A Mathematical Model for Bus Scheduling with Conditional Signal Priority

Ming Liu, Yecheng Zhao, Feifeng Zheng, Chengbin Chu

A mathematical model for bus scheduling with conditional signal priority

Ming Liu[1], Yecheng Zhao[1], Feng Chu[2⋆], Feifeng Zheng[3], and Chengbin Chu[4]

[1] School of Economics & Management, Tongji University, Shanghai, People's Republic of China
[2] IBISC, Univ Évry, University of Paris-Saclay, Évry, France
feng.chu@univ-evry.fr
[3] Glorious Sun School of Business & Management, Donghua University, Shanghai, People's Republic of China
[4] Laboratoire d'Informatique Gaspard-Monge (LIGM), UMR 8049, Univ Gustave Eiffel, ESIEE Paris, 93162 Noisy-le-Grand Cedex, France

Abstract. Inaccuracy of buses is a common situation. A common practice is that a certain amount of slack is usually added to the schedule of bus operation, so that the bus can execute the schedule in most cases. On the other hand, slack means that buses sometimes have to wait for a while at the station or slow down while driving. Since the bus cannot accelerate the driving process by itself, this method cannot make the bus fully implement the schedule. Researchers invented the Transit signal priority (TSP) and conditional signal priority (CSP), the purpose of which is to give the bus signal priority to speed up when it is delayed to a certain extent. Some previous work has studied the driving process of buses with CSP. However, there is still room for further improvement in the mathematical description of the bus driving process based on CSP. In this article, we analyze the driving state of the bus under different CSP states, that is, positive and negative. Then a series of representative and operational assumptions are given. These assumptions can be used as the basis for future research on such topics. With the assumptions, we give a mathematical model of the bus driving process using CSP. According to some performance indicators of the bus driving process obtained in the modeling process, an optimization goal is established to comprehensively improve the driving effect of the bus. Mathematical analysis and numerical solution verify the applicability of the model.

Keywords: Bus scheduling· Signal priority· Bus driving process· Mathematical model.

1 Introduction

As a kind of public transportation, buses are expected to have high schedule reliability. But due to many reasons, bus delays are common occurrences. In order

⋆ Corresponding author.

to solve this problem, decision makers often leave a certain slack when designing the bus schedule to ease it. This has caused a waste of traffic resources. In today's era where efficiency and resource conservation are increasingly concerned, the optimization of the bus driving process has become a valuable topic.

Setting a certain amount of slack for bus driving is a common way to adjust the pace of the bus to meet the schedule. In this way, the bus will face a situation where slack is useless. Many methods are used to control the driving process of buses. The classic approach involves setting a station as control point so that the bus can stay at the station for a period of time [7]. It is true that this intentional deceleration will slow down the entire driving process. But this approach does allow most buses to reach control point according to schedule.

The pursuit of efficiency makes it impossible to set slack too much. And this kind of setting cannot deal with unexpected events during driving, such as traffic accidents on the route or sudden natural disasters. Some studies take these conditions into consideration and introduce different modes of driving with failure of schedule. Given the concept of forward headway, which is the distance between current bus and the bus in front, some research study how long the bus is held based on the forward headway [3]. Another research concerns the backward headway between current bus and the behind one [5]. There are also some related papers that combine the two headway to consider, such as [4,8].

Transit signal priority (TSP) is a means used to speed up the process of public transport. Since it is difficult for a bus to recover from a late arrival by itself, TSP is an effective method to assist the bus to recover to the schedule. When the bus calls for TSP, the traffic lights will help it continue driving without stopping and waiting for the red light to end. While TSP is effective for buses, it will inevitably have a certain impact on traffic. Therefore, reducing the impact of TSP on the transportation system has also received attention. One idea is to use TSP only when the bus is late to a certain extent, in order to reduce the negative impact of TSP on the transportation system [6]. And the conception is called conditional signal priority (CSP). Other studies, such as [2], not only consider the application of CSP to request traffic signal priority to speed up the bus, but also use CSP to slow down the early bus by adding additional red lights. But the latter behavior can also be replaced by letting the bus stay at the station for a period of time, which seems to have a smaller impact on the transportation system. Anderson et al. [1] study the impact of CSP on bus reliability. Their research considered three situations, namely no holding, holding by schedule and holding by headways without schedule. Their results show that CSP can not only improve the speed of buses, but also improve the reliability of buses.

2 Problem Description and Assumptions

In this section, we analyze the driving mode of the bus according to the traffic signal priority request or without traffic signal priority request, and describe the process of the bus driving according to the CSP as an abstract process through a series of clear assumptions.

2.1 Problem Description

Suppose there is a bus driving on an infinitely long route. For example, the route of a bus is a loop, and there is enough power to drive more loops. In theory, the bus has two driving modes, one is driving at a relatively low speed V_u, and the other is driving at a relatively high speed V_c realized by requesting signal priority. The unit of speed here is km/s. In this study, we do not consider setting up any holding at stations, which means that high speed and low speed can be achieved by relying solely on the request signal light or not. Correspondingly, there are two kinds of average paces for buses, τ_u and τ_c (s/km). It can be known that $V_u = \frac{1}{\tau_u}$ and $V_c = \frac{1}{\tau_c}$. Obviously $\tau_u > \tau_c$. There is also a schedule-based pace called τ_s. In order for the schedule to be useful to the bus,

$$\tau_u > \tau_s > \tau_c,$$

otherwise the bus will not be able to achieve τ_s regardless of whether the bus is driving at a high speed or a low speed.

In order to prevent from requesting signal priority when the bus is slightly late, we artificially set a gap, namely δ. At the station x, if and only if the lateness caused by low speed is larger than δ, signal priority is requested and the bus switches to high speed. The requesting continues until lateness reached $-\delta$.

We ignore the influence between buses, including the trend of aggregation, so that we regard each bus as independent. When a bus wants to switch CSP, it can switch immediately, regardless of whether other buses need CSP. Suppose that distance between all adjacent stations is equal and each station is equipped with a traffic light. Furthermore, we assume that the entire driving process passes through an infinite number of stations, and the distance between each station is infinitely small. The entire bus driving process is therefore considered a continuous control process and distance x is assumed as continuous. Also, we ignore the time that the bus stops at the traffic lights and stations. Switching of CSP only affects the bus by switching the average pace of driving, that is, τ_u and τ_c.

The difference between actual arrival time and schedule is called lateness, written as $\varepsilon(x)$. The expectation of lateness per unit time is called drift, written as m. It is conceivable that if m is not 0, then as x increases, the absolute value of lateness drifts larger. If the traffic signal priority is not requested during the whole journey (NSP), the drift is m_u. If a traffic signal priority is requested (TSP), the drift is m_c. So the drift under NTP is

$$m_u = \tau_u - \tau_s > 0,$$

and the drift under TSP is

$$m_c = \tau_c - \tau_s < 0.$$

As mentioned earlier, the bus has only two driving modes, namely low speed $V_u = \frac{1}{\tau_u}$ and high speed $V_c = \frac{1}{\tau_c}$. When the bus departs from the departure station, if it is not the scenario that lateness is positive and greater than δ

which means $\varepsilon > 0$ and $\varepsilon = \delta$, the signal priority is not requested, that is, the CSP is negative. The bus is therefore running at a low speed V_u. With the assumption that x is continuous, the bus can immediately change its driving mode by switching CSP when it finds that the lateness is large enough. When the first occurrence of $\varepsilon > 0$ and ε is equal to the upper limit of lateness gap δ, a signal priority is requested, that is, CSP turns positive. CSP remains positive until the lateness of the bus is negative due to high-speed driving and its absolute value is greater than $-\delta$. That is, when $\varepsilon < 0$ and ε touches the lower limit of lateness gap $-\delta$, CSP turns negative. We call this duration between two CPS shifts a period. Until the next time $\varepsilon > 0$ and $\varepsilon = \delta$, repeat to request signal priority and the two periods occur alternately. If the scenario when the bus departs from the departure station is $\varepsilon > 0$ and $\varepsilon = \delta$, the bus starts with CSP positive and then acts the same as the former scenario. In both scenarios, the two periods occur the same number of times during a long journey.

2.2 Detail of Assumptions and Analysis

Based on the above analysis, the following assumptions are given.

Assumption 1: No holding at stations.

Assumption 2: The route that the bus travels is infinitely long with infinite number of stations.

Assumption 3: No influence between buses, which means that the bus is independent.

Assumption 4: Switching of CSP only affect the bus by switching the average pace of driving, that is, τ_u and τ_c.

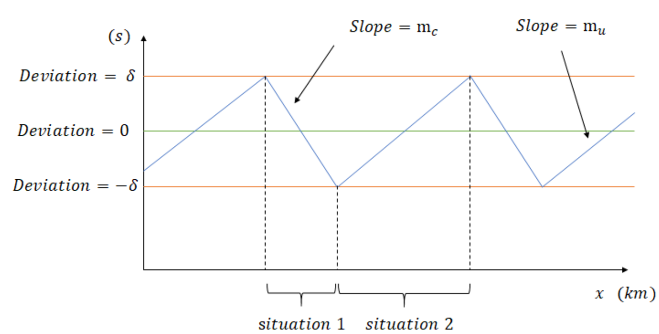

Fig. 1. Bisic driving process of the bus with CSP

Although these assumptions represent a certain abstraction of reality, they also have practical significance. For example, Assumption 2 is derived from the fact that some buses return after reaching the destination without additional rest, and some buses travel in loop routes. Without Assumption 2, the analysis and modeling of buses under CSP will be difficult. Assumption 2 is also reflected in previous research [1]. With the assumptions, the basic driving process of the bus is shown in the figure. In the actual bus driving process, there will always be behaviors such as acceleration, deceleration and stopping. Here, straight lines are used to represent the average driving of the bus under two situations.

The distance of situation 1 that CSP is positive is

$$S_c = \frac{2\delta}{-m_c},$$

and duration time of situation 1 is

$$T_c = \frac{S_c}{V_c} = \frac{S_c}{1/\tau_c} = -\frac{2\delta\tau_c}{m_c}.$$

The distance of situation 2 that CSP is negative is

$$S_u = \frac{2\delta}{m_u},$$

and duration time of situation 2 is

$$T_u = \frac{S_u}{V_u} = \frac{S_u}{1/\tau_u} = \frac{2\delta\tau_u}{m_u}.$$

The duration time of a cycle of two situations is

$$T = T_c + T_u = -\frac{2\delta\tau_c}{m_c} + \frac{2\delta\tau_u}{m_u} = 2\delta\left(\frac{\tau_u}{m_u} - \frac{\tau_c}{m_c}\right).$$

Therefore, the number of times that CSP is switched to positive per unit time, that is, the frequency that CSP is switched to positive is

$$\omega = \frac{1}{T} = \frac{1}{2\delta\left(\frac{\tau_u}{m_u} - \frac{\tau_c}{m_c}\right)}.$$

In practice, from the customer's point of view, the customer hopes that the bus can arrive at stations on time so that the bus can provide better service. In other words, δ is expected not to be too big.

At the same time, from the perspective of entire traffic, if the frequency that CSP is switched to positive is too high, the traffic will be frequently disturbed, which is an abuse of signal priority. In other words, we hope ω is not too big.

The traffic managers also do not want the distance with CSP positive too long, which is also an abuse of traffic signal priority. On the other hand, too short distance with CSP positive is a waste of CSP resource. It is a parameter specified according to people's needs that the fraction of distance where CSP is positive on the route. Here we specify that the ideal fraction is $\frac{1}{2}$. The degree of deviation of the fraction from $\frac{1}{2}$ can be expressed by the following formula.

$$\begin{aligned}\rho &= \left(\frac{S_c}{S_c + S_u} - \frac{1}{2}\right)^2 = \left(\frac{\frac{2\delta}{-m_c}}{\frac{2\delta}{-m_c} + \frac{2\delta}{m_u}} - \frac{1}{2}\right)^2 \\ &= \left(\frac{m_u}{m_u - m_c} - \frac{1}{2}\right)^2 = \left(\frac{m_u + m_c}{2\left(m_u - m_c\right)}\right)^2\end{aligned}$$

The form of the quadratic function makes that the distance of CSP positive deviate more from $\frac{1}{2}$, the more obvious the penalty is.

Based on the above analysis, the goal we need to optimize is the combination of lateness gap δ, the frequency of CSP is switched to positive ω, and the fraction of distance where CSP is positive. Here, we set the objective function as the weighted sum of the three as below.

$$\begin{aligned} F =& \alpha\delta + \beta\omega + \gamma\rho \\ =& \alpha\delta + \frac{\beta}{2\left(\frac{\tau_u}{m_u} - \frac{\tau_c}{m_c}\right)}\frac{1}{\delta} + \gamma\left(\frac{m_u + m_c}{2\left(m_u - m_c\right)}\right)^2 \\ =& \alpha\delta + \frac{\beta}{2\left(\frac{\tau_u}{\tau_u - \tau_s} - \frac{\tau_c}{\tau_c - \tau_s}\right)}\frac{1}{\delta} + \gamma\left(\frac{\tau_u + \tau_c - 2\tau_s}{2\left(\tau_u - \tau_c\right)}\right)^2 \end{aligned}$$

The whole problem can be described as

$$\begin{aligned} \min_{\delta, \tau_s} \quad F = \alpha\delta + & \frac{\beta}{2\left(\frac{\tau_u}{\tau_u - \tau_s} - \frac{\tau_c}{\tau_c - \tau_s}\right)}\frac{1}{\delta} + \gamma\left(\frac{\tau_u + \tau_c - 2\tau_s}{2\left(\tau_u - \tau_c\right)}\right)^2 \\ & s.t. \quad \delta > 0, \\ & \tau_c < \tau_s < \tau_u. \end{aligned} \tag{1}$$

Here we use three parameters α, β and γ to control the weights of θ, ω and ρ, and all three parameters are positive. It is worth mentioning that since θ, ω and ρ have different units respectively, the three parameters α, β and γ may have different orders of magnitude. The specific value of each parameter needs to be determined according to the specific situation.

3 Analysis and solution of the optimization problem

We can find the first partial derivative of F with respect to δ as follows.

$$\frac{\partial F}{\partial \delta} = \theta - \frac{1 - \theta}{2\left(\frac{\tau_u}{\tau_u - \tau_s} - \frac{\tau_c}{\tau_c - \tau_s}\right)}\frac{1}{\delta^2}$$

The second partial derivative of F with respect to δ is

$$\frac{\partial^2 F}{\partial \delta^2} = \frac{\partial}{\partial \delta}\left(\frac{\partial F}{\partial \delta}\right) = \frac{1 - \theta}{\frac{\tau_u}{\tau_u - \tau_s} - \frac{\tau_c}{\tau_c - \tau_s}}\frac{2}{\delta^3}.$$

According to these two preconditions that $0 < \theta < 1$, $\delta > 0$ and $\frac{\tau_u}{\tau_u - \tau_s} - \frac{\tau_c}{\tau_c - \tau_s} > 0$, it always exists that $\frac{\partial^2 F}{\partial \delta^2} > 0$. That is to say, $\delta = \sqrt{\frac{1-\theta}{2\theta\left(\frac{\tau_u}{\tau_u - \tau_s} - \frac{\tau_c}{\tau_c - \tau_s}\right)}}$ is the only minimum for F.

$\frac{\partial F}{\partial \tau_s}$ can be calculated as

$$\begin{aligned}\frac{\partial F}{\partial \tau_s} &= -\frac{\beta}{2\left(\frac{\tau_u}{\tau_u-\tau_s}-\frac{\tau_c}{\tau_c-\tau_s}\right)^2}\frac{1}{\delta}\frac{\partial}{\partial \tau_s}\left(\frac{\tau_u}{\tau_u-\tau_s}-\frac{\tau_c}{\tau_c-\tau_s}\right)+\gamma\frac{2\tau_s-\tau_u-\tau_c}{\left(\tau_u-\tau_c\right)^2}\\
&= -\frac{\beta}{2\delta}\frac{1}{\left(\frac{\tau_s(\tau_c-\tau_u)}{(\tau_u-\tau_s)(\tau_c-\tau_s)}\right)^2}\left(\frac{\tau_u}{\left(\tau_u-\tau_s\right)^2}-\frac{\tau_c}{\left(\tau_c-\tau_s\right)^2}\right)+\gamma\frac{2\tau_s-\tau_u-\tau_c}{\left(\tau_u-\tau_c\right)^2}\\
&= -\frac{\beta}{2\delta}\frac{\tau_u\left(\tau_c-\tau_s\right)^2-\tau_c\left(\tau_u-\tau_s\right)^2}{\tau_s^2\left(\tau_c-\tau_u\right)^2}+\gamma\frac{2\tau_s-\tau_u-\tau_c}{\left(\tau_u-\tau_c\right)^2}.\end{aligned}$$

Since the partial derivative of F with respect to τ_u has a complicated form, we use the YALMIP solver to find its numerical solution.

Here we set the values of the parameters as $\alpha = 1$, $\beta = 10000$ and $\gamma = 200$. And we set τ_c=90 s/km and τ_u=144 s/km that correspond to the speed of 40 km/h and 25 km/h. The best delta is 23.6182s and the best τ_s is 119.7508 s/km. F is 47.7554 according to them. In order to speed up the solving process, we change the constraints in Equation (1) to the following three constraints.

$$\begin{cases}1 \le \delta \le 1000\\ \tau_s \ge \tau_c + 1\\ \tau_s \le \tau_u - 1\end{cases}$$

These three constraints are used to replace the original strict inequality constraints to make the solver easier to run. At the same time, although we have restricted δ, given that the two limit values, namely 1 second and 1000 seconds, are too small or too large to appear in practical applications. The restriction on τ_s is also in line with our expectation that the journey with CSP positive will not be too close to τ_u or τ_c.

Our analysis and constraints processing are enlightening, and the results obtained can also reflect certain physical meaning and application value. The lateness gap from tens to hundreds of seconds is reasonable for buses. That is, it will not deviate too much from the schedule to affect the satisfaction of passengers, and it can also make the bus not need to request signal priority too frequently. The reasonable value of τ_u also effectively controls the proportion of the distance of the request signal priority within a reasonable range.

This result can be further improved by more information about real bus conditions and more adjustments to the above parameters.

4 Conclusion

The non-punctuality of buses has always been a phenomenon that the industry and academia are trying to change. In order to allow buses to execute schedules more stably, schedule designers usually add some slack to schedules. This results in buses sometimes having to wait at the station or slow down while driving.

But this approach does not make the bus fully implement the schedule. Transit signal priority (TSP) and its conditional form CSP are invented to accelerate the bus in its driving process.

There is still room for further improvement in the mathematical description of the process of bus driving in accordance with the CSP. In this article, we describe this process in detail. We describe the process of bus driving according to CSP as a continuous control process through a series of assumptions, and establish our bus driving process model. These assumptions can be useful to the basis for future researchers. We analyze the process of bus acceleration due to CSP positive and the process of deceleration due to CSP negative, and describe the process of bus driving as alternating acceleration and deceleration. Through analysis, we have obtained the period and frequency of the bus switching driving state. Then an objective function that takes into account gap of lateness, frequency of switching and the fraction of distance where CSP is positive is established. Through mathematical analysis and numerical solution, we verified the feasibility of the model.

Future work includes improving the expression of different parts of the objective function to better serve practical applications, and introducing other bus driving indicators that people pay attention to in the objective function.

Acknowledgement. This work was supported by the National Natural Science Foundation of China (NSFC) under Grants 72021002, 71972146, 71771048, 71432007, 71832001 and 72071144.

References

1. Anderson, P., Daganzo, C.F., Mannering, F.: Effect of transit signal priority on bus service reliability. Transportation Research Part B: Methodological **132**, 2–14 (2020)
2. Chow, A.H.F., Li, S., Zhong, R.: Multi-objective optimal control formulations for bus service reliability with traffic signals. Transportation Research Part B: Methodological **103**(sep.), 248–268 (2017)
3. Daganzo, C.: A headway-based approach to eliminate bus bunching. Transportation Research Part B: Methodological **43**(10), 913–921 (2009)
4. Daganzo, C.F., Pilachowski, J.: Reducing bunching with bus-to-bus cooperation. Transportation Research Part B: Methodological **45**(1), 267–277 (2011)
5. Iii, J., Eisenstein, D.D.: A self-cordinating bus route to resist bus bunching. Transportation Research Part B **46**(4), 481–491 (2012)
6. Janos, M., Furth, P.: Bus priority with highly interruptible traffic signal control: Simulation of san juanš avenida ponce de leon. Transportation Research Record Journal of the Transportation Research Board **1811**, 157–165 (2002)
7. Newell, G.F.: Unstable brownian motion of a bus trip. In: Statistical Mechanics and Statistical Methods in Theory and Application, pp. 645–667. Springer (1977)
8. Xuan, Y., Argote, J., Daganzo, C.: Dynamic bus holding strategies for schedule reliability: Optimal linear control and performance analysis. Transportation Research Part B Methodological **45**(10), 1831–1845 (2011)