Netw Spat Econ
DOI 10.1007/s11067-008-9080-9

Urban Traffic Jam Simulation Based on the Cell Transmission Model

Jiancheng Long · Ziyou Gao · Xiaomei Zhao · Aiping Lian · Penina Orenstein

Abstract There have been different approaches which have been proposed to understand the mechanism of traffic congestion propagation. In this paper, we use the cell transmission model and apply it to simulate the formation and dissipation of traffic jams at the microscopic level. In particular, our model focuses on jam propagation and dissipation in two-way rectangular grid networks. In the model, the downstream exit of the link is channelized to represent the interactions of vehicles in different directions. We have used traffic jam size and congestion delay to measure jam growth and dispersal. Numerical examples exploring the impact of model parameters on jam growth and congestion delay are provided. The simulation results show that there are two strategies to minimize jam size and reduce time for jam dissipation: (1) reduce the length of channelized area and, (2) allocate the stopline widths for all directions in the same ratio as the demands. Furthermore, we obtain some new results about gridlock and discuss the effect of incident position and link length on jam propagation.

Keywords Traffic jams · Cell transmission model · Jam dissipation · Channelization

J. Long (✉) · Z. Gao · X. Zhao · A. Lian
School of Traffic and Transportation, Beijing Jiaotong University,
Beijing, 100044, People's Republic of China
e-mail: 04114182@bjtu.edu.cn

P. Orenstein
Stillman School of Business, Seton Hall University,
South Orange, New Jersey, USA

1 Introduction

There are a number of traffic congestion studies in the literature but the majority of these are concerned with congestion formation on a single bottleneck (Daganzo and Laval 2005; Newell 1993; Ni and Leonard 2005) or the effect of congestion in a traffic network at the macroscopic system level (Gao and Song 2002; Gentile et al. 2007; Lam and Yin 2001). However, there are very few studies of congestion formation and dissipation at the microscopic level. It is important to study traffic jam development and dispersal at the microscopic level because this may yield qualitative insights about how to control traffic jams in general as well as form the basis for incident management techniques.

A traffic jam can start in one of three ways (Wright and Roberg 1998): (1) a temporary obstruction, (2) a permanent capacity constraint in the network itself, and (3) a stochastic fluctuation in the demand within a particular sector of the network. In general, incident-based traffic jams can be categorized under the first type. During the past decade, colleagues at Middlesex University have used a combination of theoretical analysis and computer simulation to reveal some unexpected features of jam propagation for idealized grid networks (Roberg 1994, 1995; Roberg and Abbess 1998; Roberg-Orenstein et al. 2007; Wright and Roberg 1998; Wright and Roberg-Orenstein 1999). Roberg (1994, 1995) proposed simulation models which concentrated on a holistic view of traffic jam formation due to incidents, and described a number of strategies which could be exploited to achieve a controlled dispersion of traffic jams. Wright and Roberg (1998) proposed a simple analytical model for incident-based jam growth and discussed the effect of the length of the channelized part of roads and stopline width assignment on jam formation. Roberg and Abbess (1998) applied a simulation model to investigate the diagnosis and treatment of traffic jams. Wright and Roberg-Orenstein (1999) developed simple models for traffic jams and strategies for congestion control on idealized rectangular grid networks. Roberg-Orenstein et al. (2007) developed several alternative strategies for protecting networks from gridlock and dissipating traffic jams once they had formed. The treatment focused on the installation of bans at specific network locations.

All the above mentioned works related to traffic jam formation and control have enabled a deeper insight into understanding some of the issues associated with area-wide traffic congestion. However, the studies are deficient because they sacrifice a great deal of realism, and the results are not immediately applicable to real networks. The major networks they used are idealized, uniform, one-way rectangular grid systems, which are not generally found in urban traffic networks. On the contrary, two-way roads are more commonly found in urban traffic networks, and their findings need to be evaluated in this context. Furthermore, the traffic flow model utilized in their models is oversimplified and does not adequately capture real traffic. Consequently, in this paper, we propose to extend the cell transmission model (CTM) (Daganzo

1994, 1995) to investigate the propagation of traffic jams in an idealized two-way rectangular grid network. The underlying reason for choosing the CTM is its plausibility in representing the spillback and shock wave phenomenon found in congested networks. Like the hydromechanical simulation model, the CTM is a discrete approximate model of the LWR model (Lighthill and Whitham 1955; Richards 1956), which is a macroscopical kinetic model. It can capture the realistic traffic dynamics, including important features such as shock waves, queue formation and dissipation, and dynamic traffic interactions across multiple links such as queue spillback. Because of the advantages in modelling network traffic, CTM has been used widely in dynamic traffic simulations (Long et al. 2008; Shang et al. 2007), signal control (Lo 1999, 2001), and dynamic traffic assignment (Ziliaskopoulos 2000; Lo and Szeto 2002; Lian et al. 2007).

In this paper, we are concerned only with jams arising from a single incident in a two-way grid network system. With the fundamental work of Wright and Roberg (1998), the CTM is extended to simulate traffic jam formation and dissipation. In particular, both left turning and right turning movements are considered at the downstream channelized area of each link. We examine the process of jam formation and of jam dissipation in this context. In order to compare the properties of traffic jams produced under different conditions, we introduce two measures which include jam size and congestion delay (Lo 2001). The effect of stopline width assignment and the length of the channelized area on jam formation and dissipation are analyzed by computational simulation. 'Gridlock' phenomena are also highlighted in our investigation. Furthermore, we examine the effect of incident position and link length on jam propagation. We confirm some of the results described in Wright and Roberg's work, and we also highlight some interesting new results.

This paper has been organized as follows. In the next section, a network traffic simulation model based on CTM is introduced. Simulations and results are given in Section 3. Finally, our summary and conclusions are provided in Section 4.

2 Network traffic flow propagation model

2.1 Model assumptions

Following Wright and Roberg (1998), some assumptions are made for the extended CTM network model. The assumptions mainly focus on the travellers' route choice, traffic demand, and the road characteristics. Our assumptions are more reasonable than Wright and Roberg's. In our model, it is not necessary for all the cells on each link to assume a uniform width. In addition, although the speeds of all the vehicles are determined by the flow-speed fundamental diagram, again the speeds are not assumed to be the same everywhere on the

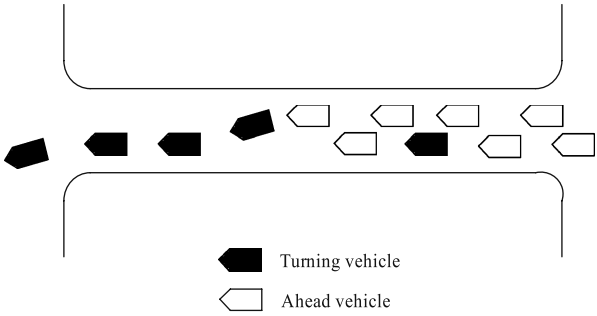

Fig. 1 Interference between turning vehicles and ahead vehicles on a road where the turning discharge is obstructed. (figure source from Wright and Roberg (1998))

network. However, once the vehicles are captured in a spillback queue their speed is assumed to be uniform. The assumptions in the model are outlined below:

- The flow between any origin-destination pair is constant. Furthermore, the route used by each driver is effectively fixed in advance, and in particular, drivers do not change their routes in response to congestion.
- The capacity per unit stopline width is the same for all directions. The reservoir capability (where the turning movements are mixed) and the outflow capacity of the channelized queue storage areas are proportional to the width of the stopline.
- Only the deterministic component of flow is taken into account, i.e. we ignore stochastic variations, together with any cyclic variation caused by alternate red and green periods at traffic signals.
- If a particular exit (e.g., left turning) is blocked, vehicles intending to drive in that direction will form a queue that spills back along the link. If the obstruction is not cleared in time, the queue may eventually spread across the other lanes and block all the traffic (Fig. 1).

2.2 Notation

In a multi-destination and strong connecting network $G = (N, A)$, N is defined as the set of nodes, A is the set of arcs (links); $a = (l, m)$ is the link formed of nodes l and m; A_l denotes the set of links leading to node l; B_m is the set of links leaving node m; v and w denote free-flow speed, the speed of the backward shock wave, respectively.

In the CTM, each roadway is discretized into homogeneous sections (or cells) and time is partitioned into intervals such that the cell length is equal to the distance travelled by free-flow traffic in one time interval δ. As shown in Fig. 2, link a is divided into two distinct zones: a downstream queue storage area where vehicles are organized into separate turning movements, and an upstream 'reservoir' where the turning movements are mixed. For a particular

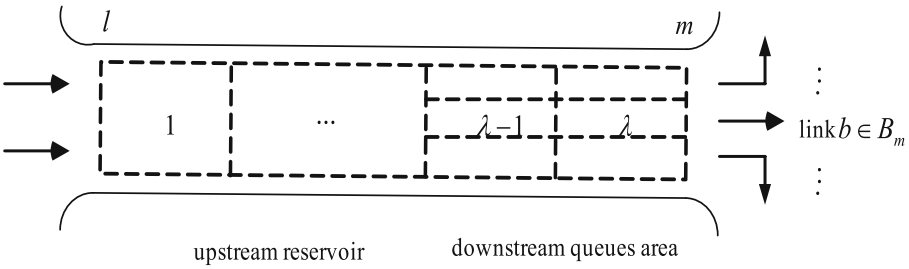

Fig. 2 Components of link *a*

cell, the downstream queue storage area consists of three divisions which form the segregated queuing areas. We assume that λ is the number of cells in each link and I is the number of cells that belong to the queue storage area, (i.e. the channelized area). It follows that the number of cells in the upstream reservoir is therefore $\lambda - I$.

The following notation has been adopted to represent the model:

$n_a^i(t)$	The number of vehicles contained in cell i of link a at the start of time interval t.
$n_{ab}^i(t)$	The number of vehicles contained in cell i of link a and take link b as the next link at the start of time interval t.
$y_a^i(t)$	The number of vehicles that flow into cell i of link a in time interval t.
$y_{ab}^i(t)$	The number of vehicles that flow into cell i of link a and take link b as the next link in time interval t.
$Q_a^i(t)$	The maximum number of vehicles that can flow into cell i of link a in time interval t.
$N_a^i(t)$	The maximum number of vehicles that can be present in cell i of link a in time interval t.
$n_{a,m}(t)$	The number of vehicles that are present in the terminal cell of link a and take node m as their destination at the start of time interval t.
$g_{l,a}(t)$	The number of vehicles that are generated at origin l and flow into the first cell of link a in time interval t.
$n_{l,a}(t)$	The number of vehicles that are present at origin l (including the waiting vehicles before time interval t) and that can flow into link a at the start of time interval t.
ϕ_{ab}	Proportion of vehicles traveling from link a to link b.
α_{ab}	Proportion of stopline width devoted to vehicles traveling from link a to link b.
$d_a^i(t)$	The sum of the congestion delay of vehicles contained in cell i of link a at time interval t.
$d(t)$	The total congestion delay of vehicles contained in the whole network at time interval t.

2.3 Model formulation

2.3.1 Overview of the cell transmission model

Daganzo (1994, 1995) proposed the CTM which simplified the solution scheme of the LWR model by adopting the following relationship between traffic flow, q, and density, k (see Fig. 3):

$$q = min\left\{vk, q_{\max}, w(k_j - k)\right\}, 0 \le k \le k_j \tag{1}$$

In Eq. (1), we let $q_{\max}$, k_j denote the inflow capacity (or maximum allowable inflow) and jam density, respectively. For all $a \in A$, the number of vehicles that flow from cell $i-1$ into cell i in time interval t can be calculated via

$$y_a^i(t) = q_a^i(t)\,\delta = min\left\{vk_a^{i-1}(t)\delta, q_a^{i,\max}(t)\delta, w(k_j - k_a^i(t))\delta\right\} \tag{2}$$

where, $k_a^i(t)$, $q_a^i(t)$, $q_a^{i,\max}(t)$ denotes the density, the inflow rate and the maximum inflow rate of cell i during time interval t, respectively.

With the relationship $n_a^i(t) = k_a^i(t)v\delta$ between the density and the number of vehicles contained in cell i, we have

$$y_a^i(t) = \min\left\{n_a^{i-1}(t), Q_a^i(t), w\left(N_a^i(t) - n_a^i(t)\right)/v\right\} \tag{3}$$

The flow conservation equation of the LWR model can be expressed as

$$n_a^i(t+1) = n_a^i(t) + y_a^i(t) - y_a^{i+1}(t) \tag{4}$$

2.3.2 The proposed model

Our model can be decomposed into two components: the *cell inflow model* and the *flow conservation model*. The former is formulated by extending the original CTM (Daganzo 1994, 1995). In Daganzo's model, the inflow of each

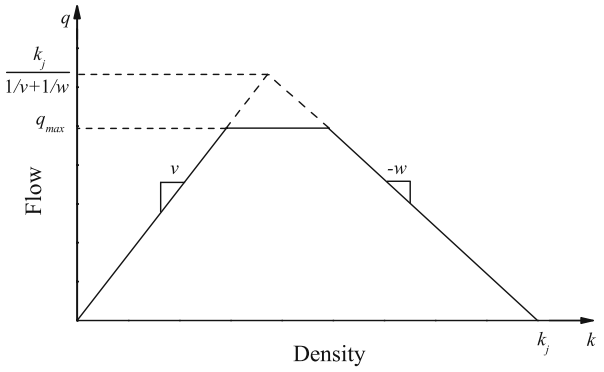

Fig. 3 The equation of state of CTM

cell can be formulated via Eq. (3) without considering a vehicle's turning movement at the downstream end of the link. Herein, we take the effect of channelization into account, and consequently, the inflow of each cell should be modified. It is obvious that the outflow of a particular cell is the inflow of its nearest downstream cell.

For every link in the network, the inflow formulation of each cell is varied due to its position on the link. Therefore, the cell inflow formulation can be classified into four categories: *inflow of upstream reservoir* ($i = 1$), *inflow of upstream cells* ($1 < i \leq \lambda - I$), *inflow of downstream queues area* ($i = \lambda - I + 1$), and *inflow of channelized cells* ($\lambda - I + 1 < i \leq \lambda$). For a particular link, the first category of cell inflow is the *boundary condition* of traffic flow propagating on the link. The other three categories of cell inflow represent the traffic flow that travels through the link. We can extend Daganzo's model in a simple way by formulating these categories of cell inflow (see Appendix for details).

Compared with the original CTM, the most remarkable difference of our model is the formulation of the boundary condition. To simplify the formulation, $y^{\lambda+1}$ is defined as the outflow of the terminal cells. CTM is extended to formulate flow propagation at intersections, given by

$$y_{ab}^{\lambda+1}(t) = min\left\{n_{ab}^{\lambda}(t), \alpha_{ab}\, Q_a^{\lambda}(t),\, Q_b^1(t), \alpha_{ab}\, w\left(N_b^1(t) - n_b^1(t)\right)/v\right\} \tag{5}$$

$$g_{l,a}(t) = min\left\{n_{l,a}(t),\, Q_a^1(t) - \sum_{b \in A_l} y_{ba}^{\lambda+1}(t)\right\} \tag{6}$$

By Eqs. (5) and (6), the inflow and the outflow of link $a = (l, m)$ can be calculated as follows, respectively,

$$y_a^1(t) = g_{l,a}(t) + \sum_{b \in A_l} y_{ba}^{\lambda+1}(t) \tag{7}$$

$$y_a^{\lambda+1}(t) = n_{a,m}(t) + \sum_{b \in B_m} y_{ab}^{\lambda+1}(t) \tag{8}$$

Note that $y_{ab}^1(t)$ can be calculated by Eq. (16) in Appendix. We also would like to point out that the inflow model formulated above can be used to describe the flow propagation in the network. The cases in which (a) vehicles depart from origins and enter the network and (b) vehicles arrive at destinations and leave the network are captured in the model. The number of vehicles from the upstream links can be calculated by Eqs. (5) and (7).

The flow conservation model, which is applied to update the number of vehicles contained in each cell, is formulated as follows

$$n_a^i(t+1) = n_a^i(t) + y_a^i(t) - y_a^{i+1}(t),\ 1 \leq i \leq \lambda \tag{9}$$

$$n_{ab}^i(t+1) = n_{ab}^i(t) + y_{ab}^i(t) - y_{ab}^{i+1}(t),\ 1 \leq i \leq \lambda \tag{10}$$

In general, $Q_a^i(t)$ and $N_a^i(t)$ have constant values due to the road configuration at cell i. The CTM can be easily modified to handle signalized networks as well as the occurrence of incident or link blockage (Lo and Szeto 2002). Traffic signal control and incidents can be modelled by modifying the value of the flow capacity $Q_a^i(t)$ of the affected cells. Specifically, for the case of a temporary obstruction, the inflow capacity of cells with obstruction can be altered according to whether an obstruction exists using:

$Q_a^i(t) = s_f$, t belongs to the period without obstruction on cell i.
$Q_a^i(t) = s_o$, t belongs to the period with obstruction on cell i.

where s_f is the saturation flow rate, s_o is the restricted flow rate by obstruction, and $s_o \leq s_f$. In this paper, the incident can be cleared just only by removing the obstruction, and we assume $s_o = 0$.

3 Computational simulation

3.1 Determination of simulation parameters

Using the network model which has been described in the previous section, we design a network traffic simulation model based on the time-step method. As shown in Fig. 4, a 16×16 two-way grid network is employed to test the model. All boundary nodes are both origins and destinations. ϕ_L, ϕ_A, ϕ_R denote, respectively, the proportion of vehicles travelling in the left turning direction, in the ahead direction and in the right turning direction. Stopline

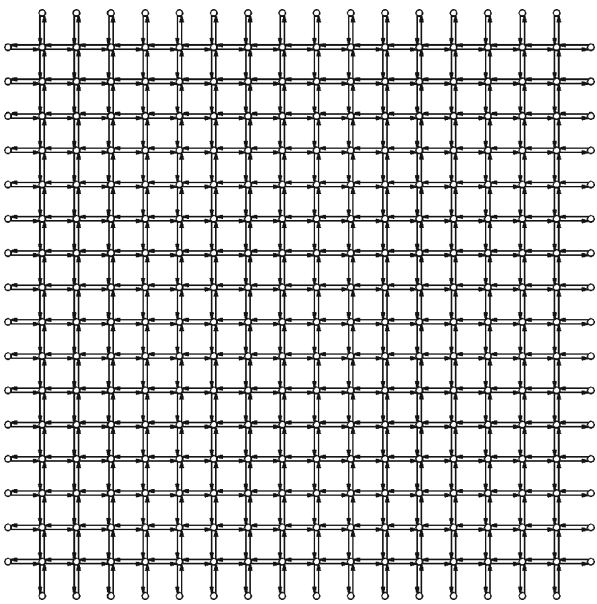

Fig. 4 16×16 two-way grid network

width assignment variables $\alpha_L, \alpha_A, \alpha_R$ denote, respectively, the proportion of the segregated queuing areas devoted to the left turning queue storage area, to the ahead queue storage area and to the right turning queue storage area. According to the definition, we have $\phi_L + \phi_A + \phi_R = 1$ and $\alpha_L + \alpha_A + \alpha_R = 1$.

We mainly focus on the difference between ahead vehicles and turning vehicles in the stopline width assignment problem, and therefore we set the stopline width devoted to both turning directions in exactly the same ratio as the demands, or $\alpha_L/\phi_L = \alpha_R/\phi_R$. The proportion of stopline width devoted to the turning directions can be expressed by α_A as follows:

$$\alpha_L = \frac{(1-\alpha_A)\phi_L}{\phi_L + \phi_R}, \quad \text{and} \quad \alpha_R = \frac{(1-\alpha_A)\phi_R}{\phi_L + \phi_R} \tag{11}$$

For the special case, if the stopline widths devoted to the ahead direction and to the turning direction are in exactly the same ratio as the demands, it follows immediately that $\alpha_A = \phi_A$, $\alpha_L = \phi_L$, and $\alpha_R = \phi_R$. This case is generally referred to as a 'balanced' layout of stopline assignment, (Wright and Roberg 1998). Equation (11) can also be applied to calculate α_L and α_R when $\alpha_A \neq \phi_A$.

The parameters for the extended CTM are set as follows:

- The length of each time interval δ: 5 s
- Jam density: 133 vehicles/km (i.e., 7.5 m for every vehicle)
- Free-flow speed: 54 km/h (i.e., 15 m/s), and backward shock-wave speed (Jiang et al. 2001): 21.6 km/h (i.e., 6 m/s)
- Number of lanes: 2
- Flow capacity: 1800 vehicles/h/lane (i.e., 2.5 vehicles/time interval/lane)
- The length of each cell is 75 m, and the holding capacity of each cell is 20 vehicles.
- The number of cells of each link: 9 (i.e., the length of every link is 675 m)

The analysis period of interest is divided into 4500 intervals (i.e., 6.25 h). The traffic demand is two vehicles per interval for each origin. The flow proportion for all directions are: $\phi_L = 0.2$, $\phi_A = 0.5$, $\phi_R = 0.3$. The initial network is empty, and some time intervals are required to allow the system to stabilize. After that we set a single incident on the 5^{th} cell of a link in the central zone of the network. The incident occurs at the 301^{st} interval, and is cleared at the 1000^{th} interval.

The traffic considered in this paper consists of identical vehicles (i.e., passenger cars). The parameter 7.5 m is the headway of a vehicle stuck in a jam, which takes into account both the length of the vehicle and the safe stopping distance. This parameter has been widely applied in the field of traffic flow theory (see Nagel and Schreckenberg (1992) for example). We set high values of turning proportions in our experiments. There are two reasons for this: (a) high turning proportions are often found at particular locations in an urban traffic network and (b) the stopline width assignment can be controlled over a suitable range, which is beneficial to illustrate the experiment results.

3.2 Congestion evaluation

We introduce two measures to explain the effect of congestion. These are (a) traffic jam size and (b) the congestion delay. The calculations of the two measures are described below.

3.2.1 Measure in jam size

Wright and Roberg (1998) measured the size of traffic jams in terms of the total number of blocked links, which does not represent the process of jam growth accurately since there may be instances where a link is partially blocked and would thus not be accounted for. Consequently, in this paper, jam size is measured with improved accuracy by examining the behavior at the cell level. We firstly define the concept of a *jammed cell* which can occur if

- the density of a cell in the upstream 'reservoir' is greater than $0.9k_j$; or if
- the density of a cell in any direction of the downstream channelized area is greater than $0.9k_j$.

It follows that *jam size* can now be measured as a function of the total number of *jammed cells*. Since $y_a^i(t) \leq w(N_a^i(t) - n_a^i(t))/v$ and $w/v < 1$, we have $n_a^i(t+1) \leq n_a^i(t) + y_a^i(t) < N_a^i(t)$ if $n_a^i(t) < N_a^i(t)$. This result shows that all cells can not be absolutely blocked if the initial network is empty. This is why we have adopted $0.9k_j$ (which is less than k_j) as a criterion to define a jammed cell.

3.2.2 Congestion delay estimation

There is a convenient way to determine congestion delay for CTM (Lo 1999). Congestion delay is defined as the additional time beyond free-flow travel time that a vehicle stays in a cell. At the cell level, the delay is determined as:

$$d_a^i(t) = n_a^i(t) - y_a^{i+1}(t) \tag{12}$$

In Eq. (12), $y_a^{i+1}(t)$ is the outflow of cell i at time interval t, and Eq. (3) ensures that $n_a^i(t) \geq y_a^{i+1}(t)$. It follows that if the exit flow from cell i at time interval t is less than its current occupancy due to congestion, then the vehicles who cannot leave the cell will incur a delay of one time step. Once the delay has been determined at the cell level, it can easily be aggregated to the network level, as given by

$$d(t) = \sum_a \sum_i d_a^i(t) \tag{13}$$

3.3 Simulation results

This paper discusses factors that influence the formation and dissipation of jams due to a single incident set up during the simulation. Under our assumptions, there are two main factors: the length of the channelized queuing areas

and the stopline width assignment. Moreover, it is also interesting to note the impact of incident position and link length on the rate of jam propagation.

3.3.1 The length setting of channelized queue areas

The length of the channelized queuing areas is denoted by the number of cells in this area. We are interested in studying the effect of the number of channelized cells on jam formation and dissipation. To simplify the representation of the changes in jam propagation relative to the number of channelized cells, the following acronyms related to jam size have been adopted:

JSIC jam size at the time interval when incident is cleared
MJS the maximum jam size before traffic flow comes back to free-flow condition
TJCE the time interval at which jam cells are completely eliminated
TMJS the time interval with MJS

Likewise, a similar set of acronyms related to congestion delay have also been adopted:

SCD sum of congestion delay due to incident
IRSCD the increase ratio of SCD compared with the case of $I = 1$
TMCD the time interval with maximum congestion delay

The results presented in Fig. 5(a) indicate that the shorter the downstream queues area is (i.e., the smaller I), the smaller the jam size and the faster it takes for the jam to dissipate. On the contrary, when I gets large, this will lead to traffic jams which extend over a larger area. Likewise, the results listed in Table 1 also show that extending the segregated queues area will not only improve jam growth rate (i.e., increase the number of cells involved in the traffic jam per time interval), but also postpone jam dispersal. When $I = 1$, only 161 time intervals are necessary to eliminate all the jammed cells following the removal of the obstruction. However, this number increases to 364 time

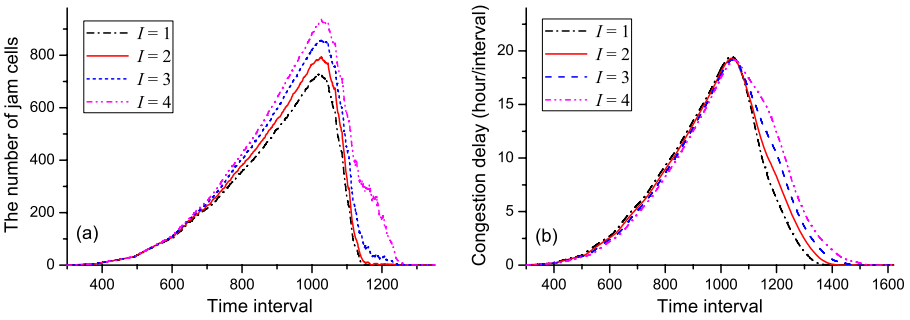

Fig. 5 Jam formation and dissipation under various length of queues area ($\alpha_A = 0.45$): (**a**) the variation of jam size, (**b**) the variation of congestion delay

Table 1 The effect of the number of channelized cells on jams propagation

	$I = 1$	$I = 2$	$I = 3$	$I = 4$
JSIC (cells)	705	766	822	883
MJS (cells)	730	795	863	937
TJS (time interval)	1027	1027	1027	1027
TJCE (time interval)	1161	1235	1271	1364
TMCD (time interval)	1036	1038	1042	1047
SCD (veh.hour)	7354.29	7571.31	7862.51	8259.90
IRSCD (%)	0.00	2.95	6.91	12.31

$\alpha_A = 0.45$

intervals when $I = 4$. Since vehicles involved in the downstream queue storage areas are organized into separate turning movements, the available space in the separate queues area can not always be filled to capacity, especially once queues have formed. Therefore, traffic queues grow faster in the channelized queuing areas than in the upstream reservoir area. We also can learn that although traffic jams can potentially clear after the incident has been removed, nevertheless the jam size may continue to expand for a number of intervals even after the incident has been removed. In this numerical study, it took 27 time intervals following the removal of an incident before the jam size began to contract.

Jam size is only one way to measure a traffic network's overall performance, which is why we have used congestion delay to further illustrate the impact of queue formation and dissipation in this context. The variation of congestion delay is presented in Fig. 5(b). The results indicate that the value of I has less influence on congestion delay prior to incident clearance. The explanation for this phenomenon is that the vehicles contained in the *jammed cells* generate most of the travel delay. Since the channelized cells are not completely filled during jam propagation, the number of vehicles contained in the *jammed cells* is approximatively equal under different values of I. Furthermore, we have examined the parameters related to congestion delay as introduced in Table 1. The results show that TMCD occurs later than TMJS, and TMCD occurs later and later, as one increases the value of I.

The results presented in Fig. 5 show that the congestion delay is still large even when the jam size reaches zero. This implies that the area around the incident is also involved in congestion in spite of the fact that the jammed cells have been completely eliminated. Following incident clearance, the congestion delay reduces faster when I is lower. We also compared SCD under various values of I as shown in Table 1. The value of IRSCD expands more rapidly when I is higher. Therefore, reducing the length of the channelized area can not only lessen the jam size, but also leads to less congestion delay during jam dissipation. Our results suggest that one should minimize the length of the channelized queuing areas to avoid wasting the space which occurs in this region. This space arises due to the interferences between vehicles traveling in different directions which causes empty areas to occur within the queues.

Therefore, the length of the channelized queue areas should not be too long to prevent these empty pockets from occurring.

Figure 5 also shows that in general, jams dissipate more quickly than they form. This result has been presented by Lo and Szeto (2002) and Lian et al. (2007). The phenomenon can be explained by the kinematic wave theory. According to this theory, the speed of the shock is

$$v_{shock} = \frac{q_1 - q_2}{k_1 - k_2} \tag{14}$$

where k_1 (q_1) and k_2 (q_2) represent the densities (flows) of the different traffic states.

When an incident occurs, a stopping wave will be generated and will travel back to the position upstream of the incident. Calculated by Eq. (14), the shock speed is $-q_0/(k_j - k_0)$, where k_0 and q_0 are the density and flow of the traffic state (i.e., free flow in this paper) before incident occurrence, respectively. Once the incident is cleared, a starting wave will generate and travel with shock speed $-q_{\max}/(k_j - k_{\text{crit}})$, where k_{crit} is the critical density with maximum flow. In this study, the initial traffic condition is free flow, and we have $q_{\max} > q_0$ and $k_{\text{crit}} > k_0$. This implies that the speed of the stopping wave is lower than that of the starting wave. Therefore, the results presented in Fig. 5 show that jams dissipate more quickly than they form. It should be pointed out that in general, jams do not always dissipate more quickly than they form. If the initial traffic condition is congestion (i.e., $k_{\text{crit}} < k_0$), the speed of the stopping wave may be higher than that of the starting wave, and jams may form more quickly than they dissipate. In addition, traffic interactions across multiple links may postpone jam dissipation.

3.3.2 Stopline width assignment

$\alpha_A < \phi_A$ implies that proportionately more stopline width is allocated to the turning traffic. By contrast, $\alpha_A > \phi_A$ implies that proportionately more stopline width is allocated to the ahead traffic. These two cases are simulated and analyzed by setting the parameter $I = 1$, with the assumption that the length of the channelized queuing area is short.

Figure 6(a) presents how the jam expands during the period when the incident is present and decreases after the incident has been removed under $\alpha_A \leq \phi_A$. It can be seen that when a smaller proportion of stopline width is allocated to the ahead queues, the resulting traffic jam was slightly larger and it took longer for the jam to dissipate. A similar effect was noticed when considering congestion delay, as displayed in Fig. 6(b). We can learn that the stopline width assignment has little effect on the process of jam formation, but plays an important role in the speed of jam dissipation. Since the length proportion of channelized queue areas (i.e., $I/\lambda = 1/\lambda$) is very small, the time required for the traffic jam to propagate through the whole link is almost

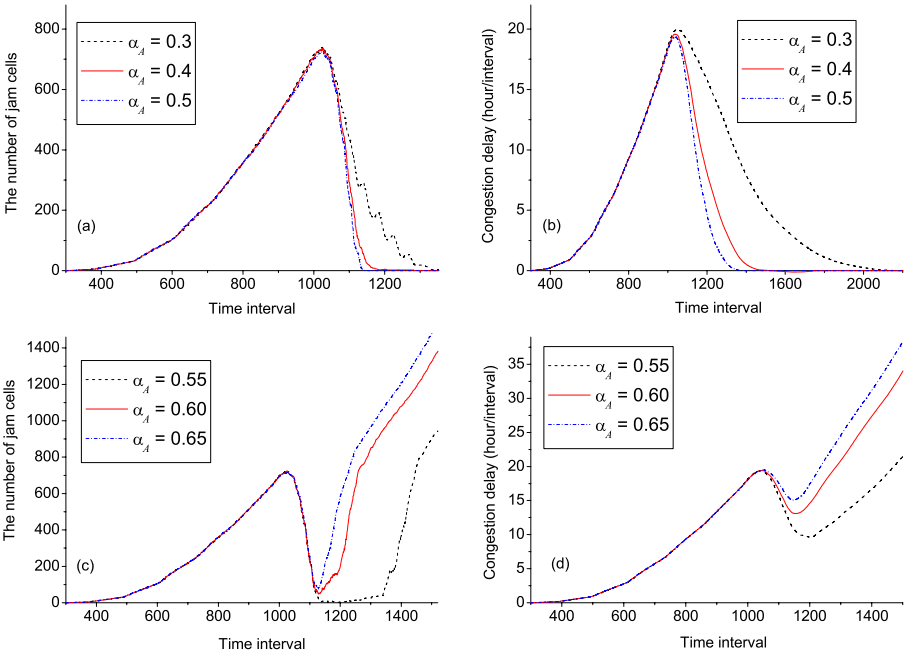

Fig. 6 Jam formation and dissipation under various channelization regimes: (**a**) the variation of jam size when $\alpha_A \leq 0.5$, (**b**) the variation of congestion delay when $\alpha_A \leq 0.5$, (**c**) the variation of jam size when $\alpha_A > 0.5$, (**d**) the variation of congestion delay when $\alpha_A > 0.5$

the same when the stopline width assignment changes. Compared with jam formation, stopline width assignment remarkably influences jam dissipation. If more stopline width is devoted to the turning movement (i.e., $\alpha_A < \phi_A$), then more space in the channelized cells will be wasted, because of queue interference between arriving vehicles. Therefore, the link exit capacity can not be fully utilized. This essentially prolongs the process of jam dissipation and invokes further congestion delay.

The results presented in Fig. 6(c) and (d) show that traffic jams can not be dispelled even after the incident has been cleared. We demonstrate this for three cases in which $\alpha_A \geq 0.55$. The traffic jam fails to clear during the dissipation process because a 'gridlock' appears during this time. Gridlock means that the tail of the queue arrives back at the starting point of the jams (Wright and Roberg 1998). Once gridlock has formed, the jams become extremely difficult to dissipate, on the contrary, they begin to grow again. From Fig. 6(c), we can see that the jam size rapidly reduces in size following the removal of the incident, but then it grows again. Likewise, Fig. 6(d) shows that congestion delay is still high even at the instance when the jam size reaches its lowest value (after incident clearance). This implies that the network remains congested and the impact of delay is still noticeable even in the absence of the incident which started the original jam.

In order to capture the process of jam growth, we have drawn a jam propagation tree ($\alpha_A = 0.6$) in Fig. 7, where the width of the black line denotes the time duration for which cells remained in the jammed condition. In Fig. 7(a), we use a statistical comparison period from the 301^{st} interval to 1000^{th} interval to represent jam growth due to an incident. In Fig. 7(b), we reset the statistical comparison period from 1141^{st} interval when a minimal jam size is reached to 1300^{th} interval to clearly represent the effect of gridlock. Figure 7(a) shows that no gridlock has formed near the source of the traffic jam before incident clearance. After incident clearance, the jam size reduces and attains its lowest value. This occurs at the 1141^{st} interval. Subsequently, we can observe how after a number of intervals have elapsed, how gridlock forms as a result of left turning vehicles (see Fig. 7(b)). Once the gridlock has formed, the jam can no longer dissipate without external intervention. On the contrary, we can observe the rapid expansion of the jammed area.

It is interesting to comment on the reappearance of a traffic jam even after the clearance of the original obstruction. Once the obstruction is removed, vehicles at the head of the queues restart and generate a starting wave. At the same time, upstream vehicles continue to arrive at the congested area. Initially, ahead vehicles can leave the congested area, and the queues appear to dissipate. However, the turning vehicles arrive at the tail of the initial queues and new queues rapidly form. The turning vehicles can not leave the area because they are restricted by the stopline width allocated to the turning directions. This causes interference between the turning movements and the ahead movements. As a result, fewer vehicles manage to leave the area, while

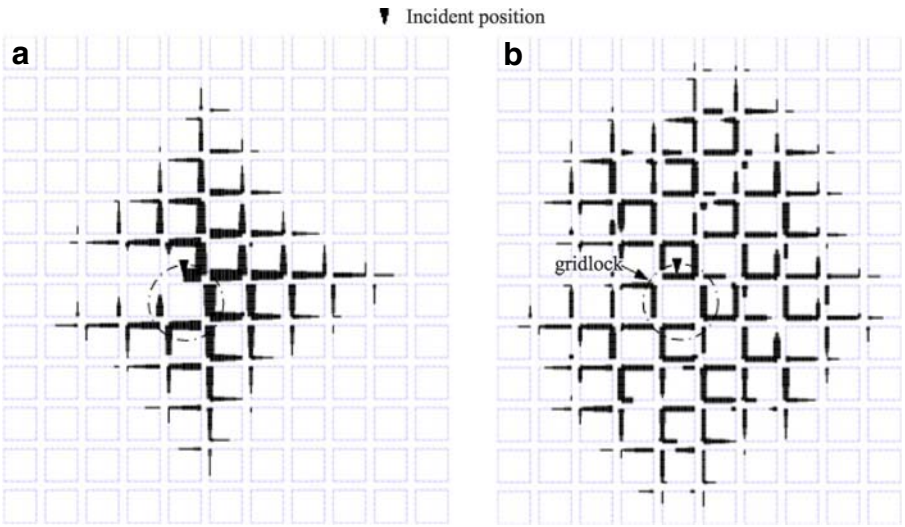

Fig. 7 The process of jam growth, the width of *black line* denotes the duration of cells under jam condition: (**a**) $\alpha_A = 0.6$, the statistical comparison period from 301^{st} interval to 1000^{th} interval; (**b**) $\alpha_A = 0.6$, the statistical comparison period from 1141^{st} interval to 1300^{th} interval

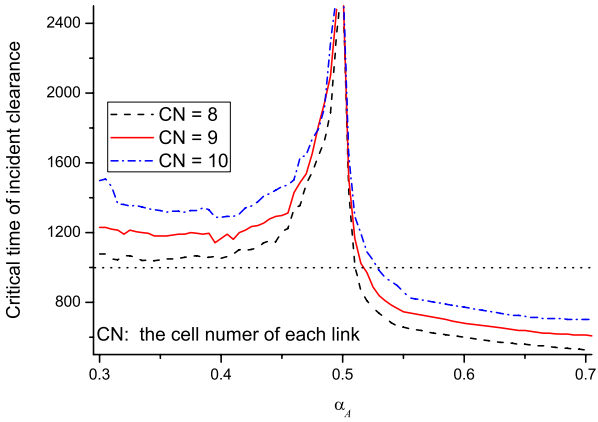

Fig. 8 Critical time of incident clearance of various channelization regimes to avoid the generation of gridlock ($\phi_A = 0.5$)

the number of arriving vehicles still increases. This leads to the reappearance of traffic jams.

As we know, traffic jams tend to propagate across a large area and it is the size of the jam which makes it more difficult to clear it. It is important to identify the source of the obstruction early in order to successfully eliminate the traffic jam. Therefore, we define the concept of the critical time, i.e., once the incident clearance time is greater than this time, the jammed traffic can never recover to its free-flow state. This critical time is referred to as *critical time of incident clearance*. It can be applied to determine the time available for the traffic engineer to control the jam. The critical time of incident clearance under various channelization regimes is graphically displayed in Fig. 8 and Fig. 9 where a sharp peak at $\alpha_A = 0.5$ can be observed. We can learn that the critical time is very sensitive to the proportion of stopline width devoted to the ahead traffic. Around $\alpha_A = \phi_A = 0.5$, the critical time reaches a maximum value. With $\alpha_A \leq \phi_A$, the critical time is lower and rapidly descends when $\alpha_A > \phi_A$. This result implies that when $\alpha_A > \phi_A$ 'gridlock' phenomena may be

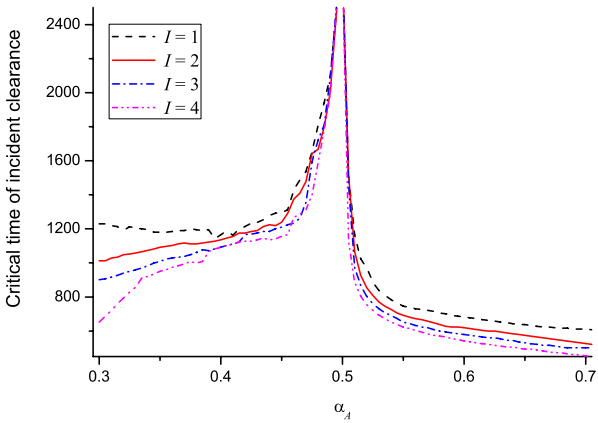

Fig. 9 Critical time of incident clearance under various length of queues area to avoid the generation of gridlock($\phi_A = 0.5$)

more frequent. Consequently, the regime of $\alpha_A > \phi_A$ is worse than $\alpha_A \leq \phi_A$. As shown, for $\alpha_A \leq 0.5$, the critical time of incident clearance is greater than 1000, and the jammed traffic can come back to the free flow condition some time after incident clearance. But for $\alpha_A \geq 0.55$, the critical time is less than 1000, and the traffic jam can not be dispersed due to gridlock formation. This result confirms our analysis in Fig. 6.

Both the cases of $\alpha_A < \phi_A$ and $\alpha_A > \phi_A$ cause the link exit capacity to be reduced due to wasted space within the channelized cells. However, if $\alpha_A \to \phi_A$ we can avoid wasting space in the channelized queue areas. According to the above analysis, the optimal stopline assignment is to allocate the widths of the segregated approach queues in exactly the same ratio as the demands, or $\alpha_A \to \phi_A$. This not only surpresses jam propagation, but it also postpones the onset of gridlock. In practice, when the proportion of ahead vehicles is unknown, it is advisable to allocate proportionately less stopline width (i.e., making $\alpha_A < \phi_A$) to reduce the probability of gridlock formation.

3.3.3 The effect of incident location and link length

Incident location and link length are also two important parameters that influence jam formation and dissipation. We change the two parameters slightly and compare our results. The SCD (sum of congestion delay due to incident) is graphically presented in Fig. 10. It shows that if more stopline width is allocated to the turning traffic (i.e., lower α_A) this will generate greater congestion delay. This implies that the value of α_A should be as close as possible to ϕ_A. Figure 10 also shows that if the incident is located near the upstream link entrance, this brings more congestion delay than when the incident occurs near the downstream exit. Incident location directly determines how long it takes for a queue to spillback to the upstream links. Once a queue has propagated to the upstream links, subsequent queues will branch out in different directions. This will obviously increase the rate of jam growth.

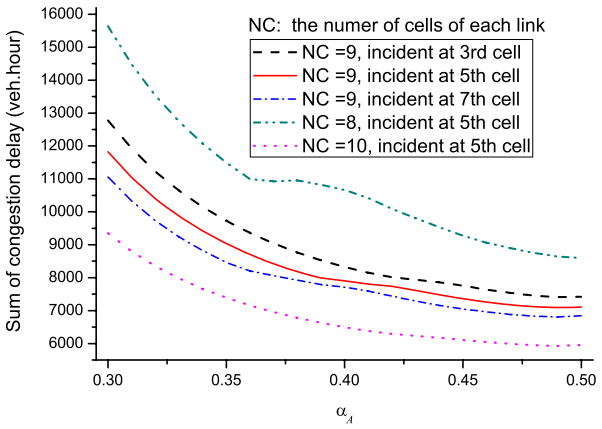

Fig. 10 The effect of incident position and link length on SCD ($\phi_A = 0.5$)

Compared with incident position, the effect of link length on jam propagation is more sensitive. A longer link (contains more cells) can prolong the period in which queues propagate through the whole link, which is beneficial when trying to prevent traffic interactions across multiple links. Moreover, link length has an important influence on the critical time of incident clearance. In Fig. 8, the critical time under various link length regimes is compared. The results indicate that longer links are helpful when dispersing traffic jams. Therefore, traffic network designers should take link length into account to guard against 'gridlock' phenomena.

3.4 Simulation with alternative flow proportions

We now explore the impact of increasing the proportion of vehicles traveling in the ahead direction assuming the same traffic demand. The other parameters remain the same as before. The flow proportions for all directions are reset as follows: $\phi_L = 0.1, \phi_A = 0.75, \phi_R = 0.15$. We repeat the same experiments as in the previous section.

The new set of simulation results provides similar results: (a) keeping the channelized area short can both reduce jam size and leads to less congestion delay during the jam dissipation phase; (b) a 'balanced' layout in stopline width assignment can not only guard against jam development, but also prevent 'gridlock' phenomena. In Fig. 11, we display the critical time of incident clearance under various channelization regimes. Note the sharp peak which occurs at $\alpha_A = \phi_A = 0.75$. The SCD is graphically presented in Fig. 12. It shows that when more stopline width is allocated to the turning traffic (i.e., lower α_A) this results in higher congestion delay. The results presented in Fig. 11 and Fig. 12 reflect our analysis presented in Figs. 8, 9 and 10 respectively and confirm our observation that it is advisable to keep the downstream queue areas short while maintaining longer links overall. This is helpful to guard against 'gridlock' phenomena and relieve congestion.

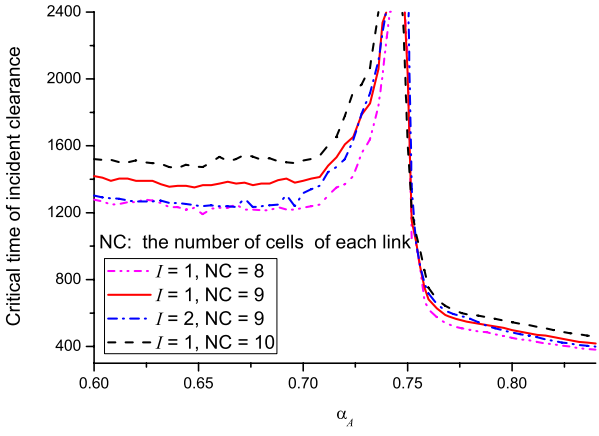

Fig. 11 Critical time of incident clearance of various channelization regimes to avoid the generation of gridlock ($\phi_A = 0.75$)

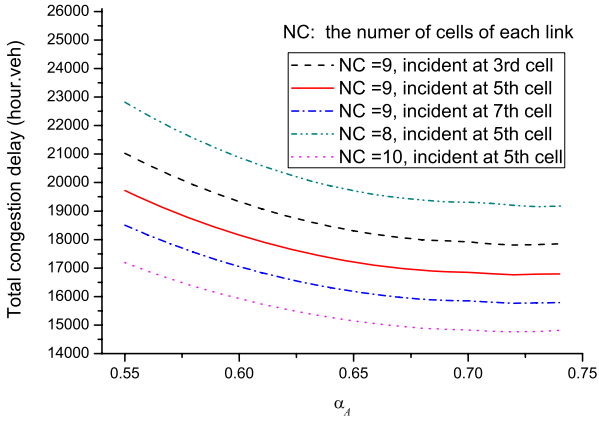

Fig. 12 The effect of incident position and link length on SCD ($\phi_A = 0.75$)

3.5 Comparison with related work

The model proposed by Wright and Roberg (1998) highlights an interesting dilemma in traffic management: a strategy that aims to minimize the rate of growth of a jam by a suitable allocation of queue storage space will actually encourage gridlock at the heart of the congested area, and conversely, a strategy that aims to defer gridlock by reserving as much storage space as possible for turning vehicles will result in queues spread over a wide area. Their basic results are: a 'balanced' layout of stopline assignment with the minimum length of segregated queue storage area will minimize jam growth rate; conversely, proportionately more stopline width allocated to the turning traffic and segregating the ahead and turning queues along the whole length of each link will maximize the time required to generate the gridlock. In our study, some results are coincident with Wright and Roberg's, yet some other new results are also derived.

On the problem of minimizing jam growth rate, the results derived from our simulation are similar to those of Wright and Roberg's: both lessening the length of the channelized area (i.e., using a lower I) and a 'balanced' layout in stopline width assignment (i.e., $\alpha_A \to \phi_A$) can prevent jam development. Moreover, our numerical studies also show that these two strategies appear to be more effective when dispersing traffic jams but do not make that much difference to suppressing jam growth.

In addition, the CTM model does generate some additional results which mainly focus on 'gridlock' phenomena. In Wright and Roberg's one-way network system, the gridlock formed during jam growth. However, in our study of the two-way system, the gridlock does not form before incident clearance. If the incident persists for a sufficient amount of time, (i.e. the traffic jam has assumed area-wide proportions), then the 'gridlock' will appear only after the incident has been cleared and the jammed traffic will never disperse without intervention. In this sense, we conjecture that one-way systems are actually worse than two-way systems, since gridlocks are part of the congestion

formation process. Gridlock phenomena always occur in one way systems, whereas two-way systems do not inherently produce gridlocks, only these might arise during the intervention phase.

The influence of the length of the channelized queue areas on gridlock generation is graphically presented in Fig. 9. The result shows that setting the value of I as low as possible can maximize the critical time of incident clearance. It also indicates that when more space is allocated to the turning vehicles ($\alpha_A < \phi_A = 0.5$) this does not seem to guard against 'gridlock' phenomena. The results both in Fig. 8 and Fig. 9 show that a 'balanced' layout in stopline width assignment is optimal to prevent 'gridlock' phenomena.

4 Conclusions and future work

In this paper, CTM is extended to simulate the propagation of jams due to an incident. The effect of the length of the channelized queue areas and the stopline width assignment on jam size and jam dissipation time has been discussed via a combination of theoretical analysis and simulation techniques. Numerical examples shows how the two strategies of (a) reducing the length of the channelized area and (b) equitable allocation of stopline width can delay queue spillback and shorten the jam dissipation time. More importantly, the two strategies can also guard against 'gridlock' phenomena. Moreover, the interesting dilemma that Wright and Roberg (1998) highlighted in traffic management is not faced in our two-way network model. Our results also show that incident position far away from link entrance will lead to less congestion delay. Furthermore, the length of the links should be taken into account when designing a traffic network. A network consisting of a group of short links is not to be encouraged if the goal is to prevent traffic jams and avoid 'gridlock' phenomena.

In this paper, we have mainly focused on investigating the propagation characteristics of area-wide traffic jams. For the process of jam dissipation, we only use incident clearance as a form of intervention. We do not consider any other forms of intervention but we envisage that the treatments of traffic jams proposed by Roberg (1994, 1995), Roberg and Abbess (1998), Wright and Roberg-Orenstein (1999), Roberg-Orenstein et al. (2007) can potentially be applied to our case. Our future work is to develop efficient diagnosis and treatment strategies for traffic jams in two-way grid networks.

Acknowledgements This work is partially supported by National Basic Research Program of China (2006CB705500), National Natural Science Foundation of China (70631001, 70801004) and the Innovation Foundation of Science and Technology for Excellent Doctorial Candidate of Beijing Jiaotong University (48040).

Appendix

In this section, Daganzo's CTM is extended to formulate three categories of cell inflow.

1) *Inflow of upstream cells*
Inflow of upstream cells can be calculated by:

$$y_a^i(t) = min\left\{n_a^{i-1}(t),\, Q_a^i(t),\, w\left(N_a^i(t) - n_a^i(t)\right)/v\right\},\ 1 < i \le \lambda - I \tag{15}$$

From Eq. (15), we have

$$y_{ab}^i(t) = \phi_{ab}\, y_a^i(t),\ 1 < i \le \lambda - I \tag{16}$$

2) *Inflow of downstream queues area*
We define $\widetilde{y}_{ab}(t)$ as the up bound of inflow of downstream queues area for vehicles travelling from link a to link b. We have

$$\widetilde{y}_{ab}(t) = min\left\{\phi_{ab}\, n_a^{\lambda-I}(t),\, \alpha_{ab}\, Q_a^{\lambda-I+1}(t),\, w\left(\alpha_{ab}\, N_a^{\lambda-I+1}(t) - n_{ab}^{\lambda-I+1}(t)\right)/v\right\} \tag{17}$$

Because of interference between turning vehicles and ahead vehicles, the total inflow of channelized queues area can be formulated as follows

$$y_a^{\lambda-I+1}(t) = min_{b\in B_m}\left\{\widetilde{y}_{ab}(t)/\alpha_{ab}\right\} \tag{18}$$

Inflow of each direction can be calculated by Eq. (18), gives

$$y_{ab}^{\lambda-I+1}(t) = \phi_{ab}\, y_a^{\lambda-I+1}(t) \tag{19}$$

$$y_{ab}^{\lambda-I+1}(t) = \phi_{ab}\, y_a^{\lambda-I+1}(t) \tag{20}$$

3) *Inflow of channelized cells*

$$y_{ab}^i(t) = min\left\{n_{ab}^{i-1}(t),\, \alpha_{ab}\, Q_a^i(t),\, w\left(\alpha_{ab}\, N_a^i(t) - n_{ab}^i(t)\right)/v\right\},\ \lambda - I + 1 < i \le \lambda \tag{21}$$

The total inflow of cell i can be calculated by Eq. (21), gives

$$y_a^i(t) = \sum_{b\in B_m} y_{ab}^i(t),\ \lambda - I + 1 < i \le \lambda \tag{22}$$

References

Daganzo CF (1994) The cell transmission model: a simple dynamic representation of highway traffic. Transp Res Part B 28(4):269–287

Daganzo CF (1995) The cell transmission model, part II: network traffic. Transp Res Part B 29(2):79–93

Daganzo CF, Laval JA (2005) Moving bottlenecks: a numerical method that converges in flows. Transp Res Part B 39(10):855–863

Gao ZY, Song YF (2002) A reserve capacity model of optimal signal control with user-equilibrium route choice. Transp Res Part B 36(4):313–323

Gentile G, Meschini L, Papola N (2007) Spillback congestion in dynamic traffic assignment: a macroscopic flow model with time-varying bottlenecks. Transp Res Part B 41(10):1114–1138

Jiang R, Wu QS, Zhu ZJ (2001) A new dynamics model for traffic flow. Chin Sci Bull 46(1):345–349

Lam WHK, Yin Y (2001) An activity-based time dependent traffic assignment model. Transp Res Part B 35(6):549–574

Lian AP, Gao ZY, Long JC (2007) A dynamic user optimal assignment problem based on link variables of the cell transmission model. Acta Automatica Sin 33(8):852–859 (in Chinese)

Lighthill MH, Whitham GB (1955) On kinematics wave: II a theory of traffic flow on long crowed roads. Proc R Soc Lond A 22:317–345

Lo HK (1999) A novel traffic signal control formulation. Transp Res Part A 33(5):433–448

Lo HK (2001) A cell-based traffic control formulation: strategies and benefits of dynamic timing plans. Transp Sci 35(2):149–164

Lo HK, Szeto WY (2002) A cell-based variational inequality formulation of the dynamic user optimal assignment problem. Transp Res Part B 36(5):421–443

Long JC, Gao ZY, Ren HL, Lian AP (2008) Urban traffic congestion propagation and bottleneck identification. Sci China F Inf Sci 55(7):948–964

Nagel K, Schreckenberg M (1992) A cellular automaton model for freeway traffic. J Phys I 2(12):2221–2228

Newell GF (1993) A simplified theory on kinematic wave in highway traffic, part I: general theory; part II: queuing at freeway bottlenecks; part III: multi-destination flows. Transp Res Part B 27(4):281–314

Ni DH, Leonard JD (2005) A simplified kinematic wave model at a merge bottleneck. Appl Math Model 29(11):1054–1072

Richards PI (1956) Shock waves on the highway. Oper Res 4:42–51

Roberg P (1954) Development and dispersal of area-wide traffic jams. Traffic Eng Control 35(6):379–386

Roberg P (1995) Distributed strategy for eliminating incident-based traffic jams from urban networks. Traffic Eng Control 36(6):348–355

Roberg P, Abbess CR (1998) Diagnosis and treatment of congestion problems in central urban areas. Eur J Oper Res 104(1):218–230

Roberg-Orenstein P, Abbess CR, Wright C (2007) Traffic jam simulation. J Maps v2007:107–121

Shang HY, Huang HJ, Gao ZY (2007) Locating the variable message signs by cell transmission model. Acta Phys Sin 56(8):4342–4347

Wright C, Roberg-Orenstein P (1999) Simple models for traffic jams and congestion control. Proc Inst Civ Eng Transp 135(3):123–130

Wright C, Roberg P (1998) The conceptual structure of traffic jams. Transp Policy 5:23–35

Ziliaskopoulos AK (2000) A linear programming model for the single destination system optimum dynamic traffic assignment problem. Transp Sci 34(1):37–49