This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Distributed Weighted Balanced Control of Traffic Signals for Urban Traffic Congestion

Na Wu, Dewei Li, and Yugeng Xi, *Senior Member, IEEE*

***Abstract*— Since urban traffic congestion has become a major problem for big cities in recent years, we propose a distributed control scheme for traffic lights in the network. First, a new criterion called traffic process ability which implies the balance between the traffic demand and traffic capacity of each road is introduced. Moreover, the congestion of a road is mitigated by utilizing the traffic process ability of the neighbors more effectively, where different weights are assigned to roads in each agent according to their importance or the real-time traffic conditions. As only local information is needed, a distributed control scheme in which the road network is divided among several agents is proposed. Furthermore, in order to accelerate the congestion dissipation process, the aggregated state of each agent is introduced into the performance index and balanced with its neighboring agents. The control signals are calculated by agents in a parallel way and the optimization problem is solved iteratively to reach a convergence. Finally, the effectiveness of the proposed control scheme is evaluated by simulation under different scenarios and the performance is compared with the traffic responsive control method SCOOT.**

***Index Terms*— Balance, distributed control, traffic congestion, traffic process ability.**

I. Introduction

URBAN traffic congestion has been an intractable problem for majority of big cities around the world over the last decades. With the progress of global urbanization, the sharp rise of traffic demand in urban cities has led to the increase of traveling time of drivers, waste of fuel consumptions, and the consistent rising of carbon dioxide emissions. Besides, the chance of collisions is also increased due to tight spacing and frequent acceleration and braking. Thus various countermeasures have been proposed to overcome the traffic congestion problem. The government has tried to reduce traffic demand by imposing parking restrictions, adopting road pricing policies, and encouraging people to use public transport. On the other hand, traffic reporting and variable message signs have also been widely utilized to provide guidance for travelers. Moreover, maximizing the effectiveness of existing transportation infrastructure is another direction for congestion mitigation.

Large literature has contributed to studying the traffic signal control algorithms. Fixed-time control strategies in [1]–[4] compute the optimal phase splits or cycle time length of intersections by analyzing off-line traffic data. Because of the dependence on historical data rather than real-time data of fixed-time strategies, they are not efficient in the cases where the traffic demand varies unexpectedly. Therefore, adaptive and traffic responsive control algorithms such as SCOOT [5] have then been developed. In order to improve the control performance, control strategies with more complicated theories and more accurate models have then been introduced. The cell transmission model (CTM) has been proposed in [6]. A multivariable regulator using linear-quadratic regulator theory has been developed in [7]. In addition, fuzzy logic has been introduced in [8] to solve the problem of uncertainty. Model predictive control that solves an optimization problem with constraints over a finite time horizon has been applied in [9] and [10] for the traffic signal control.

However, the aforementioned centralized control strategies are difficult for application in large-scale traffic networks due to the computational complexity, reliability, and communication burden. It leads to the extensive research on distributed and decentralized control strategies [11]–[14]. More recently, a feedback controller motivated by back-pressure routing, which is first applied in wireless networks, has been proposed in [15] and [16]. Basic rules for designing pressure functions have been presented in [17]. The method is further extended by [18] to a cyclic phase policy instead of activating only one phase in a cycle. Furthermore, a two-level model-based predictive controller has been developed in [19]. By decomposing the network into small subnetworks in the upper level, the traffic control can be more efficient and flexible.

Traffic distribution also plays an important role in congestion spread. Studies from [20]–[22] have shown that a heterogeneous distribution of vehicles will decrease the network flow compared with the homogeneous situations. A perimeter control policy for two urban regions based on dynamic flow characteristics has been proposed in [23] to maximize the total network throughput. A control scheme that balances the queue length of all roads in the network has been developed in [24]. However, it is implied in [25] that roads are locally interconnected with each other. The states of roads are affected

Manuscript received October 9, 2017; revised July 1, 2018, September 3, 2018, and October 14, 2018; accepted October 21, 2018. This work was supported in part by the National Science Foundation of China under Grant 61433002, Grant 61673366, Grant 61521063, and Grant 61590924, and in part by the Science and Technology Innovation Action Plan Project of Shanghai Science and Technology Commission under Grant 18511104200. The Associate Editor for this paper was Y. Wang. *(Corresponding author: Dewei Li.)*

The authors are with the Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: wunuo0419@sjtu.edu.cn; dwli@sjtu.edu.cn; ygxi@mail.sjtu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2018.2878001

by the neighboring roads. Therefore, balancing the queue length of all roads in the whole network is unnecessary and also difficult to realize. Because of the spatial characteristics of congestion spread and correlations among local roads, it is more appropriate for roads to coordinate with their neighbors than to coordinate with all the roads in the network. On the other hand, the balance control without considerations of the storage capacity of roads may not make sense.

Following the idea of [26] in which the local balance is limited inside each agent, this paper contributes to accelerating the pace of balance and thus the speed of congestion dissipation. On one hand, the scale of local balance is extended by introducing the aggregated state of each agent. On the other hand, different weights are assigned to roads so that the important roads and bottlenecks can be served preferentially. Therefore, by extending the balance scale and giving higher priorities to the critical roads, the mobility of vehicles can be improved by guiding them to less congested roads. Moreover, the utilization of green time of traffic lights can be exerted through the reasonable allocation.

The paper is organized as follows. Section II describes the new measure of traffic congestion level used in this paper which describes the balance between the traffic demand and traffic capacity. Section III presents the proposed distributed control strategy where the weighted states inside each agent are balanced. Moreover, the aggregated state of each agent is also balanced with its neighbors. Section IV evaluates the efficiency of the proposed traffic signal control approach by case study and Section V concludes.

II. Traffic Modeling for Signal Control

A. Road Network

An urban traffic network comprised of M links (roads) and N nodes (signalized intersections) can be represented by a digraph $G = (\mathcal{L}, \mathcal{V})$, where $\mathcal{V} = \{j | j \in \mathcal{V}\}$ is the set of nodes and $\mathcal{L} = \{(j,m)|\ j,\ m \in \mathcal{V}\}$ is the set of links in the network. A signalized intersection changes its traffic signals according to a predefined switching strategy by modifying the cycle length, splits, or offsets. The traffic condition of the network can be implied by quantities such as the throughput of the network, total delay time of vehicles, and the average travel speed and so forth.

B. New Measure of Congestion Level

The total queue length of all roads in the network is minimized as the objective function in many studies such as [7], [10], [12], and [22]. However, the traffic distribution, the coordination of traffic signals among intersections, and the utilization of free space are seldom considered. The efficiency of control schemes cannot be completely reflected simply by considering the total queue length. Therefore, the concept of the traffic process ability is introduced in our study as the criterion of congestion level of a road. We assume that the sampling period is T and the discrete time index is k.

The input demand of link (j,m), $d_{j,m}(k)$, is defined as the total number of vehicles intending to enter (j,m) from its upstream links during the kth period, as shown in Fig. 1(a).

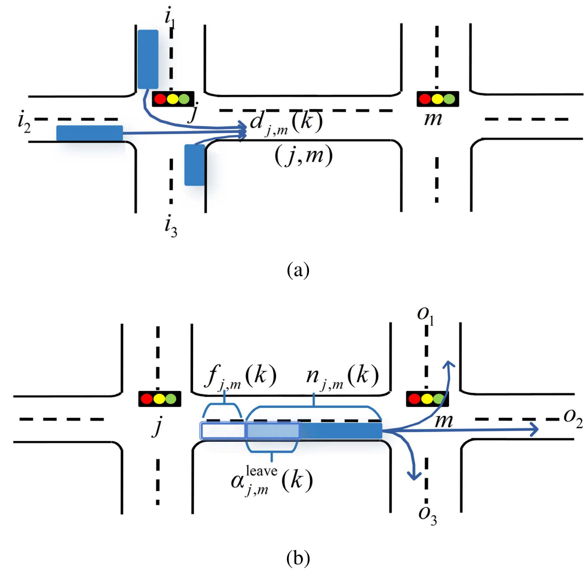

Fig. 1. (a) The input demand of link (j,m); (b) The input capacity of link (j,m).

It represents the actual requirements of upstream roads and will not be affected by the traffic condition of (j,m). Hence, $d_{j,m}(k)$ is restricted by the saturated flow rate of vehicles from the upstream roads and the current number of vehicles on the roads:

$$d_{j,m}(k) = \sum\nolimits_{i \in I_{j,m}} \min\left(n_{i,j,m}(k), \mu_{i,j,m} g_{i,j,m}(k)\right) \tag{1}$$

where the traffic stream on link (i, j) toward (j,m) is induced by the tuple (i, j, m), $n_{i,j,m}(k)$ is the number of vehicles of traffic stream (i, j, m) at time step k, $\mu_{i,j,m}(k)$ is the saturated flow rate in vehicles per second (veh/s), $g_{i,j,m}(k)$ is the green time in which (i, j, m) has right of way, and $I_{j,m}$ is the set of input nodes of link (j,m). Based on the above definition, the input demand of (j,m) can be controlled by the traffic signals of node j (through $g_{i,j,m}(k)$).

The input capacity of link (j,m), $c_{j,m}(k)$, is the number of vehicles that (j,m) can hold at the end of the kth time step,

$$c_{j,m}(k) = f_{j,m}(k) + \alpha^{\text{leave}}_{j,m}(k) \tag{2}$$

where $f_{j,m}(k)$ is the free space of link (j,m) that are not occupied by vehicles at the beginning of the kth period, represented by the empty blue box in Fig. 1(b). It is given by

$$f_{j,m}(k) = M_{j,m} - n_{j,m}(k) \tag{3}$$

where $M_{j,m}$ is the storage capacity of link (j,m) determined by the road length, jam density, and the number of lanes. Moreover, $\alpha^{\text{leave}}_{j,m}(k)$ is the number of vehicles leaving link (j,m) during the kth period denoted by the blue box with light shadow in Fig. 1(b):

$$\begin{aligned}
\alpha^{\text{leave}}_{j,m}(k) &= \sum\nolimits_{o \in O_{j,m}} \alpha^{\text{leave}}_{j,m,o}(k) \\
\alpha^{\text{leave}}_{j,m,o}(k) &= \min\Big(n_{j,m,o}(k), \mu_{j,m,o} g_{j,m,o}(k), \\
&\qquad f_{m,o}(k) + \alpha^{\text{leave}}_{m,o}(k)\Big)
\end{aligned} \tag{4}$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

where $O_{j,m}$ is the set of output nodes of link (j, m). It is assumed that the vehicles entering each link at one cycle cannot exit from it during the same cycle. From the definition, the input capacity provides a more accurate estimation of possible space for incoming vehicles by including the free space made by the outflow of (j, m) during the kth period. From (2) and (4), the input capacity of (j, m) can be controlled by the traffic signals of node m (through $g_{j,m,o}(k)$).

In this paper, the traffic process ability (TPA) is defined as the difference of the input capacity and the input demand, and selected as the state of the system model. It is a quantity that describes the degree of satisfaction to the demand of a road in terms of its capacity and suggests the balance between them. On one hand, with the consideration of outflow during the kth period in $c_{j,m}(k)$, all the available space on link (j, m) is considered. On the other hand, by restricting $d_{j,m}(k)$ to the minimum of the saturated outflow vehicles and the current number of vehicles from the upstream roads, it removes the effects from the vehicles that cannot enter (j, m) and thus are not necessary to consider. The benefit of the introduction of TPA is its ability to imply the coordination between the free space of each road and the green time of its adjacent traffic lights. Both the storage capacity of roads and the effects of adjacent traffic lights can be reflected in this state.

Let $x_{j,m}(k+1)$ denote the TPA of link (j, m) at the beginning of the $(k+1)$th period (at the end of the kth period), we have

$$x_{j,m}(k+1) = c_{j,m}(k) - d_{j,m}(k) \quad (5)$$

When the TPA of a road is negative, it implies that the number of vehicles in demand is larger than the maximum space to hold them, which will block the incoming vehicles and may cause a spill back afterwards. The smaller the TPA, the more serious this effect is. In over-saturated traffic conditions where the TPA of severely congested roads is all negative, the value implies the congestion degree and urgency for queue clearance of the road. On the roads with positive TPA, there is relatively more space for holding the incoming vehicles either resulting from larger input capacity or less input demand. It may lead to a waste of green time or inefficient use of storage capacity with more free space. As both the input demand and the input capacity in TPA can be controlled by traffic signals, the traffic conditions can be ameliorated by modifying the traffic signals to achieve the balance of TPA among roads.

Substituting (2) and (3) into (5), we have

$$x_{j,m}(k+1) = M_{j,m} - n_{j,m}(k) + \alpha^{\text{leave}}_{j,m}(k) - d_{j,m}(k) \quad (6)$$

By moving one step back, we get

$$x_{j,m}(k) = M_{j,m} - n_{j,m}(k-1) + \alpha^{\text{leave}}_{j,m}(k-1) - d_{j,m}(k-1) \quad (7)$$

Subtract (6) by (7), we have

$$x_{j,m}(k+1) = x_{j,m}(k) - n_{j,m}(k) + n_{j,m}(k-1) + \alpha^{\text{leave}}_{j,m}(k) - \alpha^{\text{leave}}_{j,m}(k-1) - d_{j,m}(k) + d_{j,m}(k-1) \quad (8)$$

With the conservation equation given by

$$n_{j,m}(k) = n_{j,m}(k-1) + \alpha^{\text{enter}}_{j,m}(k-1) - \alpha^{\text{leave}}_{j,m}(k-1) \quad (9)$$

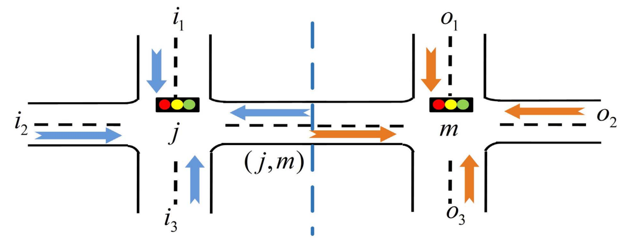

Fig. 2. The elements of agent j and agent m.

where $\alpha^{\text{enter}}_{j,m}(k)$ is the number of vehicles entering link (j, m) during the kth period, (8) can be written as

$$x_{j,m}(k+1) = x_{j,m}(k) + \alpha^{\text{leave}}_{j,m}(k) - d_{j,m}(k) + d_{j,m}(k-1) - \alpha^{\text{enter}}_{j,m}(k-1) \quad (10)$$

It is noticed that the last two terms in (10) are known and constant at the beginning of kth period. Let

$$\tau_{j,m}(k) = d_{j,m}(k-1) - \alpha^{\text{enter}}_{j,m}(k-1) \quad (11)$$

we can get the state equation of road (j, m):

$$x_{j,m}(k+1) = x_{j,m}(k) + \alpha^{\text{leave}}_{j,m}(k) - d_{j,m}(k) + \tau_{j,m}(k) \quad (12)$$

which indicates the change (improvements or deteriorations) of the satisfaction of the free space on link (j, m) to its input demand with time.

III. Distributed Weighted Balanced Control of Urban Traffic Signals

Distributed control is a control architecture that only local information is communicated among neighboring agents. It is more robust and computationally simpler than a centralized control scheme whereas more accurate than decentralized control without information exchange. Moreover, it provides convenience for reconfiguration by supporting plug-in agents only performed locally.

From (1)-(4) and (12), the TPA of roads is directly affected by the number of vehicles of the upstream and downstream roads as well as the adjacent traffic signals. This correlation among roads is also true for some other quantities such as the number of outflow vehicles. In other words, roads are affected by each other through the interconnected roads among them and cannot influence the roads geographically far away directly. It is not necessary to change the traffic signals far away for improving the traffic condition of one road in a short time, which implies the importance of focusing on local traffic conditions when mitigating congestion. This feature of traffic network is in accordance with distributed control strategies in nature. Based on the above considerations, we use a distributed control framework for traffic signal control in this paper.

A. Distributed Control Framework

For the implementation of the distributed control strategy, the network is firstly divided among agents. Each agent consists of an intersection with traffic lights and the corresponding incoming links. As shown in Fig. 2, agent m is comprised of intersection $m \in \mathcal{V}$ with its traffic lights, and incoming links $L_m = \{(j, m), (o_1, m), (o_2, m), (o_3, m)\}$.

For convenience, we define two sets of neighbors. The set of neighbors of an agent is defined as the adjacent agents $N_m = \{o|(o, m) \in \mathcal{L}\}$. Thus agent o is called the neighbor of agent m if there is a direct link (o, m) between them. The set of neighbors of a road is defined as the remaining roads that are contained in the same agent with it, $N_{j,m} = \{(o, m)|o \in N_m\}$. In Fig. 2, $N_m = \{j, o_1, o_2, o_3\}$, and $N_{j,m} = \{(o_1, m), (o_2, m), (o_3, m)\}$. Based on this definition, the upstream and downstream roads of roads in agent m are all contained in N_m, and the roads that have the competitive or shared green time with (j, m) in the same intersection are all included in $N_{j,m}$.

From (1), the input demand of link (j, m) is influenced by the number of vehicles on the upstream roads and the corresponding traffic signals. Similarly, from (4), the number of vehicles leaving link (j, m) is affected by the number of vehicles on the link and the traffic signals controlling its outflow. It can be written as

$$d_{j,m}(k) = h^{\mathrm{d}}_{j,m}(n_{i,j,m}(k), g_{i,j,m}(k)), i \in I_{j,m} \tag{13}$$

$$\alpha^{\mathrm{leave}}_{j,m}(k) = h^{\mathrm{l}}_{j,m}\left(n_{j,m,o}(k), g_{j,m,o}(k), n_{m,o}(k), g_o(k), M_{m,o}\right), \quad o \in O_{j,m} \tag{14}$$

where $h^{\mathrm{d}}_{j,m}(\cdot)$ and $h^{\mathrm{l}}_{j,m}(\cdot)$ represent the nonlinear functions abstracted from (1) and (4), and $g_o(k)$ is the vector form of traffic signals of agent o. Then the input demand and the number of outflow vehicles of agent m during the kth period can be represented by

$$d_m(k) = h^{\mathrm{d}}_m(n_j(k), g_j(k)), j \in N_m \tag{15}$$

$$\alpha^{\mathrm{leave}}_m(k) = h^{\mathrm{l}}_m(n_m(k), g_m(k), n_o(k), g_o(k), M_o), \quad o \in N_m \tag{16}$$

where $d_m(k)$, $n_m(k)$, $g_m(k)$, $\alpha^{\mathrm{leave}}_m(k)$, and M_m are the vector form of variables for traffic streams in agent m, and $h^{\mathrm{d}}_m(\cdot)$ and $h^{\mathrm{l}}_m(\cdot)$ are the corresponding vector form of nonlinear functions in (13) and (14). Combining with (12), the state equation of agent m can be given by

$$x_m(k+1) = x_m(k) + h_m\left(n_m(k), g_m(k), n_j(k), g_j(k), M_j\right) + \tau_m(k), \quad j \in N_m \tag{17}$$

where $h_m(\cdot) = h^{\mathrm{d}}_m(\cdot) - h^{\mathrm{l}}_m(\cdot)$.

Above all, we can get the model for each agent. Due to the existence of the couplings among agents and the distributed framework, each agent senses the states and then controls its traffic lights, while communicating with its neighboring agents to obtain the information required for the optimization and the coordination.

B. Distributed Weighted Balanced Control

The theory in [27]–[29] reveals the spatial and temporal characteristics of congestion propagation. The outflow of roads forms the traffic demand of their downstream links. Thus the massive discharge from a congested road will impose pressure on its downstream roads. When the downstream roads cannot handle the vehicles from the road, they will get congested. The congestion on one road spreads to its downstream roads in this way. Besides, the reduction of remaining space on the downstream roads will further restrict the outflow from the road and aggravate its congestion. On the other hand, because of the congestion on the road, vehicles from its upstream roads can be blocked. That is how the congestion on one road spreads to its upstream roads. When the congestion cannot be cleared immediately, it will further spread to the upstream and downstream roads of the newly congested roads in the same way. The pattern of congestion spread through upstream and downstream roads (adjacent roads) with time is called its spatial and temporal characteristics.

The way for congestion dissipation proposed in this paper is based on this idea. In the congestion dissipation process, the TPA is spread. For the road with less TPA compared with its neighboring roads, it reveals that it has a stronger need to clear the queue. We can improve its TPA by either increasing its input capacity or reducing its input demand. The former will lead to the increase of the input demand of the downstream roads whereas the latter will lead to the decrease of the input capacity of the upstream roads. Both the former and the latter will lead to the decrease of TPA of the upstream and downstream roads. In other words, the amelioration of TPA of one road is achieved by sacrificing the TPA of their adjacent roads, through which the TPA is transferred or spread among roads. Furthermore, this process can be extended to a larger scale and finally reach the entire network through consistent spread among adjacent roads, just like the spatial and temporal characteristics of the congestion spread. The scale is extended to the whole network progressively with time.

From [20], traffic distribution is an important factor that impacts the traffic condition of the network. A homogeneous traffic distribution will yield a higher outflow compared with the heterogeneous situation. Since the TPA of a road indicates the balance between its input demand and the input capacity, the local uniform of TPA implies the local uniform of this balance for roads. In this paper, we will try to achieve this kind of uniform by balancing the TPA among local roads. The balance mechanism will be illustrated tin the following sections.

1) Weighted Balance Inside Each Agent: the weighted TPA of roads inside each agent is balanced.

In the previous literature [9]–[13], roads are viewed as equally important in the optimization of traffic lights. Usually the additive queue length on roads or the total time spent by vehicles is selected as the objective function, where distinctions among roads brought by different congestion level or road grades are not taken into account. In fact, we are more inclined to mitigate the congestion on arterial roads or bottlenecks preferentially when the congestion appears because of the profound influence they have on the congestion evolution. However, this need is neglected when using the additive performance index.

The method proposed here tries to overcome this drawback by assigning different weights to roads, allowing for higher flexibility. The arterial roads or the bottlenecks have higher weights so that they can be served preferentially. The work reported here extends the preceding work in [26].

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

The weight for each road can be determined by different methods, in which the experience of experts, the real-time traffic conditions, or other critical factors can be considered. In this paper, we provide a heuristic method for the calculation of weights, where the potential impacts of roads on the congestion propagation are implied. The weight of link (j, m) is given by

$$\omega_{j,m}(k) = \delta_{j,m}(k) \cdot \gamma_{j,m}(k) + 1 \tag{18}$$

where $\delta_{j,m}(k)$ is an indicator reflecting the possibility of spill back caused by link (j, m). It is a 0-1 variable given by

$$\delta_{j,m}(k) = \begin{cases} 1, & \text{if } f_{j,m}(k) < \theta_{j,m} \\ 0, & \text{if } f_{j,m}(k) \geq \theta_{j,m} \end{cases} \tag{19}$$

where $\theta_{j,m}$ is a positive threshold that once the free space $f_{j,m}(k)$ is less than it, then road (j, m) cannot provide enough space for vehicles from the upstream roads and may lead to further congestion of roads in N_m. Moreover, $\gamma_{j,m}(k)$ is the total occupancy (the ratio of the queue length to its storage capacity) of the upstream roads of link (j, m), thus it indicates the degree of urgency to serve the vehicles on (j, m) to avoid congestion spread to the upstream roads. It is given by

$$\gamma_{j,m}(k) = \sum\nolimits_{i \in I_{j,m}} \frac{n_{i,j}(k)}{M_{i,j}} \tag{20}$$

The product $\delta_{j,m}(k)\gamma_{j,m}(k)$ implies the potential threatening of congestion propagation from link (j, m) to adjacent roads.

Above all, if there is not enough space on (j, m) for vehicles from the upstream roads, then $\delta_{j,m}(k) = 1$. If so, we will further investigate the occupancy level of its upstream roads to measure the degree of urgency for congestion mitigation on (j, m). When the upstream roads are also severe congested, there is a higher chance for the congestion to spread to other roads, which will lead to congestion spill back. The more severe the congestion of the upstream roads of link (j, m), the more urgent it is to relieve the congestion on it, and the lager the obtained $\omega_{j,m}(k)$ is. Otherwise, in the cases where the free space on (j, m) is sufficient, $\omega_{j,m}(k)$ is 1. Therefore, by assigning different weights to roads based on the local traffic conditions, the control strategy will be more reasonable. Besides, vehicles on other roads can also benefit from the fluency of the important roads or bottlenecks.

To achieve the local uniform of TPA, first the weighted TPA of roads inside each agent is balanced with their neighbors (the definition of neighbors of a road in an agent can be found in III-A). It can be formulated as

$$\begin{aligned} &\min\nolimits_{g_m(k)} J_{m,\text{in}} \\ &= \sum\nolimits_{j \in N_m} \sum\nolimits_{(h,m) \in N_{j,m}} \left\| \frac{x_{j,m}(k+1)}{\omega_{j,m}(k)} - \frac{x_{h,m}(k+1)}{\omega_{h,m}(k)} \right\|_2^2 \end{aligned} \tag{21}$$

By using the reciprocal of weights, (21) aims at producing control signals that leads to preferential increase of TPA of the road with urgent level compared with its neighboring road. The weighted balance process improves the TPA of the congested road by sacrificing the TPA of their neighbors, then the utilization of the free space on roads and the green time of traffic lights of the agent will be more reasonable.

2) Local Balance Among Neighboring Agents: the aggregated TPA among local agents is balanced.

In Sec. III-B1, the balance process is confined inside each agent, and the TPA of neighboring agents is not considered. With this limitation, the TPA of one agent is probably quite different than its neighboring agents, leading to the difference of TPA between one road and its upstream and downstream roads. Then the speed of congestion dissipation may also be restrained. In this paper, the scale for local balance is further extended among local agents. Similar to the balance among roads, the balance among agents will improve the TPA of one agent by sacrificing the performance of agents with higher TPA.

The aggregated TPA of each agent that represents the average level of TPA of the agent is introduced. It is defined as the average TPA of roads in the agent. More specifically, for agent m:

$$X_m(k+1) = \frac{1}{N_m} \sum\nolimits_{j \in N_m} x_{j,m}(k+1) \tag{22}$$

where $|N_m|$ is the number of neighbors of agent m and thus the number of roads in agent m.

In general, there are 3 possibilities for $X_m(k+1)$ with its neighbor $j \in N_m$:

- case 1: $X_m(k+1) < X_j(k+1)$
- case 2: $X_m(k+1) > X_j(k+1)$
- case 3: $X_m(k+1) = X_j(k+1)$

In the first case, the aggregated TPA of agent m is smaller than its neighbor j. As the difference between the input capacity and the input demand of agent j is greater, it has more space for vehicles compared with m. From the analysis in Sec. III-A, the upstream and downstream roads of link (j, m) are contained in the neighboring agents N_m. The balance of aggregated TPA between agent m and j will decrease $X_j(k+1)$ and increase $X_m(k+1)$. To increase $X_m(k+1)$, the input demand from agent j will be restricted by assigning less green time for phases in which the vehicles from roads in agent j to roads in agent m have right of way. It will decrease the outflow from agent j, thus reduce the input capacity of j. On the other hand, the input capacity of agent m will also be improved to increase $X_m(k+1)$ by giving more green time for the outflow of roads in agent m, leading to the increase of the input demand for its downstream roads in N_m. Therefore, the balance among agents changes the TPA of them simultaneously. It accelerates the speed of TPA transfer and the congestion mitigation by utilizing the green time of traffic lights and free space of neighboring agents more effectively.

In the second case, the control effects are opposite. In the third case, the TPA among neighboring agents has been balanced which is the ideal case and the purpose of our control scheme. For each agent, the realization of balance among neighboring agents implies the local uniform distribution of TPA. And the balance effect will progressively spread to the entire network by consistent local balance.

The balance process with neighboring agents can be formulated as

$$J_{m,\text{among}} = \min_{g_m(k)} \sum_{j \in N_m} \left\| X_m(k+1) - X_j(k+1) \right\|_2^2 \tag{23}$$

Above all, by the weighted balance of TPA of roads inside each agent and the balance of aggregated TPA among neighboring agents, both the green time of traffic lights and free space of roads can be fully exerted. Besides, the possibility of congestion spill back is also reduced.

C. Procedure of Optimization for Traffic Lights

In this paper, the control strategy is implemented by controlling the signaling split of traffic lights of intersections. Traffic signals are divided into several phases in which the specific traffic streams are served. We assume that the phase sequence and the cycle time at each intersection has been determined by other algorithms. In addition, the sequence of phases is fixed such that a phase appears only once in a cycle.

Because of the distributed control architecture, the optimization for each agent is solved independently. However, from (17), the coupled variables of agent m with its neighboring agents are unknown at time step k. Besides, the aggregated TPA of neighboring agents in the objective function (23) cannot be given either. Therefore, some variables must be estimated in the optimization of each agent.

At each time step, each agent communicates with its neighbors. Based on the obtained information, the optimization problem of each agent is solved and the control signals are computed. In order to achieve the convergence and global optimum of the distributed algorithm, each agent generates a control sequence of iterates. Starting from a feasible control $g_m^{(0)}(k)$, at each iteration s the agents exchange their states and decisions locally, compute the traffic signals for the next iteration, and keep working until the convergence is attained or time is up.

The optimization problem for agent m at iteration s is given by Problem m:

$$\min_{g_m(k)} J_m = \lambda_1 J_{m,\text{in}}^{(s)} + \lambda_2 J_{m,\text{among}}^{(s)} + \lambda_3 J_{m,\text{u}}^{(s)} \tag{24}$$

s.t.

$$(18) - (22), \quad \text{and}$$

$$x_{j,m}^{(s)}(k+1) = x_{j,m}(k) + \alpha_{j,m}^{\text{leave}(s)}(k) - \hat{d}_{j,m}(k) + \tau_{j,m}(k) \tag{25}$$

$$\alpha_{j,m}^{\text{leave}(s)}(k) = \sum\nolimits_{o \in O_{j,m}} \min\Big(n_{j,m,o}(k), \mu_{j,m,o} g_{j,m,o}^{(s)}(k), M_{m,o} - n_{m,o}(k) + \hat{\alpha}_{m,o}^{\text{leave}}(k)\Big) \tag{26}$$

$$\hat{d}_{j,m}(k) = \sum\nolimits_{i \in I_{j,m}} \min\big(n_{i,j,m}(k), \mu_{i,j,m} \hat{g}_{i,j,m}(k)\big) \tag{27}$$

$$\hat{g}_j(k) = g_j^{(s-1)}(k), \quad g_j^{(0)}(k) = g_j(k-1) \tag{28}$$

$$\hat{\alpha}_{m,o}^{\text{leave}}(k) = \alpha_{m,o}^{\text{leave}(s-1)}(k), \quad \alpha_{m,o}^{\text{leave}(0)}(k) = \alpha_{m,o}^{\text{leave}}(k-1) \tag{29}$$

$$\hat{X}_j(k+1) = X_j^{(s-1)}(k+1), \quad X_j^{(0)}(k+1) = X_j(k) \tag{30}$$

$$G_{\min} \le g_{m,j}^{(s)}(k) \le G_{\max}, \quad j \in F_m$$

$$\sum\nolimits_{j \in F_m} g_{m,j}^{(s)}(k) + L_m = T_m \tag{31}$$

where

$$J_{m,\text{in}}^{(s)} = \sum_{j \in N_m} \sum_{(h,m) \in N_{j,m}} \left\| \frac{x_{j,m}^{(s)}(k+1)}{\omega_{j,m}(k)} - \frac{x_{h,m}^{(s)}(k+1)}{\omega_{h,m}(k)} \right\|_2^2$$

Algorithm 1 The Optimization for Agent m

Input: $n_m(k)$, $x_m(k)$, $d_m(k-1)$, $\alpha_m^{\text{enter}}(k-1)$, $n_j(k)$, $\alpha_j^{\text{leave}}(k-1)$, $g_j(k-1)$, $X_j(k)$, $j \in N_m$

Output: $g_m(k)$

function OPTIMIZATION
- $g_j^{(0)}(k) \leftarrow g_j(k-1)$, $X_j^{(0)}(k+1) \leftarrow X_j(k)$
- $\alpha_j^{\text{leave}(0)}(k) \leftarrow \alpha_j^{\text{leave}}(k-1)$, $\forall j \in N_m$
- $e_m^{(s)} \leftarrow \infty$, $s \leftarrow 1$
- **while** $e_m^{(s)} > \epsilon$ and $s \le Q_{max}$ **do**
  - $\hat{g}_j(k) \leftarrow g_j^{(s-1)}(k)$, $\hat{X}_j(k+1) \leftarrow X_j^{(s-1)}(k+1)$
  - $\hat{\alpha}_j^{\text{leave}}(k) \leftarrow \alpha_j^{\text{leave}(s-1)}(k)$
  - solve Problem m
  - $e_m^{(s)} \leftarrow \left\| g_m^{(s)}(k) - g_m^{(s-1)}(k) \right\|_2^2$
  - $s \leftarrow s+1$
- **end while**
- $g_m(k) \leftarrow g_m^{(s)}(k)$

return $g_m(k)$

end function

$$J_{m,\text{among}}^{(s)} = \sum\nolimits_{j \in N_m} \left\| X_m^{(s)}(k+1) - \hat{X}_j(k+1) \right\|_2^2$$

$$J_{m,\text{u}}^{(s)} = \left\| g_m^{(s)}(k) - g_m^{(s-1)}(k) \right\|_2^2$$

Moreover, λ_1, λ_2, and λ_3 are the weights for the objective function and determined by trial and error, $G_{\min}$ and $G_{\max}$ are the lower and upper bounds for the green time of each phase of traffic lights in agent m, $g_{m,j}$ is the jth phase of the traffic signals, L_m is the lost time, and T_m is the cycle time. The objective function $J_{m,\text{u}}^{(s)}$ is added for accelerating the convergence speed.

The termination condition of the iteration process for the optimization of agent m is formulated by

$$e_m^{(s)} = \left\| g_m^{(s)}(k) - g_m^{(s-1)}(k) \right\|_2^2 < \epsilon \quad \text{or } s > Q_{max} \tag{32}$$

where ϵ is a small positive threshold, and Q_{max} is the maximum permitted number of iterations dependent on the computation time of the optimization and the sampling time interval. When either the difference of the control signals between two consecutive iterations is smaller than ϵ, or the maximum iteration number is reached, the iteration will be terminated and the obtained traffic signals will be implemented at the intersection. The procedure for the distributed weighted balanced control can be described by Algorithm 1.

IV. Case Studies

A. Network and Setup

In this section, a traffic network in Caohejing district in Shanghai of China is used for simulation. The simulation is based on CORridor-microscopic SIMulation program (CORSIM), which provides an approximation of the real-world traffic evolution. The network consists of 23 intersections controlled by traffic lights and 108 roads with different lengths (106-540 m). The geometry of the network is shown in Fig. 3.

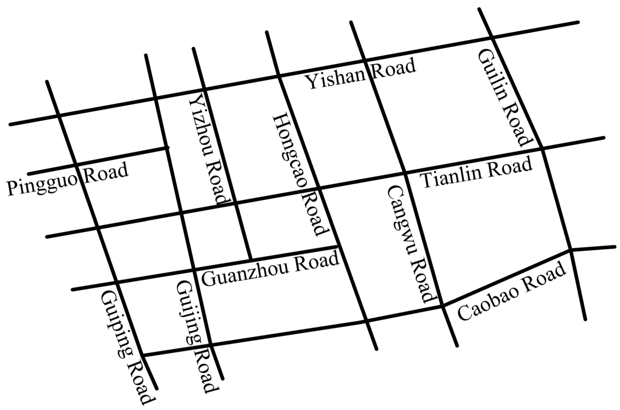

Fig. 3. The geometry of the test network.

Some parameters need to be set before the simulation. The saturated flow rate of each traffic stream μ is 16.7 veh/min. The cycle time of intersections T in the network is 80 s, and the lost time L is 0 s. In addition, each cycle is comprised of two phases, each allowing for the specific traffic streams. The lower bound $G_{\min}$ and upper bound $G_{\max}$ for each phase are 10 s and 50 s, respectively. The threshold θ in (19) is 20 vehicles.

B. Assessment

In order to evaluate the effectiveness, the performance of the proposed distributed weighted balanced control approach is compared with two other control algorithms:

1) SCOOT, which is an adaptive control algorithm for traffic lights. It estimates the average queue length of roads every 4 seconds based on the measurements and the traffic model. By minimizing the average queue length of roads in the future, it determines whether to change the green time of the next phase by a constant increment.
2) Distributed consensus control (DC), in which the TPA of roads inside each agent is balanced [26].
3) Distributed weighted balanced control (DWB), which is proposed in this paper. The weighted local TPA among roads and agents is balanced.

To compare the performance of the control strategies, five criteria are compared. Among them, the total delay time (TDT) is the accumulative difference between the actual travel time and the moving time under free-flow speed since the beginning of the simulation. It represents the time that vehicles are delayed if they cannot travel at the free flow speed. The number of congested roads (NC) counts the roads whose speed is below a threshold (here we use 5 km/h) that are in heavy congestion. Throughput error (TE) computes the difference of throughput under different control schemes. It represents the proportion of journey completion and indicates the fluency of the whole network. Here we compare the TE between DWB and SCOOT with the TE between DC and SCOOT. In addition, the distribution of the average speed (AV) of vehicles in the network is compared. The AV of vehicles on each road is computed. We divide the roads into 3 categories based on their AV, representing the free-flow roads, the lightly congested roads, and the heavily congested roads. Then we compute the proportion of each category. The traffic condition can be evaluated from a comprehensive analysis based on all the criteria.

TABLE I
TRAFFIC DEMAND SETTING OF EXPERIMENT I

Time (min)		1-60	61-80	81-160	161-200	201-240
demand	S1	1300	1200	1400	1200	1100
(veh/h)	S2	1700	1200	1700	1500	1200

C. Simulation Results

With the stochastic nature in real traffic networks, each scenario is employed by ten replications. To evaluate the effectiveness of the proposed algorithm in different scenarios, it is tested under various scenarios by the following numerical experiments.

1) Experiment I: the controller is tested in the network with even traffic demand.

This experiment emulates the network inside which there is no difference among the roads in terms of priorities. The control strategies is tested in the network with under-saturated and saturated traffic demand, respectively. The traffic demand of the two scenarios in this experiment which replicates the morning peak hour is shown in Table I. Scenario 1 (S1) denotes the under-saturated traffic situation with low traffic demand. It is used as the base scenario to evaluate the effectiveness of all controllers for comparison. Scenario 2 (S2) is the saturated situation with higher peak value of traffic demand. In the figure, the simulation results under SCOOT and DC is represented by red curve with stars and blue curve with dots, respectively. The magenta curve with triangles provides the results controlled by DWB.

The simulation results under S1 are given in Fig. 4. The curve of the queue length (QL), TDT and TE under the selected control schemes almost coincide. The difference of NC is only 1 in Fig. 4(c). Therefore, all the compared control strategies get comparable performance with under-saturated traffic situations under the low traffic demand.

In S2, the network is severe congested in the simulation implied by NC in Fig. 5(c). In Fig. 5(a), the improvement of DWB controller in terms of the maximum QL over DC and SCOOT is 8.4% and 17.7%, respectively. It implies that the DWB can effectively reduce the overall queue length in the network which will contribute to the decreasing the possibility of congestion spill back. Fig. 5(b) displays the evolution of the TDT in the network. It is shown that the difference of the accumulative TDT between DWB and the other two control schemes is increasing with time. At the end of the simulation in S2, the TDT is reduced 16.1% by DWB compared with DC, whereas it is reduced by 32.6% compared with SCOOT. The improvement demonstrates that vehicles can save much journey time under DWB, and the mobility of the network is improved. Moreover, the journey completion is also increased

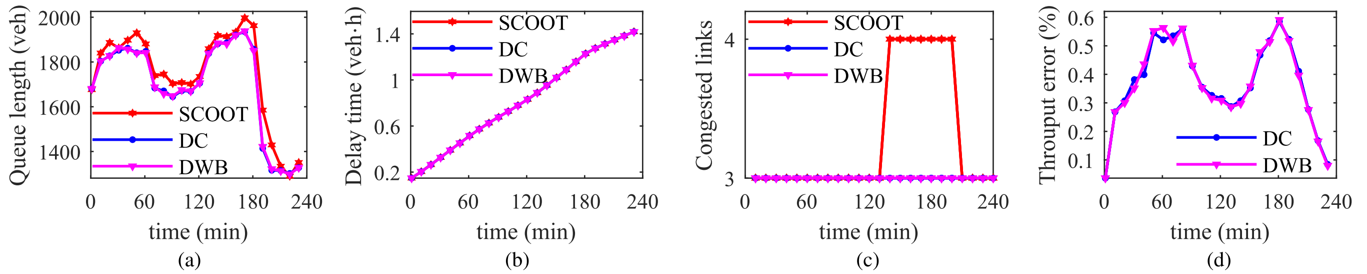

Fig. 4. Simulation Results under S1 in experiment I: (a) The queue length (QL); (b) Total delay time (TDT); (c) The Number of congested roads (NC); (d) The throughput error (TE).

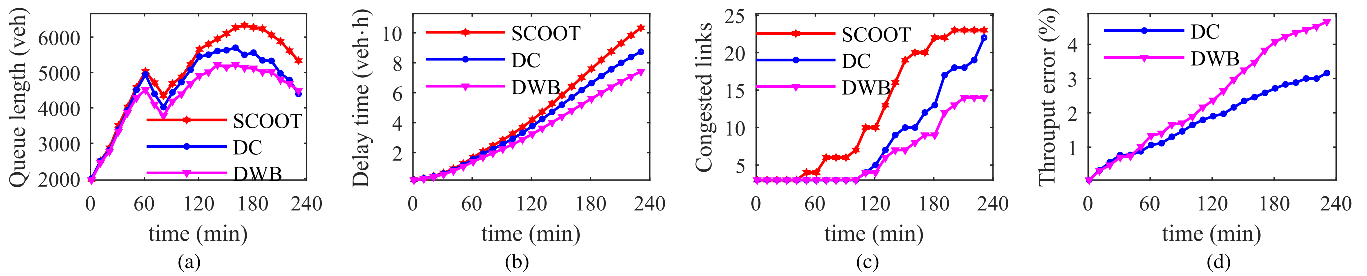

Fig. 5. Simulation Results under S2 in experiment I: (a) The queue length (QL); (b) Total delay time (TDT); (c) The Number of congested roads (NC); (d) The throughput error (TE).

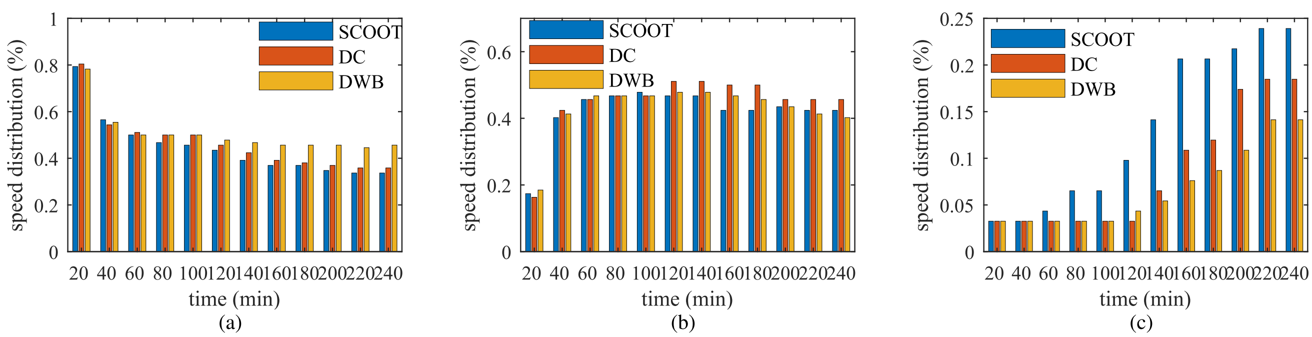

Fig. 6. The proportion of roads in each category under S2 in experiment I.

by DWB as shown in Fig. 5(d). The percent of difference of the throughput between DWB and SCOOT is denoted by the magenta curve with triangles, whereas the difference between DC and SCOOT is represented by the blue curve with dots. It can be seen that the percent of the difference is improved about 4.42% under DWB compared with SCOOT, whereas this value is 3.3% under DC. Combining with Fig. 5(a), the improvement of TE is probably resulted from the decrease of traffic density (the QL) in the network.

Besides, the evolution of the speed distribution helps us figure out how the congestion in the network evolves. The proportion of the free-flow roads (with AV $\geq$ 18 km/h), the lightly congested roads (with AV $\in$ [8, 18) km/h), and the heavily congested roads (with AV $<$ 8 km/h) in S2 is computed. The evolution of the proportion in each category reveals the evolution of the traffic condition in the network. The blue, orange, and yellow bar (the first, second, and the third bar) in Fig. 6 represent the proportion of roads in the category controlled by SCOOT, DC, and DWB, respectively. As depicted in Fig. 6(a), the proportion of the free-flow roads is initially 79% of the total and finally decreases to around 40%. The difference of the proportion of the lightly

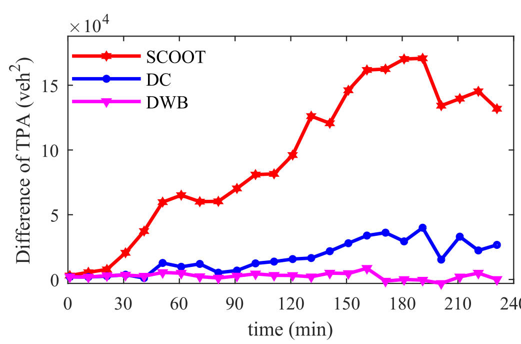

Fig. 7. The evolution of the difference of the weighted TPA in all agents under S2 in experiment I: (a)The proportion of the free-flow roads; (b) The proportion of lightly congested roads; (c) The proportion of heavily congested roads.

congested roads caused by different control schemes is not obvious. In Fig. 6(c), the proportion of the heavily congested roads increases rapidly as the congestion spreads. The speed of the congestion propagation is effectively controlled by DWB compared with SCOOT and DC, considering the gentle slope of the third bar in Fig. 6(c). Besides, the peak value for the proportion of the heavily congested roads has been reduced

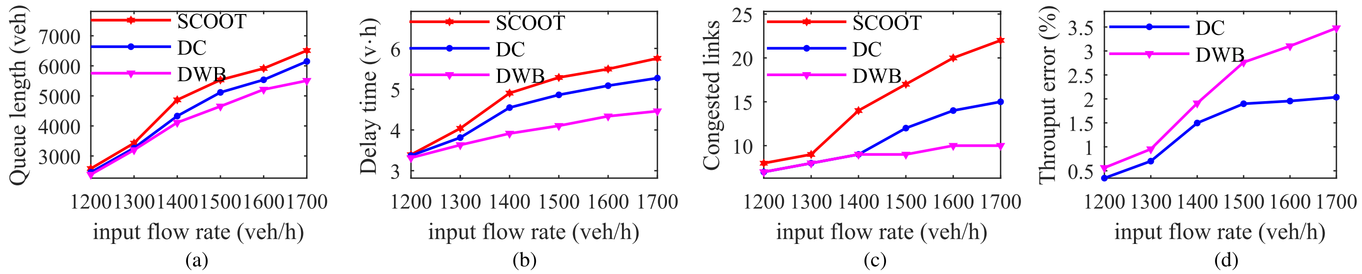

Fig. 8. Simulation results under different input flow in experiment II: (a) The queue length (QL); (b) Total delay time (TDT); (c) The Number of congested roads (NC); (d) The throughput error (TE).

to about 14% by DWB, in contrast to 24% by SCOOT and 18% by DC. In other words, DWB not only postpones the appearance of the congestion on some roads, but also relieves the congestion degree. In Fig. 6, the reduction of the proportion of the heavily congested roads is converted to the increase of the proportion of the free flow roads for DWB, indicating the speed of some roads is increased from below 8 km/h to beyond 18 km/h. The results are consistent with our local balance theory. Therefore, some roads can be protected from heavy congestion and the speed distribution are more even.

Fig. 7 gives the evolution of the sum of the difference of the weighted TPA in all agents under S2. After 120 minutes of simulation, the improvement of the proposed DWB strategy begins to emerge with more balanced TPA of roads. It is shown that in Fig. 5(c) and Fig. 5(d), the NC and TE under DWB and DC begin to diverge after 120 minutes of simulation, where the sum of the difference of the weighted TPA gets closer to zero in Fig. 7. The improvement brought by DWB can also be reflected in the speed distribution of roads in Fig. 6(c). In addition, during 60 to 80 minutes, the sum of the difference of the weighted TPA of three control algorithms is slightly reduced, whereas the QL in Fig. 5(a) is reduced simultaneously. Moreover, the NC in Fig. 5(c) and TE Fig. 5(d) stops deteriorating during 60-80 minutes. Therefore, the balance of the TPA of roads can effectively influence the traffic condition in urban traffic networks.

2) Experiment II: the controller is tested under different traffic demand.

In this experiment, the proposed control scheme is tested in 6 scenarios where in each case the traffic demand is constant during the 2-hour simulation. It varies from 1200 vehicles per hour (veh/h) in scenario 1 to 1700 veh/h in scenario 6. Thus we can evaluate the performance of the proposed control strategy when faced with different traffic demand.

The simulation results are presented in Fig. 8, where each point corresponds to the criteria computed at the end of the 2-hour simulation in each scenario. When the traffic demand is low (such as 1200 and 1300 veh/h), the effectiveness under DWB is comparable to that under SCOOT and DC. Nevertheless, the rising traffic demand leads to large performance degradation for SCOOT and DC, whereas this degradation effect is much smaller for DWB. In Fig. 8(b), TDT increases rapidly with the traffic demand for SCOOT and DC, whereas DWB can effectively limit the speed of growth. Moreover, only 3 more links are added to the congested link sets in Fig. 8(c), whereas the NC increases by 14 for SCOOT and increases by

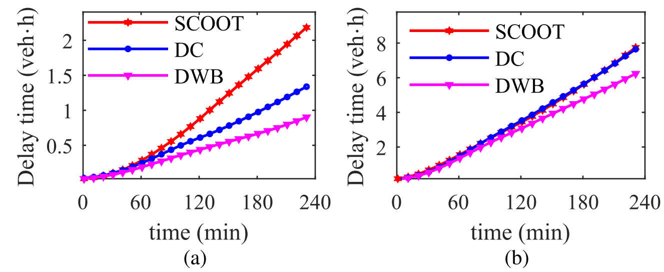

Fig. 9. Simulation results under experiment III: (a) TDT on the arterial road; (b) TDT on minor roads.

8 for DC. The TE in Fig. 8(d) reveals that even in networks with high traffic demand, DWB is capable of improving the trip completions compared with DWB. The results in this experiment demonstrate that DWB can effectively mitigate the congestion under various traffic demand.

3) Experiment III: the controller is tested in the network involving arterial roads.

In this experiment, Tianlin Road in Fig. 3 is the arterial road with 1.2 times of the traffic demand of minor roads. Moreover, the traffic demand for minor roads is the same as the settings of S2 in experiment I. According to the principle of DWB, the arterial road will be assigned with higher weights such that they can be served preferentially. The TDT on the arterial road and other minor roads during the 4-hour simulation is shown in Fig. 9.

In Fig. 9(a), the TDT on the arterial road is 2.7 veh·h under SCOOT and 1.4 veh·h under DC, whereas it is significantly reduced to 0.9 veh·h by DWB. Moreover, the TDT on minor roads is also reduced by 15% at the end of the simulation by DWB compared with SCOOT and DC. It reveals that the minor roads can also benefit from the fluency of the arterial road under DWB which leads to a win-win situation.

Above all, the performance of the proposed DWB control algorithm is demonstrated by experiment I to experiment III under various scenarios.

V. Conclusion

This paper have proposed a novel traffic signal control strategy in which the idea of congestion dissipation comes from the characteristics of the congestion propagation. Since the congestion is propagated through neighboring roads, it can also be mitigated in this way. A distributed framework is implemented since only local information is required for local balance.

In the first part, the traffic process ability of roads that depicts the balance between the input demand and the input capacity is introduced. It implies the degree of urgency for the queue clearance on the road. This quantity considers the storage capacity of roads so that free space on roads can be fully utilized. Furthermore, the network is divided among agents, each consisting of an intersection with traffic lights and the incoming links. Different from the previous control schemes which optimize the overall performance, the weighted balanced control focuses on the local traffic conditions of roads in this paper. It is implemented by balancing the weighted TPA of roads in each agent and balancing the aggregated TPA of neighboring agents at each time step, where the arterial roads or bottlenecks can be assigned with higher weights to be served preferentially, which makes the control scheme more flexible. A heuristic method for determining the weights is presented. Because of the couplings among agents, the optimization problem of each agent is solved iteratively until the convergence is achieved. Finally, the control performance of the proposed control scheme is evaluated in simulation by comparing with the traffic responsive strategy SCOOT and distributed consensus strategy. Three experiments are designed to replicate the traffic conditions in real life. All the three experiments reveal the effectiveness of the proposed scheme through the criteria.

In the future, we will explore the 2-level control scheme in which the distributed balanced control is used in lower level for achieving a balanced TPA among local roads and the perimeter control in [30] is introduced in upper level to coordinate the outflow of subregions such that they can operate at their optimal points. In addition, the convergence and stability analysis of the distributed control scheme will be investigated.

References

[1] J. D. C. Little, "The synchronization of traffic signals by mixed-integer linear programming," *Oper. Res.*, vol. 14, no. 4, pp. 568–594, 1966.

[2] D. I. Robertson, "'TRANSYT' method for area traffic control," *Traffic Eng. Control*, vol. 8, no. 8, pp. 276–281, 1969.

[3] R. E. Allsop, "SIGSET: A computer program for calculating traffic signal settings," *Traffic Eng. Control*, vol. 13, no. 2, pp. 58–60, 1971.

[4] G. Improta and G. E. Cantarella, "Control system design for an individual signalized junction," *Transp. Res. B, Methodol.*, vol. 18, no. 2, pp. 147–167, Apr. 1984.

[5] P. B. Hunt, D. I. Robertson, R. D. Bretherton, and R. I. Winton, "SCOOT—A traffic responsive method coordinating signals," Traffic Eng. Dept., Transport Res. Lab., Crowthorne, U.K., 1981.

[6] H. K. Lo, "A novel traffic signal control formulation," *Transp. Res. A, Policy Pract.*, vol. 33, no. 6, pp. 433–448, 1999.

[7] C. Diakaki, M. Papageorgiou, and K. Aboudolas, "A multivariable regulator approach to traffic-responsive network-wide signal control," *Control Eng. Pract.*, vol. 10, no. 2, pp. 183–195, Feb. 2002.

[8] S. Chiu and S. Chand, "Adaptive traffic signal control using fuzzy logic," in *Proc. IEEE 2nd Int. Conf. Fuzzy Syst.*, San Francisco, CA, USA, Mar./Apr. 1993, pp. 1371–1376.

[9] K. Aboudolas, M. Papageorgiou, A. Kouvelas, and E. Kosmatopoulos, "A rolling-horizon quadratic-programming approach to the signal control problem in large-scale congested urban road networks," *Transp. Res. C, Emerg. Technol.*, vol. 18, no. 5, pp. 680–694, Oct. 2010.

[10] S. Lin, B. De Schutter, Y. Xi, and H. Hellendoorn, "Fast model predictive control for urban road networks via MILP," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 3, pp. 846–856, Sep. 2011.

[11] A. L. C. Bazzan, "A distributed approach for coordination of traffic signal agents," *Auton. Agents Multi-Agent Syst.*, vol. 10, no. 1, pp. 131–164, Jan. 2005.

[12] R. R. Negenborn, B. De Schutter, and J. Hellendoorn, "Multi-agent model predictive control for transportation networks: Serial versus parallel schemes," *Eng. Appl. Artif. Intell.*, vol. 21, no. 3, pp. 353–366, Apr. 2008.

[13] L. B. de Oliveira and E. Camponogara, "Multi-agent model predictive control of signaling split in urban traffic networks," *Transp. Res. C, Emerg. Technol.*, vol. 18, no. 1, pp. 120–139, Feb. 2010.

[14] Z. Zhou, B. D. Schutter, S. Lin, and Y. Xi, "Multi-agent model-based predictive control for large-scale urban traffic networks using a serial scheme," *IET Control Theory Appl.*, vol. 9, no. 3, pp. 475–484, 2015.

[15] T. Wongpiromsarn, T. Uthaicharoenpong, Y. Wang, E. Frazzoli, and D. Wang, "Distributed traffic signal control for maximum network throughput," in *Proc. 15th IEEE Int. Conf. Intell. Transp. Syst. (ITSC)*, Anchorage, AK, USA, Sep. 2012, pp. 588–595.

[16] P. Varaiya, "Max pressure control of a network of signalized intersections," *Transp. Res. C, Emerg. Technol.*, vol. 36, pp. 177–195, Nov. 2013.

[17] J. Gregoire, X. Qian, E. Frazzoli, A. D. L. Fortelle, and T. Wongpiromsarn, "Capacity-aware backpressure traffic signal control," *IEEE Trans. Control Netw. Syst.*, vol. 2, no. 2, pp. 164–173, Jun. 2015.

[18] T. Le, P. Kovács, N. Walton, H. L. Vu, L. L. H. Andrew, and S. S. P. Hoogendoorn, "Decentralized signal control for urban road networks," *Transp. Res. C, Emerg. Technol.*, vol. 58, pp. 431–450, Sep. 2015.

[19] Z. Zhou, B. De Schutter, S. Lin, and Y. Xi, "Two-level hierarchical model-based predictive control for large-scale urban traffic networks," *IEEE Trans. Control Syst. Technol.*, vol. 25, no. 2, pp. 496–508, Mar. 2017.

[20] H. S. Mahmassani, M. Saberi, and A. Zockaie, "Urban network gridlock: Theory, characteristics, and dynamics," *Transp. Res. C, Emerg. Technol.*, vol. 36, pp. 480–497, Nov. 2013.

[21] V. V. Gayah and C. F. Daganzo, "Clockwise hysteresis loops in the macroscopic fundamental diagram: An effect of network instability," *Transp. Res. B, Methodol.*, vol. 45, no. 4, pp. 643–655, 2011.

[22] V. L. Knoop, S. P. Hoogendoorn, and J. W. C. Van Lint, "The impact of traffic dynamics on macroscopic fundamental diagram," in *Proc. 92nd Annu. Meeting Transp. Res. Board*, Washington, DC, USA, Jan. 2013, pp. 1–19.

[23] J. Haddad, "Optimal perimeter control synthesis for two urban regions with aggregate boundary queue dynamics," *Transp. Res. B, Methodol.*, vol. 96, pp. 1–25, Feb. 2017.

[24] Z. He, L. Wang, D. Li, and L. Zhang, "Steady-state signal control for urban traffic networks," in *Proc. Steady-State Signal Control Urban Traffic Netw. (ITSC)*, Las Palmas, Spain, Sep. 2015, pp. 463–470.

[25] K. Jang, H. Kim, and I. G. Jang, "Traffic signal optimization for oversaturated urban networks: Queue growth equalization," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 2121–2128, Aug. 2015.

[26] N. Wu, D. Li, and Y. Xi, "Balance traffic control in urban traffic networks based on distributed optimization," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, Rio de Janeiro, Brazil, Nov. 2016, pp. 428–433.

[27] C. F. Daganzo, "Urban gridlock: Macroscopic modeling and mitigation approaches," *Transp. Res. B, Methodol.*, vol. 41, no. 1, pp. 49–62, 2007.

[28] N. Geroliminis and C. F. Daganzo, "Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings," *Transp. Res. B, Methodol.*, vol. 42, no. 9, pp. 759–770, Nov. 2008.

[29] C. F. Daganzo and N. Geroliminis, "An analytical approximation for the macroscopic fundamental diagram of urban traffic," *Transp. Res. B, Methodol.*, vol. 42, no. 9, pp. 771–781, Nov. 2008.

[30] M. Hajiahmadi, J. Haddad, B. De Schutter, and N. Geroliminis, "Optimal hybrid perimeter and switching plans control for urban traffic networks," *IEEE Trans. Control Syst. Technol.*, vol. 23, no. 2, pp. 464–478, Mar. 2015.

Na Wu received the B.Sc. degree in control science and engineering from the Nanjing University of Science and Technology, Nanjing, China, in 2009. She is currently pursuing the Ph.D. degree with the Department of Automation, Shanghai Jiao Tong University, Shanghai, China.

Her work focuses on urban traffic congestion control and distributed control.

Dewei Li received the B.S. and Ph.D. degrees in automation from Shanghai Jiao Tong University, Shanghai, China, in 1993 and 2009, respectively.

He was a Post-Doctoral Researcher with Shanghai Jiao Tong University, Shanghai, China, from 2009 to 2010, where he is currently an Associate Professor with the Department of Automation. His research interests include model predictive control, intelligent control, and intelligent transportation systems.

Yugeng Xi (M'05–SM'05) received the Dr. Ing. degree in automatic control from Technical University Munich, Munich, Germany, in 1984.

Since then, he has been with the Department of Automation, Shanghai Jiao Tong University, Shanghai, China, where he became a Professor in 1988. He has authored or co-authored five books and over 300 academic papers. His research interests include predictive control, large scale and complex systems, and intelligent robotic systems.

Dr. Xi is an Editor or Associate Editor of 11 academic journals, including *Control Engineering Practice*, the *International Journal of Humanoid Robotics*, and *Acta Automatica Sinica*.