Fair Allocation of Delays in the Real-Time Control of an Autonomous Traffic Signal

Hui Li
PhD Student
Tongji University, Shanghai, China
Columbia University, New York, NY
hl2615@columbia.edu

Nicholas Maxemchuk
Professor
Columbia University, New York, NY 10027
Institute IMDEA Networks, Madrid, Spain
Phone: +1 212-854-0580
Email: nick@ee.columbia.edu
(Corresponding Author)

Yitian Gu
PhD Student
Columbia University, New York, NY
yg2261@columbia.edu

Guoxia Zhang
PhD Student
Tongji University, Shanghai, China
Columbia University, New York, NY
guoxiafly@gmail.com

Word Count: 6220 words + 3 figures + 2 tables = 7470 words

Submitted for presentation at Transportation Research Board Conference and publication in Transportation Research Record

July 28, 2011
ABSTRACT

Our objective is to use information on the vehicles approaching an intersection to control the traffic signal. Recent work in this area has used this information to minimize the average delay of vehicles controlled by the signal. We demonstrate that, when there are significant differences in the volume of traffic approaching from different directions, the average delay of the vehicles in the low volume directions can be much greater than those in the high volume direction. In order to treat the vehicles more fairly, we define two fairness mechanisms, min-max fairness and proportional fairness. Min-max fairness minimizes the maximum delay of vehicles, and is fair from the perspective of the vehicles, and proportional fairness minimizes the sum of the delays of vehicles controlled by a phase of the traffic signal, and is fair from the perspective of the traffic signal. We compare our fairness mechanisms with a real-time mechanism that minimizes the queue length of vehicles waiting at a traffic signal, and a fixed cycle traffic signal. We perform the comparisons for a range of arrival rates, and for both balanced and unbalanced loads on the approaches to the signal. We find that the min-max fairness mechanism treats the individual vehicles more fairly, but can significantly increase the average delay in comparison with the queue length mechanism. Proportional fairness, however, treats the vehicles as fairly as min-max fairness without significantly increasing the average delay. Based upon this study we recommend proportional fairness for the real time control of traffic signals.

1. INTRODUCTION

1.1 General Description

We introduce a metric that uses local, instantaneous information to control an isolated traffic signal. Using real time information to control traffic signals, instead of averages values, not only corrects differences between the current average values and the measured values, but also takes fluctuations within the average into account. Our control procedure reduces the average delay of vehicles waiting at the traffic light, and also treats the vehicles in different lanes fairly.

Wunderlich (Wunderlich et al., 2008) showed that switching the signal to give preference to the lanes with the most cars waiting can significantly reduce the average delay in comparison to traffic signals that use fixed timing. By applying Little’s law, Wunderlich shows that the average delay is minimized. We will refer to this mechanism as Minimum Queue Control. We show that this rule works well when the traffic in all of the lanes have the same arrival rates, but can result in significantly longer average delays for vehicles in lanes with lower arrival rates.

Minimizing the queue length is clearly unfair to some of the users, however the precise definition of fairness is not obvious. We will define fairness from the point of view of individual vehicles and from the perspective of the traffic signal.

From the point of view of individual vehicles, we consider a traffic signal to behave fairly when the delay caused by the signal does not depend on the arterial or the lane of the vehicle. We will equalize the delays by minimizing the maximum delay experienced by any vehicle. We define a fair system to be one in which we cannot decrease the maximum delay of a vehicle without increasing the maximum delay of a vehicle with a larger maximum delay. We will refer to this mode of operation as Min-Max Fairness. This is analogous to Max-Min fairness for capacity allocation in computer networks (D. Bertsekas and R. Gallager, 1987). While this mode of operation treats the vehicles fairly, we will show that it can increase
the average delay by switching the signals more frequently.

We consider a traffic signal to behave fairly when the inconvenience that it causes waiting vehicles is the same for the vehicles in every phase. We define the inconvenience as the sum of the waiting times of the vehicles at a red light. Our objective is to minimize the sum of the waiting times for each phase. A system is fair when we cannot reduce the sum of the waiting times for a phase without increasing the sum of the waiting times for a phase with a larger sum. We will refer to this type of fairness as Proportional Fairness because the average waiting time for vehicles in a phase is inversely proportional to the average number of vehicles waiting at the traffic signal. There are some similarities between this mechanism and proportional fairness in computer networks (F. Kelly et al., 1998), although the computational techniques are very different.

We show that proportional fairness achieves an average delay that is very close to that achieved by minimum queue control, and that the maximum delay is very close to that achieved by min-max fairness. In minimum queue control, a single vehicle that arrives on a low volume street and tries to cross a high volume avenue may never receive an opportunity to progress. Proportional fairness improves the fairness of minimum queue control because the waiting time of even a single vehicles on a low density street will eventually exceed the sum of the waiting times for the vehicles on the busy avenue. Proportional fairness approaches fairness min-max fairness for individual vehicles because the vehicles with the longer waiting times make a bigger contribution to the sum of the waiting times for their phase.

The paper is organized as follows: In section 2, we describe our model of the intersection, the measurements that we make to control the traffic signal, the assumptions that we make about the operation of vehicles, and the metrics that we use to evaluate the control procedures. In section 3 we describe the implementation of the three real-time traffic light control procedures. Section 4 presents the simulations. We describe how the simulations and measurements are made, the range of arrival rates that we consider, and the structure of the balanced and unbalanced loads.

Section 5 is the conclusion. We summarize the main results and recommend using proportional fairness rather than either minimum queue control or min-max fairness. Our recommendation is based upon the metrics that we use to evaluate the mechanisms and the entire range of simulations that we have conducted.

1.2 Previous work

Traffic signal control has been extensively studied. In 1958, Webster (Webster, 1958) introduced equations to determine optimal signal cycle length and green phase timing of the fixed-time control. While this control mechanism is widely used, the fixed design fails to handle the variations in the traffic flows, and the measurements are frequently outdated. With the advances in inexpensive sensor and wireless communication, significant efforts have been devoted to improve signal control by using more current average data and instantaneous measurements.

In 2008, Wunderlich (Wunderlich et al., 2008) proposed a real time traffic control mechanism to minimize the average waiting time at a traffic signal. His mechanism, called Longest Queue First Maximal Weight Matching (LQF-MWM), minimizes the average delay by minimizing the individual queue lengths. The phases of the traffic signal are actuated to service the longest queues first. By Little’s Law, $L = \lambda W$, where $\lambda$ is the average arrival rate of vehicles at the intersection, $W$ is the average time
that a vehicle spends in the system, and L is the average number of vehicles in the system. Wunderlich argues that minimizing the individual queue lengths at the intersection results in the minimum average delay. This technique is the basis of the minimum queue control mechanism used in tour comparisons. We show that this strategy can treat vehicles unfairly.

Liu (Henry X. Liu, et.al, 2002) presented an adaptive signal control system with an online signal control systems. The timing plan is optimized based on the delay estimated from vehicle re-identification technology.

Shelby (Steven G. Shelby, 2004) conducted a novel simulation study focused on real time computational capabilities of adaptive-control algorithms used in OPAS, PRODYN, ALLONS-D, and COP, the intersection control algorithm of RHODES. He drew a conclusion that certain aspects of algorithmic techniques from OPAC, PRODYN, and ALLONS-D are unsuitable for much more computationally demanding tasks, while specific features of the COP-97 and ALLONS-D algorithms present significant advantages in computational complexity and reduced traffic delay, respectively.

A novel adaptive control logic called PODE was introduced and tested on a single intersection model by Yi (Yi.et.al, 2008). PODE requires real-time input from multiple detectors, and the special features of PODE are flexible interval length and self-adjustment. PODE performs consistently better than fixed control and reduces the average delay by nearly 50% in congested traffic.

Haijema (Haijema et al, 2008) prescribed a control mechanism to determine the appropriate time for the traffic light to switch from green to yellow based on the current state of traffic light and the number of vehicles waiting in every queue. They model the problem as a Markovian decision process. Some other researchers have proposed using fuzzy logic to adjust the green light durations and the phase sequence (Ehsan et al., 2010; Lai Guan Rhung et al., 2009). These techniques yield up to a 70% decrease in the average waiting time compared with fixed control.

In 2005, Gershenson (Gershenson, 2005) proposed several self-organizing traffic control methods. A technique that divides vehicles into platoons decreases the time it takes vehicles to traverse a route by up to 30% when compared with fixed green waves. Gershenson’s techniques use simple local rules at each traffic light, and does not communicate between lights.

Gettman (Douglas Gettman, et.al, 2007) presents a data-driven real-time adaptive control algorithm for tuning traffic signal offsets in a coordinated traffic signal system. The algorithm makes small incremental changes to the offset at each signal by evaluating the amount of traffic that is captured by the green interval of the coordinated phases. Simulation tests have quantitatively shown that this tuning approach can improve arterial progression performance. There are also some other methods that use dynamic programming and rolling horizons, to control and coordinate the network of signals (Porche, et al., 1996; Mirchandani et al., 2001).

2. SYSTEM MODEL

2.1 Intersection configuration

In this evaluation we use a simple intersection with traffic approaching from four directions, on two, two-way streets. Vehicles that arrive at the intersection from each direction can leave in three directions, straight, left or right. The traffic light at the intersection has four phases as shown in Fig.1. On the
east-west street (A, B), during phase I vehicles can go straight or turn right, and during phase II they can turn left. Phase III and phase IV control vehicles on the north-south streets in the same manner.

FIGURE 1 Intersection Model and the four-phases

We will use the four phase model in all of the results presented in this paper. It is possible to consider traffic lights with fewer phases or additional phases. For instance, four additional phases can be defined by allowing the traffic on one approach to move forward, turn right, or turn left, while all of the traffic on all of the other approaches is stopped. Or, two phases can be eliminated by combining the two phases on the north-south and the east-west streets, allowing left turns when the driver feels that it is safe. In addition, we can allow a right turn on a red signal, or use different combinations of rules for the north-south and east-west traffic. Vehicles that are able to turn right on a red signal are removed from the queue. They do not add to the real time metrics that control the signal, and are not included in the statistics that determine the delay and fairness of the signal. Except when there is a separate right turn lane, the effect of right on red is only significant when the traffic density is low, and the signal does not change to accommodate waiting vehicles. While all of the approaches described above are useful in specific traffic conditions, they do not lead to a better understanding of the real-time mechanisms, which can be applied to any combination of phases, and are not considered in this work.

There are several ways to gather the real-time information that is used in the real time procedures. In (Wunderlich et al., 2008), Wunderlich assumes that each vehicle has an in-vehicle information system (IVIS) which can determine the vehicle’s location and communicate with the traffic signal. This technique is accurate, but requires a large investment in new equipment in every vehicle before it can be used at any intersections. In this work, we approximate the real time information by using a small number of sensors located on the roadway. Sensors are located in each of the lanes entering and leaving the
intersection. In Figure 1 there are 16 sensors, indicated as solid lines labeled from (S1,i) to (S8,i), (S1,o) to (S8,o). By collecting information on the time that vehicles pass each sensor, we approximate the queue length and delays in each lane.

We assume that all of the vehicles that enter a lane leave the lane. When vehicles enter or leave parking areas between two sensors, the queue length will either be overestimated or underestimated. In this paper, we will not take this problem into consideration, but can introduce heuristics to correct for occasional changes in the number of vehicles. The number of vehicles in a queue at any point in time equals the number of vehicles that have entered the roadway up to that time, minus the number of vehicles that have departed.

We define the delay that the traffic signal causes for a vehicle as the time that it spends in a roadway approaching an intersection minus the time that it would take to traverse that roadway without being delayed by a traffic light or cross traffic. We assume that vehicles do not pass other vehicles in the queue waiting at the traffic signal. The delay of each vehicle in the queue can be calculated, and the vehicle at the head of the queue has the largest delay of any vehicle in the queue. When there are multiple lanes, the FIFO assumption becomes weaker. The average delay is accurate, but the longest delay may be underestimated.

2.2 Metrics
Both the delays and fairness of the traffic signal control mechanisms are evaluated. We use the average delay, the maximum delay, and the standard deviation of delays to assess the procedures. We evaluate each metric for the intersection, and for each lane entering the intersection. The smaller the value for the intersection, the better the performance of the rule. The difference between the values for different lanes indicates when the system may be unfair. However, the system is only unfair when we have decreased the value for one lane while increasing the value for a lane with a larger value.

For instance, on a two-way street the traffic signal may simultaneously control the traffic in both directions. If the traffic volume in one direction is much greater than that in the other, the delays in the high volume direction will be greater than those in the low volume direction. This is not unfair because we have not increased the delay in the high volume direction to decrease the delay in the low volume direction.

3. REAL-TIME PROCEDURES
The sensors continuously monitor the vehicles entering and leaving each lane. The number of vehicles in a lane at any time, \( t \), is the number of vehicles that have entered that lane up to \( t \), minus the number of vehicles that have left that lane up to \( t \). In addition to the number of vehicles that pass a sensor, we also record the time that the \( j^{th} \) vehicle enters and leaves the lane, \( a_j \) and \( b_j \). If the \( j^{th} \) vehicle is in the lane at \( t \), the delay that it has experienced because of the traffic signal is the \( \max(0, t-a_j, T) \), where \( T \) is the time for a vehicle to drive between the sensors marking the beginning and end of the lane at the legal speed limit. Vehicles that have a delay greater than zero are considered to be waiting in the queue of that phase.

For each lane we calculate the number of vehicles in the queue and the delays of those vehicles. Each phase of the traffic signal provides a green signal for several lanes. We calculate the delay or number of queued vehicles for a phase as the sum of these values for all lanes that receive a green signal during that
phase. For instance, in phase I we calculate the values for two lanes, the lanes that go straight or make a right turn in both directions on the east-west street.

We give a green signal to the phase with the largest value, where the value that we use depends on the rule that we are using to control the signal. Once we change the traffic light, we keep it green for a minimum amount of time, as long as there are vehicles approaching a green signal. We keep the light green for a minimum amount of time, even though the value that controls the signal indicates that it should be changed because the throughput of the intersection is reduced when the light changes. However, if there aren’t any vehicles approaching the intersection in the lanes with a green signal during the minimum green interval, and there are vehicles waiting in the other lanes, then the signal is changed.

The three values that we investigate to control the traffic signal are:

1. N: the number of vehicles that have entered the queues corresponding to a phase,
2. M: the maximum delay of any vehicle in the lanes corresponding to a phase, and
3. D: the sum of the delays of all the vehicles in the lanes corresponding to a phase.

The first value is Wunderlich’s metric. The second value turns the signal green for the vehicle with the longest delay. This is the metric that we use to approximate min-max fairness for individual vehicles. The third value is an approximation of proportional fairness.

The values are only approximations to the fairness metrics because they do not consider the increased delay caused changing the light or the effect on vehicles that have not yet been delayed. There are times when switching using the second value will increase the maximum delay, or switching using the third value may increase the inconvenience caused by the traffic signal, as will be noted in the simulations. We can construct more complex functions that include the cost of switching the traffic signal and the effect on traffic that has not entered the queue, and may obtain additional decreases in the average delay and improvements in fairness. However, the simple metrics that we are using provide significant improvements and provide an understanding of the reason for the improvements.

4. SIMULATIONS

We use the Green Light District (GLD) simulator to compare the real time traffic light control algorithms with each other, and with a fixed control strategy. The GLD simulator is an open source simulator that was developed by the Intelligent Systems Group at the University of Utrecht (Cools et al., 2007; Wiering et al., 2004). The simulator is written in JAVA, and is easily modified to model the new control procedures that we are investigating.

The traffic on each street arrives at a specified rate, according to a Poisson process, and each vehicle turns left with a specified probability. In the simulations we consider both balanced loads, with the same arrival rates on the north-south and east-west streets, and unbalanced loads with higher arrival rates on one of the streets. We have elected not to consider traffic imbalances from opposite directions on the east-west or north-south street. This type of imbalance would effect a traffic signal with 8 phases, but not with 4. The differences between the delays in the two directions does not result in unfairness, as noted in section 2.2. In the simulations we pick arrival rates that create low utilization, with small waiting lines, moderate utilization, and high utilization, that approach the capacity of the intersection.

We randomly select a fraction of the vehicles arriving on each street to turn left. In the simulations, 10% of the cars turn left.
The fixed control strategy uses a 90 second cycle time. The fraction of the cycle time assigned to each phase is proportional to the fraction of the arrivals in that phase. For instance, if 40% of the arrivals are on the east-west street and proceed straight or turn right, then the light is green for phase I for 36 seconds in each 90 second cycle. According to Webster’s equation (Webster, 1958), the optimum cycle length is a function of lost times and flow ratios. The delays experienced by the fixed cycle strategy can be reduced by adjusting the cycle time with the load. The optimum cycle time is designed to minimize the average delay, and increases with increasing load. This increases the average delays for low utilization phases more than for high utilization phases, and increases the unfairness with increasing load. We elected to keep the cycle time fixed. This does not minimize the average delay, but also does not increase the unfairness.

Each time the traffic signal changes, there is a period of time when no vehicles pass through the intersection, which decrease the throughput of the intersection. In the simulations we have set the lost time to 2 seconds (Furth, P., B. Cesme, T. Muller, 2009). As rate of the phase changes increases, the fraction of the time that is lost also increases. In the real time procedures once a light turns green, it remains green for 10 seconds, unless there are no cars approaching the green light, and there are cars waiting for one of the other phases.

Each point in the results that are reported is based upon 20 simulation runs. In each run we operate the system for 3,600 seconds before starting to take statistics, so that the queues, which are initially empty, stabilize. We then take measurements for 12,000 seconds.

4.1 Simulation of balanced traffic loads

4.1.1 Scenario description

In this section the arrival rates on all four approaches to the intersection is the same. The traffic load is balanced. The arrival rate is normalized to the rate at which cars can traverse the intersection if there is no traffic signal, the vehicles travel at the legal speed limit, and the separation between vehicles is the recommended safe distance. With a balanced load, the fraction of the time that the signal is green for the two phases of the east-west street, or the two phases of the north-south street is $\frac{1}{2}$. Therefore, the normalized arrival rate on any street is at most 0.5. Otherwise, more vehicles will arrive than can traverse the intersection, and the queues will grow indefinitely. The arrival rate into each street must actually be less than 0.5, because there are intervals between phases when no vehicles cross the intersection.

We plot the average delay, the maximum delay, and the standard deviation of the delay for the three real-time procedures and for the traffic signal with a fixed cycle time. In both the balanced and unbalanced simulations we assume that the fixed control procedure knows the arrival rates on all of the streets. In a realistic system, where these rates are continuously changing, the real-time procedures respond to the changing rates, and the improvements over the fixed procedures will be greater.

4.1.2 Average delay

In figure 2(a) we plot the average delay of vehicles as a function of the arrival rate, for the three real-time procedures and for fixed cycle traffic signals. We note that at low to moderate utilizations, corresponding to arrival rates between 0.1 and 0.3, all of the real time schemes are nearly equal, and reduce the average delay of fixed strategies by as much as an order of magnitude. The reason for the improvement is that the fixed signal will force vehicles to wait at a red light when there aren’t any
vehicles passing through the green light. When no cars are approaching a green signal, all of the real-time procedures immediately provide a green signal to the waiting vehicles.

**FIGURE 2** Simulation results in balanced traffic load
As the load increases and the utilizations increase from 0.35 to 0.45, the average delay of min-max fairness increases, and is greater than that for the fixed signal. Under high loads, there are vehicles waiting in all of the phases, and the vehicle at the head of the queue has the longest delay. When that vehicle leaves the queue, a vehicle at the head of another queue may have the longest delay. The rule prevents the signal from changing for 10 seconds, but the signal changes as frequently as possible, and wastes 2 seconds with each change. By contrast, minimum queue control has a 20-30% improvement over the fixed control strategy at the highest load. Proportional fairness also provides an improvement over the fixed strategy over the entire range of arrival rates.

4.1.3 Maximum and standard deviation of delay

Plots 2(b) and 2(c) for the maximum and standard deviation of the delay expose the problem with minimum queue control. When we minimize the average delay, the maximum delay exceeds that of the fixed procedure over the entire range of arrival rates, and the standard deviation of the delays exceeds that of the fixed signal at high utilizations. The reason that minimum queue control behaves badly according to these metrics is that there are only 10% as many vehicles turning left as proceeding straight or turning right. Since we are switching based on the number of vehicles that have accumulated at the red signal, the phases corresponding to left turns take 9 times as long to accumulate the same number of vehicles, and the vehicles in these phases must wait much longer. In an extreme case a single vehicle may arrive on a side street and may never get a chance to cross a busy thoroughfare.

By contrast, both min-max fairness and proportional fairness have smaller maximum and standard deviation of delay than the fixed procedure. Min-max fairness has a smaller maximum delay than proportional fairness, as expected, but the standard deviation of delay is very close for both procedures, and proportional fairness has a smaller standard deviation at high loads, where the average delay of min-max fairness becomes large.

4.1.4 Overall comparison and analysis

Based on these results, proportional fairness is the most promising procedure for balanced loads. Its average delay is very close to minimum queue control, its maximum delay is very close to min-max fairness, and its standard deviation of delay is close to the best of the other three mechanisms. Proportional fairness provides significant improvements, up to an order of magnitude improvement, over the fixed traffic signal for all three metrics for utilizations ranging from very low, 20% of the maximum throughput, to very high, close to 100% of the maximum throughput of the intersection.

4.2 Simulation of unbalanced traffic load

4.2.1 Scenario description

In order to determine the effect of unbalanced loads on the real-time algorithms we have considered situations where 75% of the traffic arrives on the east-west street and 25% on the north-south street, and also when 90% of the traffic arrives on the east-west street and 10% on the north-south street. As before, 10% of the traffic arriving at the intersection makes a left turn. We compare the average delay, the maximum delay, and the standard deviation of delay of the three real-time procedures with a fixed traffic light controller with a 90 second cycle time that is tuned to the arrival ratios. Note that the time allocated for left turns on the low density street with the 9:1 ratio is only about 1 second. In a real system, the small number of vehicles making left turns on this street would not be assigned a phase of the traffic light, but
would make the turns when it is safe. We use the four phase model in this case for consistency with the other results.

We perform the evaluation at low, moderate and high loads. The throughputs are 20%, 50%, and 80% of the maximum throughput that can be achieved if there weren’t any lost time when the traffic signals change. The arrival rates in each direction on the streets are listed in table 1:

<table>
<thead>
<tr>
<th>Unbalance ratio</th>
<th>Traffic density</th>
<th>Arrival rate of EW</th>
<th>Arrival rate of NS</th>
</tr>
</thead>
<tbody>
<tr>
<td>3:1</td>
<td>Low</td>
<td>0.15</td>
<td>0.05</td>
</tr>
<tr>
<td>3:1</td>
<td>Medium</td>
<td>0.375</td>
<td>0.125</td>
</tr>
<tr>
<td>3:1</td>
<td>High</td>
<td>0.6</td>
<td>0.2</td>
</tr>
<tr>
<td>9:1</td>
<td>Low</td>
<td>0.18</td>
<td>0.02</td>
</tr>
<tr>
<td>9:1</td>
<td>Medium</td>
<td>0.45</td>
<td>0.05</td>
</tr>
<tr>
<td>9:1</td>
<td>High</td>
<td>0.72</td>
<td>0.08</td>
</tr>
</tbody>
</table>

4.2.2 General findings

Most of the results for unbalanced loads are expected from the discussion of balanced loads. Minimum queue control treats the vehicles on low density streets unfairly in order to lower the average delay of all of the vehicles. Min-max fairness switches the signal too often under heavy loads, wastes time, and causes an increase in the average delay. Proportional fairness provides low average delay without treating the vehicles on the low density streets as unfairly as minimum queue control, and is the recommended procedure.

The unexpected result is the peak in the maximum delay for minimum queue control in figure 3(e). Looking into the simulation more carefully provided an explanation for the peak. The maximum delay occurs for vehicles making a left turn on the low density street. At low loads, vehicles in this lane turn when there are no vehicles in the lanes corresponding to the other phases, which results in low delays. At moderate loads there are almost always vehicles in the other lanes, but the arrival rate in this lane is very low, and the maximum of the delays to accumulate the necessary number of vehicles in this lane can be very large. At heavy load, the vehicles in this lane build up more quickly, and the time to acquire the signal is more consistent.

4.2.3 Fairness analysis

In table 2, we note the difference in average delays for the cars that have a green light during the four phases of the traffic signal, for low moderate and heavy loads, when 75% of the traffic is on the east-west road. Tables for the 90/10 split provide similar conclusions and are not included here.

Phase I turns the signal green for the 67.5% of the cars on the east-west road that proceed straight or turn right. Phase II is the green signal for the 7.5% of the cars on the east-west road that turn left. Similarly, Phase III is the green signal for the 22.5% of the cars on the north-south road that go straight or turn right, and Phase IV is the green signal for the 2.5% of the cars on the north-south road that turn left.

We note that when we use minimum queue control there are very large differences between the average delays of the vehicles in the different phases. For instance, at low utilizations, scenario 1, the average delay of vehicles in phase IV is more than 20 times longer than the average delay in phase I. The arrival rate in phase I is 27 times that in phase IV. At high utilizations, scenario 3, the average delay in phase IV is 11.7 times that in phase I, showing that the unfairness decreases when the arrival rate in the low volume phases increases.
FIGURE 3 Simulation results in unbalanced traffic load
### TABLE 2 Simulation Results of 3:1 unbalanced load

<table>
<thead>
<tr>
<th>Control Method</th>
<th>Intersection Results</th>
<th>Average Delay of Each Phase</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Maximum delay(s)</td>
<td>Std of delay</td>
</tr>
<tr>
<td>Fixed-control</td>
<td>146</td>
<td>20.1</td>
</tr>
<tr>
<td>Minimum Queue Control</td>
<td><strong>352</strong></td>
<td>12.6</td>
</tr>
<tr>
<td>Min-Max Fairness</td>
<td>19</td>
<td>3.2</td>
</tr>
<tr>
<td>Proportional Fairness</td>
<td>24</td>
<td>3.1</td>
</tr>
</tbody>
</table>

**Scenario 1: Low Utilization**

For fixed control, at low utilizations the average delay for phase IV is 5.6 times greater than the average delay of phase I. This occurs because there is a longer wait between green signals in phase IV than in phase I, 87.75 seconds as compared with 29.25 seconds, so that cars in phase IV must wait longer. If the cycle time is reduced, the absolute value of the delay would decrease, but the ratio would remain the same, since the red/green ratio is determined by the traffic volumes. The difference between the average wait at low and high utilizations remains about the same. The average delay for fixed control is 3.7 times greater than the average delay for minimum queue control at low utilizations and only 1.4 times as large at high utilizations. This shows the need for using real time control procedures to prevent vehicles from waiting at a red signal when there aren’t any vehicles passing through the green light.

Min-max fairness provides the smallest difference between the average delays for the different phases, although phase IV still has an average delay that is 2.5 times larger than phase I at low utilizations. Proportional fairness is not much worse than the min-max fairness, with a ratio of about 3 for the same comparison. At high utilizations the average delay for min-max fairness is about 44% greater than the average delay for minimum queue control, while proportional fairness is only 1% greater.

The unfairness of the different control procedures is compared by the maximum and standard deviation of the delays. The maximum delay of minimum queue control is 18.5 times greater than the min-max fairness at low utilizations and 8.5 times greater at high utilizations. The standard deviation of the delay is

| Fixed-control          | 208                  | 23.5                       | 23.5             | 14.0 | 55.4 | 34.9 | 78.3 |
| Minimum Queue Control  | **515**              | 26.5                       | **16.9**         | **10.0** | 43.0 | 17.2 | **116.7** |
| Min-Max Fairness       | 61                   | 12.4                       | 24.2             | 25.3 | 23.6 | 20.5 | **29.2** |
| Proportional Fairness  | 161                  | 12.4                       | **17.1**         | **14.7** | 28.3 | 17.6 | **46.3** |
also 3.9 and 2.1 times greater for the same comparisons. Fixed control also behaves unfairly, with a standard deviation of delay that is 6.3 times greater than min-max fairness at low utilizations, and 1.9 times greater at high utilizations. By contrast, the standard deviation of the delay for proportional fairness is within 0.1 of that for min-max fairness at both low and high utilizations.

5. CONCLUSIONS

In this paper, we propose a new real time mechanism to control a traffic signal, called proportional fairness. We compare this mechanism with a real-time control mechanism that has been proposed to minimize the average delay experienced by vehicles, a fixed traffic control mechanism that is tuned to the average arrival rates on the streets, and a second real-time mechanism that we introduced to minimizes the maximum delay experienced by any vehicle, called min-max fairness. The comparisons are performed at a range of arrival rates varying from lightly used intersections, where there are very few vehicles waiting at red lights, to heavily used intersections, where the queues of waiting vehicles are about to increase indefinitely. The comparisons are performed when the traffic on the cross streets are about the same, and when one of the cross streets has significantly more traffic than the other.

All of the real-time mechanisms decrease the average delay of the fixed control procedure at low utilizations. At low utilizations fixed control can be wasteful by forcing vehicles to wait at a red light while there are no vehicles going through the green light. At low utilizations, all of the real time control procedures switch the signal so that one of the phases with vehicles waiting has a green light. Real time procedures that turn a light green on low volume intersections when there are vehicles waiting are becoming common.

Minimum queue control decreases the average waiting time of fixed signals over the range of utilizations that we have studied, but can treat some vehicles unfairly, particularly when there is a significant difference between the arrival rates of the different phases.

We introduced min-max fairness to treat individual vehicles more fairly. Whenever we can change the traffic signal, the vehicle that has been waiting longest goes first. This procedure reduces both the maximum delay and the standard deviation of the delay when compared with both minimum queue control and fixed control. However, it increases the average delay, particularly at high utilizations, by switching the phases of the signal more frequently.

We introduced a new fairness procedure, called proportional fairness, which does not consider fairness from the point of view of individual vehicles, but from the point of view of the traffic signal. It equalizes the sum of the delays of the vehicles waiting in any phase. The vehicles in a phase with fewer vehicles may wait longer, but a single vehicle waiting at a red signal will eventually accumulate a longer delay than a larger number of vehicles in a more heavily travelled lane.

Proportional fairness preserves the best characteristics of both of the other real-time control mechanisms. It decreases the average delay close to the levels of minimum queue control and has a standard deviation of delay that is close to that of min-max fairness. As a consequence, proportional fairness is the preferred mechanism of those mechanisms studied, especially under unbalanced traffic load.

Acknowledgements:
This work is funded at Columbia University by the National Science Foundation. It is an interdisciplinary grant with Satish Ukkusuri at Purdue University. Ms Li and Ms. Zhang are supported by the China Scholarship Council.

This work builds on a number of earlier projects that have been conducted at Columbia University by Hyunwoo Nam, Haluk Celebi, Fan Zhang, Xueyan Zheng, and David Englehardt.

References:


