavy congestion occurs. This paper proposes a methodology to predict queue spillovers in city street networks with signalized intersections using data from loop detectors (counts and occupancy). The key idea of the developed methodology is that when spillovers from a downstream link, vehicle arrivals from the upstream signal line, queues discharge at rates smaller than the saturation flow. The paper underlines the significant effect of spillovers in congestion and analyzes an extension of the method to capture variations in vehicle lengths. It also shows that a macroscopic diagram connecting spillovers with car density exists for large scale urban networks which are homogeneously congested. The application of the methodology on an arterial site and the comparison with field data shows that it accurately identifies spillovers in city street networks with signal-controlled intersections.

Keywords
Spillovers – Arterials – Traffic signals – Macroscopic Fundamental Diagram
1. INTRODUCTION

Quick and reliable estimation of traffic conditions is of critical importance for advanced traffic management systems, of which the basic objectives are to provide road users with timely and reliable traffic information and to improve traffic through adaptive tuning of control strategies based on current congestion. Substantial research effort has been devoted to develop accurate and reliable techniques for estimation of various congestion measures such as queue lengths and travel times.

It is common knowledge that high occupancy is related directly to congestion. In freeways higher values of occupancies are relevant to smaller observed speeds, while in arterials long physical queues contribute to increased occupancy. When these physical queues exceed the link length block the arrivals from the upstream link and lead the system to restricted mobility and service inefficiency.

Many studies have been developed for traffic control and queue management of oversaturated arterials (e.g., [1], [2]). Later, many researchers developed control schemes including objectives for avoidance of traffic spillovers [3], [4]. To develop reliable control strategies, monitoring of the system and knowledge of intense traffic phenomena which drastically affect the state of the system (queue spillovers or incidents), are important.

Nevertheless, literature is limited in methods to identify spillovers in city street networks with signal-controlled intersections. This paper develops a robust method to identify spillovers based on data collected from loop detectors. As the formulation of the problem is elegant and sophisticated, it can be applied in large scale congested networks and provide important information for highly oversaturated links and active bottlenecks in a city street network.

Within the past two decades, loop detector technology has become an indispensable tool for traffic operation and surveillance systems. The output of most current single inductive loop detectors is a simple relay or semiconductor closure, signifying the presence or absence of a vehicle. Although the loops are read many times a second, the data (traffic volumes and occupancy) are typically reported back to the centralized transportation management center (TMC) at intervals of 30 seconds or more aggregated observational periods. However, for the purposes of traffic management it is frequently useful to have data on mean vehicle speeds, but this is not directly available from single loop detectors. While detector occupancy is related in a simple fashion to vehicle speed and length, the latter variable is not measured on the vehicles that pass.

Given the widespread use and relatively low cost of single loop detectors, methods for obtaining speeds from the count-occupancy data produced by these detectors are of particular interest. The first related work used the landmark speed estimation formula proposed by Athol [5]

$$\bar{u}_i = \frac{N_i \cdot L_{eff}}{T \cdot o_i},$$  

where $\bar{u}_i$ is space mean speed for interval $i$, $N_i$ and $o_i$ are volume and occupancy in $i$, $T$ is time length per interval and $L_{eff}$ is speed estimation parameter with units of distance, known as average effective vehicle lengths (EVL) of the traffic stream.
In practice, $L_{\text{eff}}$ has been assumed to be a constant value. For example, Washington State Department of Transportation (WSDOT) uses $L_{\text{eff}}=20-25\text{ft}$ [6]. In reality, $L_{\text{eff}}$ varies as the average EVL changes with vehicle composition. Hellinga [7] proposed an algorithm that uses dual-loop measured vehicle lengths to calculate $L_{\text{eff}}$ and applied that value to estimate speed at adjacent single loop stations. Wang and Nihan [8] studied the relationship between lane occupancy and speed and concluded that $L_{\text{eff}}$ could be considered constant only when all vehicle lengths were approximately equal.

The remainder of this paper is organized as follows: The next section proposes a methodology to predict spillovers in city street networks with signalized intersections and applies some statistical analysis to catch variations in vehicle lengths. Section III underlines the significant effect of spillovers in congestion by providing results from a simulation of San Francisco downtown district. Section IV verifies the accuracy of the method using real world data while the last section contains conclusions and ideas for future work.

2. METHODOLOGY ANALYSIS

Loop detectors is one of the most prevailing technologies for road-based traffic surveillance. They are usually placed sufficiently upstream from the intersection stopline (system detectors), so measured flows and occupancies are not affected by the presence of queues at the traffic signal. This assumption is violated in cases of heavier traffic or oversaturated conditions as the growing queues block the detector line. This section analyzes an approach to predict heavy traffic conditions and especially spillovers in city street networks with signal-controlled intersections using occupancy and counts data from detectors. The analysis is based on selection of data in time intervals of one cycle, but can be applied comfortably for more disaggregated data. LWR theory [9], [10] is applied, which explicitly considers the temporal and spatial formation of queues.

One of the basic ideas of the LWR theory is that shock waves are generated by the traffic signal, which causes (i) congested conditions to develop near the stop line during the red interval, and (ii) capacity conditions to occur in the period during which the queue is discharging at the saturation flow rate. When the queue has dissipated, (iii) the rest of the platoon that arrives during the green time crosses the intersection stopline without any interference from the traffic signal. Figure 1 describes the development of shockwaves in an isolated signal for a triangular flow-density relationship with parameters $u_f$ (free flow speed), $u$ (congested wave speed) and $k_j$ (jam density). Because the system detector is placed upstream of the stopline the effect of red phase and queues is identified with a time lag.

The key idea of the developed methodology is that when spillovers from the downstream signal block vehicle arrivals from the upstream signal line, queue discharges at rates smaller than the saturation flow. The first part of this section sets out the relation between growing queues and high values of occupancy, while the second part explicates the threshold values of occupancy for which spillovers occur.

Identification of growing queues

When growing queues pass the detector line, platoon arrivals are not known, as for an amount of time (interval $t_2$ in fig. 1) vehicle counts are zero and detector is fully occupied (occupancy values near to 1). For a triangular flow-density relationship, drivers arrive from the upstream signal and reach the detector at free flow speed $u_f$ if queues do not constrain their mobility. By using Athol formula, the critical value of occupancy for which queue length
overcomes loop detector line $\bar{\sigma}_c$ is given by equation (2). The importance of this simple formula is that if the variations of the free flow speed between drivers are small this value is stable independently of the time interval data are collected; $\bar{q}$ is the average flow measured by the detector.

$$\bar{\sigma}_c = \frac{L_{\text{eff}} \cdot \bar{q}}{u_f}.$$  \hspace{1cm} (2)

Cycle $c$ is divided in four time intervals using detector position as a reference space point. Interval $t_1$ is the time queues do not reach the detector line while vehicles arrive during red; $t_2$ is the time the detector is occupied because of the congested conditions developed near the stop line during the red phase, $t_3$ is the time queue discharges at the saturation flow and $t_4$ is the remaining green time platoon that arrives cross stopline without any interference from the traffic signal. $L_d$ is the distance between the detector and the stop line, and $L_q$ the maximum queue length.

\textbf{Figure 1:} Shockwaves in an isolated signal

When average occupancy $\bar{\sigma}$ is greater than $\bar{\sigma}_c$, growing queues pass the detector and $t_2>0$. Then, average occupancy $\bar{\sigma}$ is given by equation 3. The first term is the free flow occupancy whilst the second one expresses the congested conditions around the detector. The term $o_{\text{stop}}$ is the jam occupancy with a value near to 1 and $N$ is the vehicle counts during that cycle. Free flow speed $u_f$ can be calibrated from data for undersaturated conditions while effective vehicle length is approximately equal to the vehicle length plus the detector length (about 20-25ft).

$$\bar{\sigma} = \frac{\sum_{i=1,2,3,4} t_i \frac{L_{\text{eff}} q_i}{u_f} + t_2 \cdot o_{\text{stop}}}{c} = \frac{L_{\text{eff}} \cdot N + t_2 \cdot o_{\text{stop}}}{c} = \frac{L_{\text{eff}} \cdot \bar{q}}{u_f} + \frac{t_2 \cdot o_{\text{stop}}}{c}.$$  \hspace{1cm} (3)
The time loop detector is occupied, $t_2$, after some manipulations and assuming $o_{stop} \approx 1$ is

$$t_2 = \frac{c}{o_{stop}} \left( \bar{\phi} - \frac{L_{eff} \cdot \bar{q}}{u_f} \right) \approx c \cdot \left( \bar{\phi} - \frac{L_{eff} \cdot \bar{q}}{u_f} \right).$$  \tag{4}

The estimation of $t_2$, which is based on values of observable quantities, is an important tool for identifying different traffic conditions. Firstly, when $t_2$ takes negative values queues do not reach detector line for the whole cycle. This information provides an unfailing sign that the related link is uncongested and that the vehicle accumulation in that link never exceeds the value $L_d \cdot k_j$, where $L_d$ is the distance between the detector and the stopline and $k_j$ is the jam density. Furthermore, as $t_2 > 0$ increases, (i) queue reaches the detector and (ii) the arrival rate during that cycle is more intense. Thus, the rate queues are growing in that link is faster as this expressed by shockwave speed $\bar{w}$ in figure 1. Next subsection analyzes the methodology to identify the spillovers from the estimated value of $t_2$ using counts and occupancy data.

**Identification of spillovers**

The role of spillovers in city street networks with signal-controlled intersections is significant. It is already known for example that spillovers past merges can lead to gridlock on ring roads and other networks with closed loops [11]. Spillovers are a result of highly oversaturated conditions, but simultaneously is the cause for increasing intensely queued traffic and creation of more congestion as a chain reaction. An illustration of the above statement is addressed in figure 2, which shows a time-space diagram for three consequent signals. Vehicles arrive at the first upstream signal at rate $A$ and during the through red phase of the signals turning traffic enters the through direction with rate $B < A$. It is easily recognizable that the upstream signal is operating at undersaturated conditions as residuals queues do not occur. But growing queues in the downstream signal block the arrivals from the middle signal and after two cycles spillovers reach the upstream undersaturated signal. Spillovers can also occur when turning vehicles fill up the available storage length of turn bays and block the through movements.

It is clear that spillovers reduce the efficiency of the system and more importantly, they reduce the capacity rate when demand is high and the green time is needed to serve traffic movements. We could consider blocking traffic as interpolation of growing red phases during the green time.

The key idea in the identification of spillovers is the observation that queue discharges at rates smaller than the saturation flow. So, the methodology focuses on coinstantaneously recognizing the presence of queues and discharging rates smaller than saturation flow $s$. To achieve this goal time interval $t_2$ is considering using equation 4. If blocking traffic does not exist then the maximum value $t_2$ can take is equal to the duration of the red phase $r$ for the cycle $c$. This is because the maximum speed of the shockwave is the congested wave speed $u$ which occurs when the arrival rate is equal to the saturation flow $s$. By applying this observation ($t_2 = r$) to equation (3) “blocking occupancy” $o_{sp}$ is estimated:

$$o_{sp} = \frac{L_{eff} \cdot \bar{q} + r}{u_f \cdot c}.$$  \tag{5}
Therefore, if the measured occupancy is greater than the critical value $o_{sp}$ (blocking occupancy), then it is concluded that spillovers from the downstream signal block the departures from the first upstream one. This condition is sufficient for the occurrence of spillovers but not necessary. It is a way to identify the spillovers when oversaturated conditions take place and queues exceed loop detector’s line. Consequently, the described methodology captures the existence of “active” spillovers. By that term we mean spillovers, which result intense congestion problems and underutilization of the green phase. This kind of spillovers are accountable for drastic decrease in vehicles mean speed and possible extension of spillovers to more links upstream if the demand continues to be high. For example, if the analyzed methodology is applied in traffic conditions presented in figure 2, blocking traffic for the middle signal is identified with a delay of one cycle.

**Method adaptation to capture vehicle length variations**

As analyzed in the previous section, using occupancy and count data from detectors we can predict the existence of spillovers assuming a constant effective vehicle length $L_{eff}$. This section provides an extension to the methodology when variations in the effective vehicle length occur.

Higher value of occupancy is caused not just because of long queues which pass the detector line but also because of long vehicles which travel in free flow speed and occupy the detector more time because of their length. The following analysis shows that the effect of vehicle length variations in the estimation of $o_{sp}$ is not significant.

Vehicle length distribution follows a bimodal distribution having two separated peaks, one higher for the short vehicles and one smaller for long vehicles (trucks, buses etc.) similar to this of figure 3.
Yeo and Skabardonis [12] found, by analyzing single-vehicle lengths observed by cameras, that both small vehicles (SV) and long vehicles (LV) lengths are normally distributed at 0.01 significance levels. Therefore, SV lengths are assumed to follow the $N(\mu_1, \sigma_1^2)$ distribution (call it $x_1$), and LV lengths to follow the $N(\mu_2, \sigma_2^2)$ distribution (call it $x_2$), where $\mu_1$ and $\sigma_1$ are the mean and standard deviation of SV lengths, and $\mu_2$ and $\sigma_2$ are the mean and standard deviation of LV lengths. Now, we can approximate the vehicle length density function $\pi$ as:

$$\pi = \xi x_1 + (1-\xi)x_2,$$

where $\xi$ is the Bernoulli distribution with probability of success $p$, the percentage $p$ of SVs. By assuming that $\xi$, $x_1$ and $x_2$ are independent, the average vehicle length is simply $p\mu_1+(1-p)\mu_2$.

Regarding the variance we have that:

$$\text{var}[\xi x_1 + (1-\xi)x_2] = \text{var}[\xi x_1] + \text{var}[1-\xi]x_2 + 2 \text{cov}(\xi x_1, (1-\xi)x_2) = \text{var}[\xi x_1] + \text{var}[1-\xi]x_2 + 2 \text{cov}(\xi x_1, x_2) - 2 \text{cov}(\xi x_1, \xi x_2),$$

$$= \text{var}[\xi x_1] + \text{var}[1-\xi]x_2 - 2 \text{cov}(\xi x_1, \xi x_2). \quad (7)$$

Using statistical analysis and after some manipulations we have that for $i=1, 2$:

$$\text{cov}(\xi x_i, \xi x_2) = E[\xi^2 x_i x_2] - E[\xi x_i] \cdot E[\xi x_2] =$$

$$= E[\xi^2]E[x_i x_2] - E^2[\xi] \cdot E[x_i x_2] =$$

$$= E[\xi^2]E[x_i x_2] - E[\xi] \cdot E[x_i x_2] =$$

$$= p \cdot (p^2) \mu_1 \mu_2. \quad (8)$$

$$\text{var}\{\xi x_i\} =$$

$$= E[\xi^2 x_i^2] - E^2[\xi x_i] = E[\xi^2] \cdot E[x_i^2] - E^2[\xi] \cdot E[x_i] =$$

$$= E[\xi] \cdot (E^2[x_i] + \text{var}[x_i]) - E^2[\xi] \cdot E^2[x_i] =$$

$$= p \cdot (\mu_i^2 + \sigma_i^2) - p^2 \cdot \mu_i^2. \quad (9)$$
From equations (7), (8) and (9) the variance of the vehicle length distribution $\pi$ is given by:

$$\text{var}\left[\xi x_i + (1-\xi)x_2\right] =$$

$$= p\cdot(\mu_1^2 + \sigma_1^2) - p^2 \cdot \mu_i^2 + (1-p)\cdot(\mu_2^2 + \sigma_2^2) -$$

$$(1-p)^2 \cdot \mu_2^2 - 2\left(p-p^3\right)\mu_i\mu_2 =$$

$$= p\cdot(\sigma_1^2 - \sigma_2^2) + \sigma_2^2 + (p-p^3)\cdot(\mu_i - \mu_2)^2.$$  \hspace{1cm} (10)

Table 1 shows the estimated value for “blocking occupancy” by ignoring or considering variations in vehicle lengths. In the first case, formula 5 was applied using average values for the effective vehicle length, $L_{eff}$. In the second case instead of the average, 95th percentile value of vehicle length was applied for $L_{eff}$.

The results show that the variation in vehicle lengths is not of critical importance for the estimation of the blocking occupancy (the absolute error is less than 2% in most of the cases) and the described methodology could be applied in cases where vehicle lengths vary.

### Table 1: Critical blocking occupancy with or without variations in vehicle lengths for different values of $p$ and $g/c$ ($\mu_1=6m$, $\sigma_1=0.7m$, $\mu_2=13m$, $\sigma_2=2m$, $u_f=15.65m/sec=35mph$, $\bar{q}=0.5\text{vh}/\text{sec}/(g/c)$)

<table>
<thead>
<tr>
<th>$p$</th>
<th>$g/c$</th>
<th>No variation in lengths</th>
<th>With variation in lengths</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.2</td>
<td>0.4</td>
<td>0.6</td>
</tr>
<tr>
<td>0.99</td>
<td>.839</td>
<td>.678</td>
<td>.516</td>
</tr>
<tr>
<td>0.95</td>
<td>.841</td>
<td>.681</td>
<td>.522</td>
</tr>
<tr>
<td>0.90</td>
<td>.843</td>
<td>.686</td>
<td>.528</td>
</tr>
<tr>
<td>0.85</td>
<td>.845</td>
<td>.690</td>
<td>.535</td>
</tr>
</tbody>
</table>

3. THE EFFECT OF SPILLOVERS IN URBAN TRAFFIC

In reference [13], the authors have found that for neighborhoods of a city (with a size comparable to a trip length) where congestion is (roughly) evenly distributed, total output of the system (rate vehicles reach their destinations) is only a function of the total accumulation and has a shape as this of figure 4. They also stress that the maximum output produces the greatest mobility benefits in terms of vehicle-kilometers traveled per time unit.

Their analysis is based on a simulation of the San Francisco Downtown Area. Further analysis of the simulated network is presented in this section to highlight the effect of spillovers in urban traffic. A diagram of the area is shown in figure 4.
This test site is a 2.5 square mile area of Downtown San Francisco (Financial District and South Of Market Area), including about 100 intersections with link lengths varying from 400 to 1,300 feet. The number of lanes for through traffic varies from 2 to 5 lanes and the free flow speed is 30 miles per hour. Traffic was simulated in this study site for a period of 4 hours with time- and space-dependent demand, starting from low flows increasing to higher flows until the system gridlocks. Many runs were made with significantly different demand profiles (different color for each run in figure 5).

Figure 5: Output vs. accumulation pairs for different runs in the San Francisco network aggregated every two cycles. (taken from Geroliminis and Daganzo, 2007)
Figure 6: Demand description: (a) Total Trips Originated per node per run; (b) Total Trips Ended per node per run

Figure 6 shows the total number of trips generated and completed per node in a period of 4 hours for two different runs. Runs differed vastly in the geographical distribution of demand. For example, in run 2 more than 70% of the demand origins were internal to the SFBD whereas in run 1, 80% of the traffic was external, entered the SFBD from the north. Trip endings followed similar profiles (mostly external in run 1, mostly internal in run 2).

Our conjecture is that the existence of spillovers is highly related to the decrease in the system output and constraints mobility. If the system reaches a congested state in the decreasing part of output-accumulation curve and demand continues to be high, then accumulation increases and
this can lead the system to gridlock. To validate this conjecture, the number of spillovers is calculated in each cycle as the simulation runs. This is the number of vehicles that could get served if queues from the downstream link, did not spillback and block the arrivals.

Figure 7: Output vs. spillovers pairs for two different runs in San Francisco Network

Figures 7 and 8 show “total output from the system” vs. “number of spillovers” pairs and “number of spillovers” vs. “accumulation” pairs (number of vehicles in the system) aggregated per two cycles for two different runs. It is clear that when the number of spillovers increases the performance of the system decreases drastically, as fewer vehicles reach their destinations per cycle. We observe that for each additional vehicle in the system almost one more spillover occurs (for values of accumulation in the decreasing part of curve in figure 4). These figures verify that any control strategy which could contribute to the avoidance of spillovers and maintain accumulation in the non-decreasing part of figure 4, would result in mobility improvements. This should be an item of some research priority.

Figure 8: Spillovers vs. accumulation pairs for two different runs in San Francisco Network
4. METHOD VERIFICATION USING REAL WORLD DATA

The site and the data set

The selected test site is 1.42 mile long stretch of a major urban arterial (Lincoln Avenue) near to Los Angeles International Airport. A diagram is shown in figure 9. The study section includes 7 signalized intersections with link lengths varying from 500 to 1,600 feet. The number of lanes for through traffic per link is three lanes per direction. Additional lanes for turning movements are provided at intersection approaches. The free flow speed is 40 mph. Traffic signals are all multiphase operating as coordinated under traffic responsive control. System cycle lengths range from 100 seconds early in the analysis period (6:00 to 6:30 am) to a maximum of 150 sec during the periods of highest traffic volume (7:30 to 8:30 am). System loop detectors are located on each lane approximately 250 ft upstream of the intersection stopline. Detector data every 30 seconds (vehicle count and occupancy) and signal timing data for the study period were obtained from the ATSAC (Automated Traffic Surveillance and Control) central traffic control system database. Floating cars runs were performed at 7 minute headways. Vehicle location and speed were recorded on each second using GPS units. The study period enabled us to obtain data for a wide range of traffic conditions (from low volume off-peak conditions to peak period conditions). For more information about the field study, obtaining comprehensive database of operating conditions to the study area, the reader should refer to [14].

Figure 9: Schematic Diagram of the study site (Lincoln Avenue, Los Angeles)

Data analysis

Traffic demand is high especially during the peak hour. Traffic volumes are heavily directional with the higher volumes in the northbound direction. The average travel speeds on the test section are 25 mph during the off-peak times and drop to about 10 mph during the peak hour in the heavily traveled northbound direction. The proposed methodology applied to the above site using the occupancy and count data to identify the spillovers spatially and temporarily. Also, average speeds per cycle for each link were calculated using the GPS data from floating cars for comparison purposes. The results show that spillovers cause very low observed speeds during the peak hour (7.00-9.00). [Floating cars stopped repetitively during the green phase of the signal.]
Figure 10: Time-series of measured and blocking occupancy at Washington intersection (left axes), average speed of floating cars (right axes)

Figure 10 illustrates (i) values of the measured occupancy from the detectors, (ii) critical occupancy estimated from eq. 5 at Washington intersection and (iii) such average speeds. We observed that floating cars average speeds at this intersection were measured less than 9mph when spillovers identified.

Figure 11: (a) Spillover plot with space and time (b) Occupancy contour plot in the study site
Figure 11a plots different traffic states in the Lincoln site for the 4 hours study period. Green color indicates the existence of spillovers, while blue the opposite. The interesting result is that locations of spillovers are very clearly presented. By considering the effect of spillovers in delays and decrease of mobility, the described method detects the critical congested intersections in a network. An intervention in these critical points could result the avoidance of long queues and the efficient operation during the peak hour. As shown in figure 10b, a simple contour plot of occupancies in the study site is unable to give information about blocking traffic.

5. Final remarks
This paper described and validated a methodology for identifying queue spillovers in city street signalized networks. It also presented the significant effect spillovers have in the system output and efficiency. Ongoing and future research involves the development of strategies for monitoring traffic in congested urban networks. Effective monitoring is essential for developing observation-based control. Avoidance of spillovers in congested parts of a city can lead to better utilization of the system and increase in mobility and accessibility.

Acknowledgments
The author thank Alexander Skabardonis from UC Berkeley for comments and suggestions throughout the study.

References