TUTKIE working paper #990501
Revised 17 April 2003

Detecting Dynamic Traffic Assignment Capacity Paradoxes: Analysis of Non-Saturated Networks

Takashi AKAMATSU
Graduate School of Information Sciences, Tohoku University
Aoba, Sendai, Miyagi 980-8579, JAPAN (E-mail: akamatsu@plan.civil.tohoku.ac.jp)

Benjamin HEYDECKER
Centre for Transport Studies, University College London
Gower Street, London, WC1E 6BT, ENGLAND (E-mail: ben@transport.ucl.ac.uk)

Abstract

In Akamatsu and Heydecker (2003), we presented a necessary and sufficient condition for the occurrence of capacity paradoxes in general saturated networks, in which there is a queue on each link. The present paper extends that analysis to the more usual case of non-saturated networks, in which there are queues on some links but not on others. First we formulate dynamic user equilibrium (DUE) assignment in non-saturated networks. We then show how non-saturated networks can be reduced by direct network transformations to corresponding saturated ones in a way that will not affect occurrence of the capacity paradoxes: the reduced network, which is saturated, can then be tested for DUE assignment paradoxes to determine whether or not they will occur in the original one. This technique therefore yields a convenient method to examine whether or not the paradox occurs from information on the queuing patterns on the links of the network. Finally, as an application of the theory, we consider a range of example networks and associated applications, including investigation of a freeway ramp metering strategy. The analysis of these networks leads to an interesting finding: it is likely that in many situations metering or closing a freeway entrance ramp can play an effective role to reduce travel times not only on the freeway but also in the whole network including arterial streets.

Key Words: *Traffic Assignment, Queuing Networks, Dynamic User Equilibrium, Capacity Paradox, Ramp Metering*

Contents

In Akamatsu and Heydecker (2003), a theory of "capacity paradoxes" under dynamic user equilibrium (DUE) assignment was presented. Defining the paradox as the situation that increase (decrease) in capacity of a link leads to an increase (respectively decrease) in total travel time in a network, we established a necessary and sufficient condition for the paradox to occur in "*saturated networks*" with general structure, in which there is a queue present on each link. We then gave a graph theoretic interpretation of the condition, which enables us to identify network structures in which the paradox always occurs regardless of capacity and demand patterns. These results were derived under the seemingly restrictive assumption of "saturated networks", in which all links have queues.

In the present paper, we extend the theory to the more general case that includes a variety of "*non-saturated networks*" in which there are queues on some links but not on others. The key result of this analysis is to establish that in respect of the occurrence of DUE assignment capacity paradoxes, certain non-saturated networks can be reduced to equivalent saturated ones. This enables us to exploit the theory developed in the earlier paper (Akamatsu and Heydecker, 2003).

Throughout the present paper, we adopt the assumptions and notation of the earlier paper, with the exception of the assumption that the network is saturated in the sense that every link is congested. The main assumptions are summarised as follows:

(a) For a traffic assignment principle, we assume the dynamic user equilibrium (DUE) assignment: the DUE is defined as the state where at each time, no user can reduce his/her travel time by changing his/her route unilaterally;

(b) We consider only networks with a one-to-many travel pattern;

(c) For a link travel-time model, we employ a First-in-First-Out (FIFO) principle and the deterministic queue concept;

(d) We consider all links of the network that carry flow at some time in the DUE state: networks of this form can readily be extracted from arbitrary ones by excluding links to which no assignment is made.

The organisation of the paper is as follows. In the next section, we identify conditions for DUE assignments in non-saturated networks, and then we introduce a procedure for network reduction that does not affect occurrence of capacity paradoxes; this forms the key to our analysis of the properties of DUE solutions in certain non-saturated networks. In Section 2, we derive a necessary and sufficient condition for the capacity paradoxes to occur in non-saturated networks. In Section 3 we analyse a range of example networks, whilst in Section 4, as an application of the theory, we consider a simple ramp metering problem. Finally, concluding remarks are presented in Section 5.

1. DYNAMIC EQUILIBRIA ON NON-SATURATED NETWORKS

1.1 Notation and Non-Saturated Networks

We first summarise the main notation used in this paper; the notation is almost the same as that in the earlier paper (Akamatsu and Heydecker, 2003). Our model is defined on a transportation network $G[N, L, W]$ consisting of the set L of directed links with L elements, and the set N of nodes with N elements. The structure of a network is represented by a reduced node-link incidence matrix **A**, which is an (N-1) × L matrix obtained by removing the row corresponding to the unique origin from a standard incidence matrix. Under the assumptions mentioned in **Introduction** above, the equilibrium states in a network can be described by two kinds of variables, y_{ij}^s and t_i^s, decomposed with respect to origin departure-time s: t_i^s is the earliest arrival time at node i for a vehicle that departs from origin o at time s; y_{ij}^s is the link flow rate *with respect to s*. An N-1 dimensional column vector with elements dt_i^s / ds, and an L dimensional column vector with elements y_{ij}^s are denoted as $\dot{\mathbf{t}}(s)$ and $\mathbf{y}(s)$, respectively. The DUE travel time to traverse link (i,j), for users who depart the origin at time s (and arrive at the entrance of the link at time t_i^s), is denoted by c_{ij}^s, and $\dot{\mathbf{c}}(s)$ is an L dimensional column vector with elements dc_{ij}^s / ds. The capacity of link (i,j), m_{ij}, is given constant, and **M** is an L by L diagonal matrix whose a^{th} diagonal element represents the capacity of link a. The OD flow rate departing origin o at time s for each destination

d, $dQ_{od}(s)/ds$, is given, and $\dot{\mathbf{Q}}(s)$ is defined as an (N-1) dimensional vector with elements $dQ_{od}(s)/ds$.

In this paper, we deal with "*non-saturated networks*" in which there are queues on some links but not on others. We partition the link set L into two subsets, L_Q and L_F, based on the queuing state of links corresponding to the assignment with origin departure-time s:

if { a non-zero queue exists on link (i, j) ***or***

a queue starts to form when traffic that departs the origin at time s reaches it

(ie, $y_{ij}(s) \geq \mu_{ij}\dot{\tau}_i(s)$)}

then $(i,j) \in L_Q$; ***otherwise*** $(i,j) \in L_F$.

The number of links in the sets L, L_Q and L_F are denoted as L, $\mathrm{L_Q}$ and $\mathrm{L_F}$, respectively. Corresponding to this partition of the link set, we also split the link variables $\mathbf{y}$, $\dot{\mathbf{c}}$, link capacity matrix **M,** and the node-link ((N-1) by L) incidence matrix **A** into two parts, respectively:

$$\mathbf{y}(s) = \left[\frac{\mathbf{y}_Q(s)}{\mathbf{y}_F(s)}\right], \quad \dot{\mathbf{c}}(s) = \left[\frac{\dot{\mathbf{c}}_Q(s)}{\dot{\mathbf{c}}_F(s)}\right], \quad \mathbf{M} = \begin{bmatrix} \mathbf{M}_Q & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_F \end{bmatrix}, \qquad \mathbf{A} = [\mathbf{A}_Q \mid \mathbf{A}_F].$$

Throughout this paper, we analyse the DUE solution under the following assumptions:

1) there is a single origin of all traffic,

2) all links in L are used (ie $\mathbf{y}$(s)>0 for any s),

3) the sets L_Q and L_F are fixed (ie the queuing state of each link in L does not change) during the period of our analyses, and

4) the sets L_Q and L_F are given in advance.

A network that satisfies these assumptions is called a *non-saturated network* in this paper. For non-saturated networks, first we present a formulation of the DUE assignment in Section 1.1, and then we show how in some cases a non-saturated network can be reduced by elimination of links on which no queueing takes place to form an equivalent saturated networks" in Section 1.2.

1.2 Formulation

The DUE condition for non-saturated networks as defined above is formulated as follows. First, the shortest path condition and the flow constraints in the DUE state are the same as in the saturated networks. These can be stated, respectively, as:

$$\dot{\mathbf{c}}(s) + \mathbf{A}^T \ \mathrm{t}(s) = \mathbf{0} \qquad \forall s \tag{1a}$$

$$-\mathbf{A}\ \mathbf{y}(s) = \dot{\mathbf{Q}}(s) \qquad \forall s \tag{2a}$$

As for link travel times, we should distinguish unsaturated links that carry some flow in the DUE state from saturated links; the change in travel time on link $(i, j) \in L_F$ (*ie* an unsaturated link), dc_{ij}^s / ds, is always zero because the travel time on link $(i, j) \in L_F$ remains at the free-flow value:

$$\begin{cases} \dot{c}_{ij}(s) = y_{ij}^s / \mathrm{m}_{ij} - \dot{t}_i^s & (i, j) \in L_Q \\ \dot{c}_{ij}(s) = 0 & (i, j) \in L_F \end{cases} \qquad \forall s \tag{3a}$$

or equivalently,

$$\begin{cases} \dot{\mathbf{c}}_Q(s) = \mathbf{M}_Q^{-1}\ \mathbf{y}_Q(s) - \mathbf{A}_{Q+}^T\ \mathrm{t}(s) \\ \dot{\mathbf{c}}_F(s) = \mathbf{0} \end{cases} \qquad \forall s \tag{3b}$$

where $\mathbf{A}_{Q+}$ is the matrix obtained by setting all negative elements of $\mathbf{A}_Q$ to zero. Thus, we can formulate the DUE assignment in a non-saturated network as the simultaneous solution of equations of (1a), (2a) and (3).

To investigate the properties of the equilibrium solution, it is convenient to rewrite (1a) and (2a) in terms of the partitioned link sets L_Q and L_F:

$$\begin{cases} \dot{\mathbf{c}}_Q(s) + \mathbf{A}_Q^T\ \mathrm{t}(s) = \mathbf{0} \\ \dot{\mathbf{c}}_F(s) + \mathbf{A}_F^T\ \mathrm{t}(s) = \mathbf{0} \end{cases} \tag{1b}$$

$$-\mathbf{A}_Q\ \mathbf{y}_Q(s) - \mathbf{A}_F\ \mathbf{y}_F(s) = \dot{\mathbf{Q}}(s) \tag{2b}$$

For the links in L_Q, as in saturated networks (see Section 2 of Akamatsu and Heydecker), we have the following equations from (3b) and (1b):

$$\mathbf{y}_Q(s) = -\mathbf{M}_Q \mathbf{A}_{Q-}^T \, \mathrm{t}(s) \tag{4}$$

where $\mathbf{A}_{Q-}$ is a matrix obtained by setting all positive elements of $\mathbf{A}_Q$ to zero.

For the links in L_F, from (3b) and (1b),

$$\mathbf{A}_F^T \, \mathrm{t}(s) = \mathbf{0} \tag{5a}$$

Note here that $\dot{t}_o = 1$ always holds at the origin (from the definition of $\dot{t}_i(s)$) and that $\dot{t}_i$ for a node i that is connected to the origin by a single link has the same value as $\dot{t}_o$ ($ie \dot{t}_i = \dot{t}_o = 1$) *whenever link* (o, i) *is not saturated*. However, (5a) cannot reflect this because neither $\mathbf{A}_F$ nor $\mathbf{A}_Q$ contains a row corresponding to the origin. To overcome this limitation of (5a), we replace it with the following modified expression:

$$\mathbf{A}_F^T \, \mathrm{t}(s) = \mathrm{d} \tag{5b}$$

where d is an L_F dimensional vector with elements d_a:

$$\delta_a = \begin{cases} -1 & \text{if link } a \in L_F \text{ emanates from the origin node} \\ 0 & \text{otherwise} \end{cases}$$

Note that d = **0** holds when all links emanating from the origin are saturated.

Substituting (4) into (2b), we get

$$\mathbf{A}_Q \mathbf{M}_Q \mathbf{A}_{Q-}^T \, \mathrm{t}(s) - \mathbf{A}_F \mathbf{y}_F(s) = \dot{\mathbf{Q}}(s). \tag{6}$$

Thus, we can express the DUE conditions in a non-saturated network as a single system summarising the equations (5) and (6):

$$\left[\begin{array}{c|c} \mathbf{A}_Q \mathbf{M}_Q \mathbf{A}_{Q-}^T & -\mathbf{A}_F \\ \hline \mathbf{A}_F^T & \mathbf{0} \end{array}\right] \begin{bmatrix} \mathrm{t}(s) \\ \mathbf{y}_F(s) \end{bmatrix} = \begin{bmatrix} \dot{\mathbf{Q}}(s) \\ \mathrm{d} \end{bmatrix} \tag{7}$$

1.3 Equilibrium Solutions

Existence of the DUE solution in non-saturated networks seems to be evident because (7) is a system of linear equations and the number of variables is equal to that of the equations.

However, the solution for (7) should be consistent with a given classification of links that indicates whether each link in L belongs to L_Q or L_F; more specifically, the solution for links in L_F should satisfy the definition of non-saturated state:

$$y_{ij} < m_{ij}\, t_i(s) \qquad \forall (i,j) \in L_F . \tag{8}$$

Clearly, this condition is not necessarily satisfied by the solution of (7) for an arbitrary partition pattern of a link set, and hence non-existence of the DUE solution remains a possibility.

In what follows, we shall examine the uniqueness of the DUE solution, assuming that L_F is properly identified in advance (*ie* the set L_F that is consistent with the DUE solution is given). Thus we concentrate our discussion on the uniqueness of the solution for (7). The discussion can be classified into two cases depending on whether or not the matrix $\mathbf{V} \equiv \mathbf{A}_Q \mathbf{M}_Q \mathbf{A}_{Q-}^T$ is invertible:

Case (a): V is invertible.

When $\mathbf{V}$ is invertible, the equation (7) is equivalent to

$$\begin{bmatrix} \mathbf{V} & -\mathbf{A}_F \\ \mathbf{0} & \mathbf{A}_F^T \mathbf{V}^{-1} \mathbf{A}_F \end{bmatrix} \begin{bmatrix} \mathrm{t}(s) \\ \mathbf{y}_F(s) \end{bmatrix} = \begin{bmatrix} \dot{\mathbf{Q}}(s) \\ -\mathbf{A}_F^T \mathbf{V}^{-1} \dot{\mathbf{Q}}(s) + \mathrm{d} \end{bmatrix}. \tag{9}$$

Furthermore, the matrix $\mathbf{W} \equiv \mathbf{A}_F^T \mathbf{V}^{-1} \mathbf{A}_F$ is also invertible when $\mathbf{V}$ is invertible. Hence, we obtain the explicit solution for (9):

$$\mathbf{y}_F(s) = \mathbf{W}^{-1} \{ -\mathbf{A}_F^T \mathbf{V}^{-1} \dot{\mathbf{Q}}(s) + \mathrm{d} \}, \tag{10}$$

$$\mathrm{t}(s) = \mathbf{V}^{-1} \{ \dot{\mathbf{Q}}(s) + \mathbf{A}_F \mathbf{y}_F(s) \}, \tag{11}$$

and the flows $\mathbf{y}_Q$ on the links in L_Q can be obtained by substituting (11) into (4). Thus, we see that the DUE assignment has a unique solution in this case.

Note that the solution in (11) when $\mathrm{d} = \mathbf{0}$ (*ie* when all the links incident from the origin are saturated) can be represented as

$$\mathrm{t}(s) = [\mathbf{I} - \mathbf{P}]\, \mathbf{V}^{-1} \dot{\mathbf{Q}}(s), \tag{12}$$

where $\mathbf{P} \equiv \mathbf{V}^{-1} \mathbf{A}_F \mathbf{W}^{-1} \mathbf{A}_F^T$. Because the matrix $\mathbf{P}$ satisfies $\mathbf{P} = \mathbf{P}^2$, $\mathbf{P}$ is a projection matrix, and hence, **I** - **P** is also a projection matrix; in terms of geometry, **I** - **P** projects any vector (with respect to metric **V**) onto the null space of $\mathbf{A}_F^T$, whilst **P** projects any vector onto the column space of $\mathbf{A}_F$. This also means that the vector $\mathbf{V}^{-1}\dot{\mathbf{Q}}(s)$ can be decomposed into two orthogonal components $[\mathbf{I}-\mathbf{P}]\mathbf{V}^{-1}\dot{\mathbf{Q}} = t$ and $\mathbf{P}\,\mathbf{V}^{-1}\dot{\mathbf{Q}} = \mathbf{V}^{-1}\mathbf{A}_F\mathbf{y}_F$, which shows that t is normal to $\mathbf{V}^{-1}\mathbf{A}_F\mathbf{y}_F$ and hence that $\mathbf{V}\,t$ is normal to $\mathbf{A}_F\mathbf{y}_F$. Thus, we see from (12) that whenever V is invertible, the solution in a non-saturated network can be obtained by projecting the solution in a saturated network, $\mathbf{V}^{-1}\dot{\mathbf{Q}}(s)$, onto the null space of $\mathbf{A}_F^T$.

Case (b): V is not invertible.

This case occurs whenever a network contains a "non-saturated node" at which none of the incident links is saturated (*eg* node 8 of the network shown in Figure 1). Existence of non-saturated nodes will decrease the rank of the matrix **V**. The magnitude of this decrease is precisely the same as the number N_N of non-saturated nodes: the reason for this is the same as the "pure origin" case in saturated networks with a many-to-one OD pattern discussed in Akamatsu (2000), and hence the explanation is omitted here. This means that the rank of **V** is $N-1-N_N$.

On the other hand, the rank of $\mathbf{A}_F$ is $N_F - N_P$, where N_F is the number of nodes in N_F, the set of initial and terminal nodes of links in L_F, and N_P is the number of connected sub-graphs included in a graph $G_F[N_F, L_F]$. Therefore the rank of the coefficient matrix in the left-hand side of (7) is at most $(N-1-N_N) + (N_F - N_P)$. This suggests that the solution of (7) is not necessarily unique because the number of unknowns in (7), $N-1+L_F$, can exceed the rank of the coefficient matrix, $(N-1-N_N)+(N_F - N_P)$. To illustrate this possibility, we show a simple example for indeterminacy of the solution with respect to $\mathbf{y}_F(s)$. Consider the case where no links in L are saturated (*ie* $L=L_F$ and L_Q is null), then the equilibrium conditions do not tell us anything about $\mathbf{y}_F$ beyond flow conservation, $-\mathbf{A}_F\mathbf{y}_F(s) = \dot{\mathbf{Q}}(s)$: this shows that the equilibrium solution $\mathbf{y}_F$ is indeterminate. However, we can determine the equilibrium solution uniquely with respect to $t(s)$ even if we cannot obtain a unique $\mathbf{y}_F(s)$. This fact can be established by a simple network transformation technique, which we will discuss in the following section.

1.4 Reduced Networks

To discuss the uniqueness of the equilibrium solution $\dot{t}(s)$ in non-saturated networks with non-invertible **V**, we introduce a new transformed network, which we call a "*reduced network*". The transformation technique is also useful for connecting our earlier theory of capacity paradoxes in saturated network with one in non-saturated networks as well as in analysing the non-invertible case.

A reduced network is constructed by unifying the initial and terminal nodes of each unsaturated link of an original network into a single node. For example, consider the network with both unsaturated and saturated links as shown in Figure 1(a). The links (6, 8), (7, 8) and (8, 9), depicted by broken lines, are unsaturated, and the other links, depicted by solid lines, are saturated. Note here that there is no need to distinguish between the variables $\dot{t}_i$ for the initial and terminal nodes of each unsaturated link, say, $\dot{t}_6$ and $\dot{t}_8$ for link (6, 8), as different unknown variables because $\dot{t}_6 = \dot{t}_8$ holds for the unsaturated link ($\dot{c}_{68} = 0$) under the DUE state ($\dot{t}_8 - \dot{t}_6 = \dot{c}_{68}$). Hence, it is natural to unify the two nodes 6 and 8 in respect of this analysis. Repeating such unification of nodes for each unsaturated link (6, 8), (7, 8) and (8, 9), we obtain the reduced network depicted in Figure 1(b), from which all the unsaturated links have been removed.

Fig.1. Transformation of a non-saturated network to the corresponding reduced one

Having constructed a reduced network, G[N_R, L_R, W_R], from a non-saturated network, G[N, L, W], we consider the DUE assignment on G[N_R, L_R, W_R]. Because the reduced network is saturated, the solution is governed by

$$\left(\mathbf{A}_R\mathbf{M}_R\mathbf{A}_{R-}\right)\dot{t}_R(s) = \dot{\mathbf{Q}}_R(s)\,, \qquad (13)$$

where the variables with subscript R denote that they are defined on a reduced network. Note that the uniqueness of the solution $\dot{t}_R(s)$ for (13) is guaranteed as shown in Akamatsu and

Heydecker (2003). The solution for the original network G[N, L, W] is readily found from (13) for the reduced network. To show this, we introduce an $(\mathrm{N}-\mathrm{L}_\mathrm{F})\times\mathrm{N}$ matrix **R** defined as follows: the element in the i^{th} row and j^{th} column is 1 if node i in N_R corresponds to node j in N, 0 otherwise; then $\dot{t}_i(s)$ for each node in the set N is given by

$$\dot{\mathbf{t}}(s) = \mathbf{R}^T\ \dot{\mathbf{t}}_R(s), \tag{14}$$

and link flows $\mathbf{y}_Q$ on the saturated links in G[N, L, W] are determined uniquely by using this in (4).

As proved in **Appendix A**, the unique solution ($\dot{\mathbf{t}}(s)$, $\mathbf{y}_Q$) obtained by (13), (14) and (4) is consistent with the DUE conditions in a non-saturated network G[N, L, W]. In particular, the solution satisfies (7) notwithstanding that the flows $\mathbf{y}_F$ on non-saturated links may be indeterminate: we note that it is possible that substituting a value of $\mathbf{y}_Q$ into (2b) does not determine the value of $\mathbf{y}_F$ uniquely. Thus we can conclude that, by using the reduced network, the dynamic equilibrium solution ($\dot{\mathbf{t}}(s)$, $\mathbf{y}_Q$) can be obtained even if **V** is not invertible and that $\dot{\mathbf{t}}(s)$ is uniquely determined though $\mathbf{y}_F$ might not be in all cases.

A few remarks are in order concerning reduced networks. First, the technique detailed above, reducing a network to a saturated one for which (13) is then solved, is valid not only for case (b) but also for case (a) of **1.3**, in so far as we aim only to obtain $\dot{\mathbf{t}}(s)$ and $\mathbf{y}_Q$. This is evident because this technique can be applied whether or not the matrix **V** is invertible.

Second, *if the sets L and L_Q do not change during the period of analysis*, then no saturated link of the original network is deleted in constructing the reduced network.

However, it is possible to identify cases in which saturated links are eliminated in the procedure of unifying nodes. Consider a saturated link whose initial and terminal nodes are the terminal nodes of two non-saturated links that have the same initial node (see Figure 2); these nodes would be unified into a single node, and therefore, the saturated link may be removed (or forms a loop at the single node that results from this). Networks with combinations of saturated and non-saturated links such as this cannot arise in the DUE state *unless the sets L and L_Q change during the analysis period*. We illustrate this by considering the network example in

Figure 2. Suppose that both routes $(1,3)$ and $(1,2,3)$ are used in the DUE state; then for the users of route $(1,3)$, $\dot{t}_3(s) = \dot{t}_1(s) + \dot{c}_{13}(s) = 1 + 0 = 1$ because link $(1,3)$ is not saturated. On the other hand, for the users of route (1, 2, 3), $\dot{t}_3(s) = \dot{t}_1(s) + \dot{c}_{12}(s) + \dot{c}_{23}(s) = 1 + 0 + \dot{c}_{23}(s) \neq 1$ because link (2, 3) is saturated; these contradict each other unless the particular case arises that link (2, 3) is saturated with $l_{23}(s) = m_{23}$ so that $\dot{c}_{23}(s) = 0$. Furthermore, as shown in the third remark below, this particular case (link (2, 3) is saturated and $\dot{c}_{23}(s) = 0$) can occur only when the set L_Q changes.

Fig.2. An example of a queuing pattern that causes deletion of a saturated link

Third, our theory in the present paper can be applied directly to intervals throughout which the the sets L and L_Q remain unchanged. We can extend the application of this analysis by decomposing an analysis period into intervals of this kind, which are delimited by instants at which the sets L and L_Q do change. Contrary to the assertion in the second remark above, this allows a class of queuing patterns that leads to deletion of saturated links in constructing the reduced network to occur. We now present a simple example depicted in Figure 2 to show how this can occur. Suppose that the free-flow travel time on link 2 exceeds the sum of those on links 1 and 3, whilst the capacity of link 3 is less than the demand for travel to node 3 and the capacities on links 1 and 2 are sufficiently large that they will not be congested. In the DUE state of this network, link 2 is not used at all until the travel time of the route using links 1 and 3 ($c_{12} + c_{23}$) grows to the free-flow travel time of link 3 (c_{13}); link 2 starts to be used at the first instant, s, when $c_{12} + c_{23}$ reaches c_{13}; thus, the set L must change at time s in order for the queuing pattern in Figure 2 to occur. To include this particular class of queuing patterns into our analysis, we unify the initial and terminal nodes of each unsaturated link *even if the two nodes are also the initial and terminal nodes of a saturated link*, and repeat such procedure until all unsaturated links have been deleted. In this extended construction of a reduced network, some saturated links of an original network may be deleted, but only those for which the cost is constant over time because it is stabilised by the assignment in equilibrium to an alternative set of links, each of which is unsaturated. An example of this is the reduction of the network in Figure 2 to a single node by deletion not only of unsaturated links (1,2) and (1,3) but also of the saturated link (2,3) to which they provide an alternative.

However, in so far as our concern is the analysis of the capacity paradox, this deletion of saturated links will cause no problem; for example, we see immediately that change in the capacity of the saturated link (2,3) causes no paradox, because $\dot{t}_3$ is not at all affected by the change in the capacity due to the requirement of $\dot{t}_1 = \dot{t}_2 = \dot{t}_3$; that is, we can ignore any deleted links, regardless of whether they are saturated or unsaturated, in detecting capacity paradoxes. Furthermore, we can also obtain equilibrium flow y_{ij} on each deleted link in L_Q by the following formula:

$$y_{ij} = m_{ij}\, \dot{t}_j = m_{ij}\, \dot{t}_i = m_{ij}\, \dot{t}_{U(i,j)},$$

where $U(i, j)$ denotes the node (in the reduced network) that is obtained by unifying nodes i and j of the original network, and $\dot{t}_{U(i,j)}$ is the time-derivative of equilibrium arrival time to node $U(i, j)$ in the reduced network. The validity of this formula is apparent: $\dot{t}_i = \dot{t}_j = \dot{t}_{U(i,j)}$ always holds since the deleted link in L_Q connects the two nodes that are also connected by an unsaturated link.

Finally, we should note that the DUE flow pattern on a reduced network can contain loops (consisting of several links) although the corresponding flow pattern on the original non-saturated network cannot. This implies that we cannot apply the procedure DUE_SN described in Akamatsu and Heydecker (2003) to solve equations (13) because DUE_SN is designed for saturated acyclic networks; instead, to solve (13) we can use a standard procedure a system of linear equations.

2. CONDITIONS FOR OCCURRENCE OF CAPACITY PARADOXES

As in Akamatsu and Heydecker (2003), we define the paradox to be the situation where increasing (decreasing) capacity of a certain link causes the increase (respectively decrease) of total travel time C in the whole of a network. That is, the paradox occurs if and only if

$$\frac{\partial C}{\partial \mu_a} = \int_0^T \dot{\mathbf{Q}}(s)^T \frac{\partial \tau(s)}{\partial \mu_a} ds \quad = \quad \int_0^T \dot{\mathbf{Q}}(s)^T \left\{ \int_0^s \frac{\partial \tau(t)}{\partial \mu_a} dt \right\} ds \quad \geq \quad 0 \tag{15}$$

A formula for calculating $\partial C / \partial \mu_a$ can be obtained from the sensitivity of the DUE solution $\tau(s)$ with respect to the change in capacity of link a, classifying the problem into two cases depending on whether or not the matrix $\mathbf{V}$ is invertible:

2.1 Invertible Case

Let us first consider the two equilibrium solutions, $\tau(\mu)$ and $\tau(\mu + \Delta\mu)$ respectively, for the capacity patterns μ and $\mu + \Delta\mu$. From (7), the respective solutions are governed by the following equations:

$$\begin{bmatrix} \mathbf{V}(\mu) & -\mathbf{A}_F \\ \mathbf{A}_F^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tau(\mu) \\ \mathbf{y}_F(\mu) \end{bmatrix} = \begin{bmatrix} \dot{\mathbf{Q}}(s) \\ d \end{bmatrix}, \tag{16a}$$

$$\begin{bmatrix} \mathbf{V}(\mu) + \mathbf{V}(\Delta\mu) & -\mathbf{A}_F \\ \mathbf{A}_F^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tau(\mu + \Delta\mu) \\ \mathbf{y}_F(\mu + \Delta\mu) \end{bmatrix} = \begin{bmatrix} \dot{\mathbf{Q}}(s) \\ d \end{bmatrix}. \tag{16b}$$

We then compare the solutions; subtracting (16a) from (16b), we have

$$\mathbf{V}(\mu)\{\tau(\mu + \Delta\mu) - \tau(\mu)\} - \mathbf{A}_F \{\mathbf{y}_F(\mu + \Delta\mu) - \mathbf{y}_F(\mu)\} = -\mathbf{V}(\Delta\mu)\ \tau(\mu + \Delta\mu), \tag{17a}$$

$$\mathbf{A}_F^T \{\tau(\mu + \Delta\mu) - \tau(\mu)\} = \mathbf{0}. \tag{17b}$$

Consider the case that $\Delta\mu = [0,...,0, \mu_a, 0...,0]$ (*ie* only the capacity of link a changes, and is increased by $\Delta\mu_a$). Dividing both sides of equations (17) by $\Delta\mu_a$, considering the identity $\mathbf{V}(\Delta\mu)/\Delta\mu_a = \mathbf{A}\,\mathbf{I}_a \mathbf{A}_-^T$, where $\mathbf{I}_a$ is an $L \times L$ matrix whose a^{th} diagonal element is 1 and all other elements are 0, and taking the limit of $\Delta\mu_a \to 0$, we obtain

$$\mathbf{V}(\mathrm{m})\frac{\partial \mathrm{t}(\mathrm{m})}{\partial m_a} - \mathbf{A}_F \frac{\partial \mathbf{y}_F(\mathrm{m})}{\partial m_a} = -\mathbf{A}_Q \mathbf{I}_a \mathbf{A}_{Q-}^T \mathrm{t}(\mathrm{m}), \tag{18a}$$

$$\mathbf{A}_F^T \frac{\partial \mathrm{t}(\mathrm{m})}{\partial m_a} = \mathbf{0}. \tag{18b}$$

Note here that the coefficient matrix of the left hand side of (18) is the same as (7). Hence, it follows immediately that the solution $\partial \mathrm{t}(s)/\partial m_a$ of (18a) can be obtained by projecting $\mathbf{V}^{-1}\{-\mathbf{A}_Q \mathbf{I}_a \mathbf{A}_{Q-}^T \mathrm{t}(\mathrm{m})\}$ onto null space of $\mathbf{A}_F$ as in (12):

$$\begin{aligned}\frac{\partial \mathrm{t}(\mathrm{m})}{\partial m_a} &= -[\mathbf{I}-\mathbf{P}]\,\mathbf{V}(\mathrm{m})^{-1}\{\mathbf{A}_Q \mathbf{I}_a \mathbf{A}_{Q-}^T \mathrm{t}(\mathrm{m})\} \\ &= -[\mathbf{I}-\mathbf{P}]\,\mathbf{V}(\mathrm{m})^{-1}\mathbf{A}_Q \mathbf{I}_a \mathbf{A}_{Q-}^T [\mathbf{I}-\mathbf{P}]\,\mathbf{V}(\mathrm{m})^{-1}\dot{\mathbf{Q}}(s)\end{aligned} \tag{19}$$

From this sensitivity formula for the DUE solution and the definition of *TC*, we have the following proposition:

Proposition 1(a). *The capacity paradox in a non-saturated network with an invertible matrix* **V** *occurs if and only if*

$$\frac{\partial C}{\partial m_a} = -\int_0^T \dot{\mathbf{Q}}(s)^T [\mathbf{I}-\mathbf{P}]\,\mathbf{V}^{-1}\mathbf{A}_Q \mathbf{I}_a \mathbf{A}_{Q-}^T [\mathbf{I}-\mathbf{P}]\,\mathbf{V}^{-1}\mathbf{Q}(s)\,ds \geq 0. \tag{20}$$

Note that the paradox cannot occur on a link in L_F (*ie* a non-saturated link) whilst the link sets L_Q and L_F remain fixed (*ie* the queuing pattern does not change) because $\partial T/\partial m_a$ for link $a \in L_F$ is always zero. Thus it can never occur on a link in L_F in respect of capacity increase, and could only occur on one if capacity is reduced to below the flow that is currently assigned.

2.2 Non-Invertible Case

We now consider the case in which the matrix **V** is not invertible. In this case, the flow pattern $\mathbf{y}_F$ on links in L_F is not necessarily determined uniquely. However, this does not cause any problems in deriving the condition for occurrence of capacity paradoxes. The reason for this is as follows:

1) $\partial C/\partial m_a$ can be calculated using information only on $\partial \mathrm{t}(s)/\partial m_a$ as can be seen from (15) which include no explicit information on $\mathbf{y}_F$;

2) $\partial \mathrm{t}(s)/\partial m_a$ is affected by neither $\mathbf{y}_F$ nor $\mathbf{M}_F$ because $\mathrm{t}(s)$ is independent of both $\mathbf{y}_F$ and $\mathbf{M}_F$ as we have seen in Section 1;

3) from 1) and 2), we conclude that $\partial C/\partial m_a$ is independent of both $\mathbf{y}_F$ and $\mathbf{M}_F$.

In the light of the fact that the paradox occurrence is independent of both $\mathbf{y}_F$ and $\mathbf{M}_F$, it is convenient to consider the problem defined in a reduced network. Because the DUE assignments in a reduced network have the same form as that in a saturated network, it is evident that the necessary and sufficient condition for the paradox to occur corresponds to (16) of Akamatsu and Heydecker (2003) when applied to the reduced network:

Proposition 1(b). *The capacity paradox occurs in a non-saturated network if and only if*

$$\frac{\partial C}{\partial m_a} = -\int_0^T \dot{\mathbf{Q}}_R(s)^T \mathbf{V}_R^{-1} \mathbf{A}_R \mathbf{I}_a \mathbf{A}_{R-}^T \mathbf{V}_R^{-1} \mathbf{Q}_R(s)\, ds \geq 0, \tag{21}$$

where $\mathbf{V}_R \equiv \mathbf{A}_R \mathbf{M}_R \mathbf{A}_{R-}^T$.

Remarks:

(1) For the reason why the right end term in the integral in (20) and (21) is not $\dot{\mathbf{Q}}(s)$ but $\mathbf{Q}(s)$, see **Appendix B**.

(2) The result in Akamatsu and Heydecker (2003) enables us to be sure that the matrix $\mathbf{V}_R$ is invertible even when $\mathbf{V}$ is not.

3. ANALYSIS OF EXAMPLE NETWORKS

We now explore the application of these analyses of non-saturated networks. We consider separately the case of a network in which the matrix **V** is invertible and that in which it is not.

3.1 Invertible Case

Example (1). Consider the network shown in Figure 3(a), where node 1 is the only origin; nodes 2 and 3 are destinations; the maximum exit flow rate (capacity) of link a is given by m_a ($a = 1, 2, 3$); links 1 and 2 are saturated but link 3 is not saturated.

Fig.3(a). An example network in which the matrix **V** is invertible

The reduced incidence matrix **A** and the matrix $\mathbf{A}_{Q-}$ for this network are given by

$$\mathbf{A} = \left[\begin{array}{cc|c} -1 & 0 & 1 \\ 0 & -1 & -1 \end{array}\right] = [\mathbf{A}_Q \mid \mathbf{A}_F], \quad \mathbf{A}_{Q-} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \mathrm{d} = 0\,.$$

Direct calculation yields

$$\mathbf{V} \equiv \mathbf{A}_Q \mathbf{M}_Q \mathbf{A}_{Q-}^T = \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix}, \qquad \mathbf{V}^{-1} = \begin{bmatrix} 1/m_1 & 0 \\ 0 & 1/m_2 \end{bmatrix},$$

$$\mathbf{W} \equiv \mathbf{A}_F^T \mathbf{V}^{-1} \mathbf{A}_F = \frac{1}{m_1} + \frac{1}{m_2}, \quad \mathbf{W}^{-1} \mathbf{A}_F^T \mathbf{V}^{-1} = \frac{1}{m_1 + m_2} [m_2 \quad -m_1], \tag{22a}$$

$$\mathbf{P} \equiv \mathbf{V}^{-1} \mathbf{A}_F \mathbf{W}^{-1} \mathbf{A}_F^T = \frac{1}{m_1 + m_2} \begin{bmatrix} m_2 & -m_2 \\ -m_1 & m_1 \end{bmatrix}, \quad [\mathbf{I} - \mathbf{P}]\mathbf{V}^{-1} = \frac{1}{m_1 + m_2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}. \tag{22b}$$

Substituting (22b) into (12), we find

$$\dot{t}_2(s) = \frac{1}{m_1 + m_2} \left\{ \dot{Q}_{12}(s) + \dot{Q}_{13}(s) \right\} = \dot{t}_3(s)\,. \tag{23}$$

Reference to the original network shown in Figure 3(a) shows that this arises because link (2, 3) is used in the DUE assignment but is not saturated. From (23) and (4), we have that in the DUE state, the flows y_1 and y_2 on the saturated links are given by

$$y_1(s) = \frac{\mu_1}{\mu_1 + \mu_2}\left\{\dot{Q}_{12}(s) + \dot{Q}_{13}(s)\right\}, \quad y_2(s) = \frac{\mu_2}{\mu_1 + \mu_2}\left\{\dot{Q}_{12}(s) + \dot{Q}_{13}(s)\right\}. \qquad (24a)$$

The DUE flow y_3 on the non-saturated link (2, 3) is given by substituting (22a) into (10):

$$y_3(s) = \left(\frac{1}{m_1 + m_2}\right)\left(-m_2\,\dot{Q}_{12}(s) + m_1\,\dot{Q}_{13}(s)\right). \qquad (24b)$$

We note that the same solution can be obtained by using the reduced network shown in Figure 3(b).

We now proceed to verify the requirement for the existence of the DUE solution; the solution should satisfy $0 \le y_3(s) \le m_3 \dot{t}_2(s)$ because we have treated link 3 as being non-saturated. This requirement is equivalent to

$$\left(\frac{m_1 - m_3}{m_2 + m_3}\right)\dot{Q}_{13}(s) \;<\; \dot{Q}_{12}(s) \;<\; \frac{m_1}{m_2}\dot{Q}_{13}(s). \qquad (25)$$

If this condition is not satisfied, then the DUE solution for this network will not have the queuing pattern that is shown in Figure 3(a).

Fig.3(b). The reduced network for the one in Fig.3(a)

Next, let us examine whether or not the paradox occurs in this network. Because link 3 is not saturated, we see immediately that marginal changes in the capacity of link 3 will not affect the DUE assignment and hence cannot cause the paradox. As for links 1 and 2, substituting the expressions in (22) for $\mathbf{V}^{-1}$ and $\mathbf{P}$ into (20), we have

$$\frac{\partial C}{\partial m_1} = \frac{\partial C}{\partial m_2} = -\frac{1}{(m_1 + m_2)^2}\int_0^T \{Q_{12}(s) + Q_{13}(s)\}\{\dot{Q}_{12}(s) + \dot{Q}_{13}(s)\}\,ds. \qquad (26)$$

Because (26) is always negative, this shows that the paradox can never occur in this network.

Example (2). We now consider a second example network in which the matrix **V** is invertible. The network shown in Figure 4(a) has the same topology as that in Figure 3(a) and differs only in the queuing pattern: in this case links 1 and 3 are saturated but link 2 is not saturated.

Fig.4(a). A second example network in which the matrix **V** is invertible

The matrices $\mathbf{A}_F$, $\mathbf{A}_Q$ and $\mathbf{A}_{Q-}$ for this network are given by

$$\mathbf{A}_F = \begin{bmatrix} 0 \\ -1 \end{bmatrix}, \qquad \mathbf{A}_Q = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}, \qquad \mathbf{A}_{Q-} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \qquad \mathrm{d} = -1.$$

Direct calculation yields

$$\mathbf{V} \equiv \mathbf{A}_Q \mathbf{M}_Q \mathbf{A}_{Q-}^T = \begin{bmatrix} m_1 & -m_3 \\ 0 & m_3 \end{bmatrix}, \quad \mathbf{V}^{-1} = \begin{bmatrix} 1/m_1 & 1/m_1 \\ 0 & 1/m_3 \end{bmatrix}, \tag{27a}$$

$$\mathbf{W} \equiv \mathbf{A}_F^T \mathbf{V}^{-1} \mathbf{A}_F = \frac{1}{m_3}, \qquad \mathbf{W}^{-1} \mathbf{A}_F^T \mathbf{V}^{-1} = [0 \quad -1], \tag{27b}$$

$$\mathbf{P} \equiv \mathbf{V}^{-1} \mathbf{A}_F \mathbf{W}^{-1} \mathbf{A}_F^T = \frac{1}{m_1} \begin{bmatrix} 0 & m_3 \\ 0 & m_1 \end{bmatrix}, \qquad [\mathbf{I} - \mathbf{P}]\ \mathbf{V}^{-1} = \frac{1}{m_1} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}. \tag{27c}$$

The DUE flow y_2 on the non-saturated link 2 is given by substituting (27b) into (10):

$$y_2(s) = \dot{Q}_{13}(s) - m_3. \tag{28}$$

Substituting (27a) and (28) into (11) yields

$$\dot{t}_2(s) = \frac{1}{m_1}(\dot{Q}_{12}(s) + m_3), \qquad \dot{t}_3(s) = 1. \tag{29}$$

From (28) and (4), we obtain the DUE flows y_1 and y_3 on the saturated links 1 and 3 as:

$$y_1(s) = \dot{Q}_{12}(s) + m_3, \qquad y_3(s) = m_3 \dot{t}_3(s) = m_3. \tag{30}$$

It is also possible to obtain the same solution by using the reduced network shown in Figure 4(b). We note that the reduced network for this case contains a loop, unlike the saturated networks considered in Akamatsu and Heydecker (2003). However, this does not prevent us from applying the desired analysis.

Because we supposed link 2 to be non-saturated, the DUE solution should satisfy $0 < y_2(s) < m_2 t_1'$. This requirement reduces to $m_3 \le \dot{Q}_{13}(s) < m_2 + m_3$.

Fig.4(b). The reduced network for the one in Fig.4(a)

Next, we investigate occurrence of the capacity paradox in this network, and show that in this case it can occur. Because link 2 is not saturated, we see immediately that marginal changes in the capacity of link 2 cannot cause the paradox. As for links 1 and 3, substituting $\mathbf{V}^{-1}$ and $\mathbf{P}$ calculated in (27) into (20), we have

$$\frac{\partial C}{\partial m_1} = -\frac{1}{m_1}\int_0^T Q_{12}(s)\,\dot{Q}_{12}(s)\,ds\,, \tag{31a}$$

$$\frac{\partial C}{\partial m_3} = \frac{1}{m_1}\int_0^T Q_{12}(s)\,ds\,. \tag{31b}$$

This means that changing the capacity of link 3 always causes the paradox whilst marginal changes to the capacity of link 1 will not do so.

We can understand this by direct reference to the network shown in Figure 4(a) as follows. In this network, because link 2 is used in the DUE assignment, the cost incurred by all travellers from the origin node 1 to destination node 3 is equal to the cost of using link 2. Because link 2 is not saturated, the cost of using it remains constant at the free-flow value, so that the cost of travel from the origin node 1 to destination 3 also remains constant. Because link 3 is used in the DUE assignment, the route (1, 3) carries some flow, so from the equilibrium condition, at each time s we have $c_1(s) + c_3(s) = c_2(s)$. Now an increase in the capacity of link 1 will reduce the cost of travel if the flow remains constant. Any reduction in cost of using link 1 will cause a complementary increase in the cost of using link 3 in order to achieve constancy of their sum as is required for equilibrium: this will be achieved by an increase in the assignment to the route (1, 3) which will moderate, but not nullify, the reduction in cost on link 1. Thus the increase in capacity of link 1 leads to a reduction in the equilibrium cost of travel to destination 2 and an increase in assignment to route (1, 3), but does not change the equilibrium cost of travel to destination 3. Changes to the capacity of link 1 therefore cause variation in costs of the sign that would be expected from naïve consideration. On the other hand, an increase in capacity of link 3 will reduce the cost of travel on that link at constant flow. Any reduction in the cost of using link 3 will cause a complementary increase in that on link 1 in order to achieve constancy of their sum as is required for equilibrium: as before, this will be achieved by an increase in the assignment to route (1, 3). In this case, the resulting increase in cost on link 1 will affect all travellers from the origin to destination 2. Accordingly, changes to the capacity of link 3 will not affect the equilibrium cost of travel to destination 3 but, because of their effect on assignments to destination 3, will influence the cost of travel to destination 2 in a way that causes paradoxical effects.

3.2 Non-Invertible Case

Example (3). We consider an example network in which the matrix $\mathbf{V}$ is not invertible but in which the reduced matrix $\mathbf{V}_R$ is. The network shown in Figure 5 (a) has the same topology as those in Figures 3(a) and 4(a) but differs in the queuing pattern: in this case, links 2 and 3 are saturated but link 1 is not.

Fig.5 (a). An example network in which the matrix **V** is non-invertible

The matrices $\mathbf{A}_F$, $\mathbf{A}_Q$ and $\mathbf{A}_{Q-}$ for this network are given by

$$\mathbf{A} = \begin{bmatrix} -1 & 0 & 1 \\ 0 & -1 & -1 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_F & \mathbf{A}_Q \end{bmatrix}, \quad \mathbf{A}_{Q-} = \begin{bmatrix} 0 & 0 \\ -1 & -1 \end{bmatrix}.$$

Hence

$$\mathbf{V} \equiv \mathbf{A}_Q \mathbf{M}_Q \mathbf{A}_{Q-}^T = \begin{bmatrix} 0 & -m_3 \\ 0 & m_2 + m_3 \end{bmatrix} . \tag{32}$$

This matrix **V** is clearly non-invertible, although the topology of the original network is the same as in Examples (1) and (2) with an invertible V. Despite this singularity, we can obtain the equilibrium solution as follows. First, note that we do not have to regard t_2 as an unknown variable to be solved from (6). Because in this case link 1 is not saturated, $dc_1(s)/ds = 0$ holds and it follows that

$$\dot{t}_1(s) = \dot{t}_2(s) = 1 . \tag{33a}$$

Considering equation (6) applied at node 3, we get

$$\dot{t}_3(s) = \frac{\dot{Q}_{13}(s)}{m_2 + m_3} , \tag{33b}$$

and substituting this into (4), we obtain

$$y_2(s) = \frac{m_2}{m_2 + m_3}\dot{Q}_{13}(s) , \qquad y_3(s) = \frac{m_3}{m_2 + m_3}\dot{Q}_{13}(s) . \tag{34a}$$

The DUE flow y_1 on the non-saturated link 1 can then be calculated from (2b) as:

$$y_1(s) = \dot{Q}_{12}(s) + y_3(s) . \tag{34b}$$

This flow should satisfy $0 \le y_1(s) \le m_1\, \dot{t}_1(s)$, which leads to the requirement on the problem specification

$$\dot{Q}_{12}(s) + \frac{m_3}{m_2 + m_3}\dot{Q}_{13}(s) < m_1 \tag{35}$$

in order for the DUE solution to give rise to this queueing pattern. Note that the same solution can be obtained by using the reduced network shown in Figure 5 (b).

Fig.5 (b). The reduced network for the one in Fig.5(a)

We now examine occurrence of the capacity paradox in this case. Because link 1 is not saturated, marginal changes in the capacity of link 1 will not affect costs and hence cannot cause paradoxical behaviour. As for links 2 and 3, substituting $\mathbf{V}_R^{-1} = (m_2 + m_3)^{-1}$ into (21), we have

$$\frac{\partial C}{\partial \mu_2} = \frac{\partial C}{\partial \mu_3} = -\frac{1}{(\mu_2 + \mu_3)^2} \int_0^T Q_{13}(s)\, \dot{Q}_{13}(s)\, ds = -\left(\frac{dQ_{13}(T)}{\mu_2 + \mu_3} \right)^2 < 0 . \tag{36}$$

This means that the paradox can never occur in this network with this queueing pattern.

Example (4). Finally, we consider a second example network in which the matrix $\mathbf{V}$ is not invertible. The network is shown in Figure 6 (a), which has the same topology as the famous Braess' (1968) network with a single OD pair (1, 4); as for the queuing pattern, we assume that links (1,2), (2,3) and (3,4) are saturated but links (1,3) and (2,4) are not.

Fig.6 (a). A second example network in which the matrix **V** is non-invertible

The matrices $\mathbf{A}_F$, $\mathbf{A}_Q$ and $\mathbf{A}_{Q-}$ for this network are given by

$$\mathbf{A} = \left[\begin{array}{ccc|cc} 1 & 0 & 0 & 1 & 0 \\ -1 & 1 & 0 & 0 & 1 \\ 0 & -1 & 1 & -1 & 0 \\ 0 & 0 & -1 & 0 & -1 \end{array}\right] = \left[\mathbf{A}_Q \mid \mathbf{A}_F\right], \qquad \mathbf{A}_Q = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix}.$$

Hence, we have

$$\mathbf{V} \equiv \mathbf{A}_Q \mathbf{M}_Q \mathbf{A}_{Q-}^T = \begin{bmatrix} 0 & -m_{12} & 0 & 0 \\ 0 & m_{12} & -m_{23} & 0 \\ 0 & 0 & m_{23} & -m_{34} \\ 0 & 0 & 0 & m_{34} \end{bmatrix}. \tag{37}$$

Although this matrix is singular, we can obtain the equilibrium solution by using the reduced network shown in Figure 6 (b).

Fig.6 (b). The reduced network for the one in Fig.6(a)

In the reduced network, nodes 1 and 3 in the original network are unified into a single node A because link (1,3) in the original network is not saturated (*ie* $dc_{13}(s)/ds = 0$), which implies

$$\dot{t}_1 = \dot{t}_3 = 1. \tag{38a}$$

Similarly, nodes 2 and 4 in the original network are unified into a single node B, and it follows

$$\dot{t}_2 = \dot{t}_4 . \tag{38b}$$

The flow conservation at node B of the reduced network yields

$$(m_{12} + m_{34})\dot{t}_4 = \dot{Q}_{14} + m_{23}\dot{t}_3 \tag{39}$$

Hence, the equilibrium solution is given by

$$\dot{t}_2 = \dot{t}_4 = \frac{\dot{Q}_{14} + m_{23}\dot{t}_3}{m_{12} + m_{34}}, \tag{40}$$

Substituting this into (4), we obtain

$$y_{12} = m_{12}\dot{t}_2 = \frac{m_{12}}{m_{12} + m_{34}}(\dot{Q}_{14} + m_{23}), \quad y_{34} = m_{34}\dot{t}_4 = \frac{m_{34}}{m_{12} + m_{34}}(\dot{Q}_{14} + m_{23}), \tag{41a}$$

$$y_{23} = m_{23}\dot{t}_3 = m_{23} . \tag{41b}$$

The DUE flows on non-saturated links can then be calculated from (2b) as:

$$y_{13} = \dot{Q}_{14} - y_{12} = \frac{\dot{Q}_{14}m_{34} - m_{12}m_{23}}{m_{12} + m_{34}}, \quad y_{24} = \dot{Q}_{14} - y_{34} = \frac{\dot{Q}_{14}m_{12} - m_{34}m_{23}}{m_{12} + m_{34}} . \tag{41c}$$

These flows in non-saturated links should satisfy $y_{13} < m_{13}\dot{t}_3$ and $y_{24} < m_{24}\dot{t}_4$ in the DUE state. These requirement reduce to

$$\dot{Q}_{14} < m_{13} + \frac{m_{12}}{m_{34}}(m_{13} + m_{23}) \text{ and } \dot{Q}_{14} < \frac{m_{23}}{m_{12} - m_{24}}(m_{24} + m_{34}) .$$

The solution (40) clearly shows occurrence of the capacity paradox in this case: OD travel time t_4 increases with the increase in m_{23}, which implies that changes in the capacity of link (2,3) always causes the dynamic capacity paradox corresponding to Braess' static one.

From these examples, we can see a simple but useful rule that *paradox always occurs if a controlled link* (in which we are changing the capacity) *forms a part of loops (cycles) in a reduced network*. Recall that we have seen the occurrence of the capacity paradox in Examples (2) and (4). In Example (2), links 1 and 3 of the reduced network (Fig.4(b)) indeed form a loop; and in Example (4), links (1,2) (or (3,4)) and (2,3) of the reduced network (see Fig.6(b)) form a loop. We shall discuss this point again in the next section, where ramp metering applications in more complicated networks are presented.

4. APPLICATIONS - Examination of Ramp Metering Operations in a Ladder Network

In this section, we consider a simple ramp metering/closing problem as an application of the theory developed so far. Note that ramp metering/closing in freeway entrance/exit ramps can be represented in our dynamic assignment model as decreasing the capacity of certain links, and that any occurrence of the capacity paradox in the links would have important implications for the effectiveness of ramp metering in managing congestion. We note that the usual intention of these strategies is to reduce the cost of travel in some or all of a road network by making appropriate temporary reductions in the capacity on the freeway entry ramps. When we allow for consequent reassignment, as might occur in response to a recurrent strategy operating to a fixed-time schedule, this traffic management strategy relies for its success on behaviour that in other contexts might be regarded as paradoxical: the reduction in capacity on the freeway entry ramps is required to result in a reduction in total cost of travel. A more local view would be that the beneficiaries should be travellers who join the freeway upstream of the entry ramps that are controlled in this way, but that is clearly inequitable. This observation means that the theory developed in Sections 1 and 2 can be used to analyse whether or not ramp metering/closing operations in a network will be effective when the total cost of travel in the network is considered.

4.1 Example Network and Possible Queuing Patterns

As an example application of this analysis, we consider here the network shown in Figure 7: links 1, 2 and 3 represent freeway sections, links 7 and 8 represent arterial streets, and links 5, 4 and 6 represent entrance/exit ramps to/from the freeway. Each link a (a=1,2,...,8) has capacity m_a; node o is the origin, and nodes c, d, e and f are destinations. We assume that queuing pattern and OD flow rate (measured at the origin) for each OD pair (o, d), $\dot{Q}_d(s)$, are given for all s during the study period. We will analyse the effect of entry ramp metering, which in this case is represented by modifying the capacity of link 5.

Fig. 7. A simple ladder network

We begin with enumerating all the possible queuing patterns in this network, and then examine whether or not the paradox occurs in link 5 for each of the queuing patterns. Note that we do not have to consider all combinations of queuing states (*ie* 2^8=256 queuing patterns). First, we can exclude the queuing patterns in which link 5 is not saturated because we have already observed that the paradox never occurs if the link is not saturated. Second, we assume link 1 to be saturated because the queuing state of link 1 is independent of the occurrence of the paradox in link 5. Thus, we analyse the possible 64 (=2^6) queuing patterns listed in Table 1, where queuing state in each links is denoted by 1 (saturated) or 0 (non-saturated), and each queuing pattern is designated by one of a consecutive numbers from 0 to 63. [For example, if links 2 and 6 (as well as links 1 and 5) are saturated and the remaining links 3,4,7 and 8 are not saturated, the queuing pattern is coded as "001001" and the pattern number is "9" (=1+2^3).]

Table 1: Queuing patterns for examining the paradox in link 5

4.2 Classifying Queuing Patterns – Effectiveness of Ramp metering

The DUE solution for each of the queuing patterns can be easily obtained by applying the theory developed in Chapters 1 and 2. The detailed solutions are described in Akamatsu (1999). By inspecting the solution, we can classify each queuing pattern into one of four cases according as whether the metering on link 5

(1) always increase total travel time, or

(2) has no influence on total travel time, or

(3) can reduce total travel time depending on **M** and **Q**(s), or

(4) always reduces total travel time regardless of **M** and **Q**(s).

The resulting classification is shown in the "Class." column of **Table 1**, and the number of queueing patterns for each classification is summarised in **Table 2** Somewhat surprisingly, this reveals that the ramp metering is likely to have positive effect in considerably many states: the queuing patterns falling into case (4) account for slightly over half (33) out of the possible (64) patterns, whilst those falling into case (1) are only 3 in number.

Table 2: Classification of queuing patterns

4.3 Characterisation of Queuing Patterns

We shall make some remarks on the features of each case classified in **4.2**. For case (1), the classification reveals the following facts. First, the ramp metering for this case does not reduce travel time for any of the OD pairs. Second, this case arises when links 4 and 7 are both non-saturated as shown in **Figure 8(a)**. The reason why the metering in this case increase travel time can be seen by considering the reduced networks for these queuing patterns depicted in **Figure 8(b)**: nodes *a* and *b* are connected by links 2 and 5 in the reduced networks; this means that the capacity of link 5 cannot affect the travel time for destinations *d* and *e*, and that introduction of metering on link 5 necessarily decreases total capacity between nodes *a* and *b*, which can only lead to an increase in travel time for destinations *c* and *f*.

Fig.8(a). Queuing patterns where metering on link 5 never reduces congestion

Fig.8(b). The reduced networks for the queueing patterns in Fig.8(a)

For case (2), four queuing patterns (3, 11, 35 and 43) shown in **Figure 9(a)** are similar to those of case (1) in that links 4 and 7 are both non-saturated. The essential difference between the four patterns and those in case (1) is seen from the reduced networks in **Figure 9(b)**: nodes *b* and *c* are unified in the former queuing patterns (9, 33 and 41, whilst nodes *b* and *c* are not unified in the latter patterns (3, 11, 35 and 43). Recall that the value of t_i^* in a reduced network without loops can be determined by proceeding backwards from pure destinations towards an origin. This implies that

$\dot{t}_c$ in these four queuing patterns is independent of the capacity of link 5. Therefore, metering on link 5 cannot affect the travel time for destination *c* as well as other all destinations in these four queuing patterns, while the metering increases the travel time for destination *c* in case (1).

Fig. 9(a). Queuing patterns where metering on link 5 does not influence the total travel time

Fig.9(b). The reduced networks for the queueing patterns in Fig.9(a)

The remaining 15 queuing patterns in case (2) share certain similarities in that link 5 is deleted in the reduced networks. As we have seen in section **2.2**, paradox occurrence is independent of the capacity of any links that are deleted to form a reduced network. It follows from this fact that metering on link 5 can have no influence on equilibrium travel time for these queuing patterns.

For case (3), we note that link 2 is saturated and link 3 is non-saturated in all the queueing patterns (**Figure 10** shows this fact in the reduced networks, where nodes *b* and *c* are unified into a single node while node *a* is not unified with *b*). Since this feature also applies to the queuing patterns that fall into case (1), we see that metering of link 5 increases the travel time for destination *c*. However, unlike the queuing patterns of case (1), the travel time for other destinations in case (3) can be reduced by the metering. For example, consider the queuing pattern 13 whose reduced network is shown at the top of **Figure 10**. Equating inflows and outflows at each node of the reduced network, we see that equilibrium travel time for destinations *c*, *d*, *e*, and *f* is governed by

$$\dot{t}_b = \dot{t}_c, \quad \dot{t}_d = \dot{t}_e = \dot{t}_f, \quad \begin{cases} (m_4 + m_6)\,\dot{t}_d = \dot{Q} + m_5 \dot{t}_b \\ (m_2 + m_5)\,\dot{t}_b = m_6 \dot{t}_d \end{cases},$$

where $\dot{Q} \equiv \dot{q}_c + \dot{q}_d + \dot{q}_e + \dot{q}_f$. Simple calculation yields the following solution:

$$\dot{t}_d = \dot{t}_e = \dot{t}_f = \left[m_4 + m_6 - \frac{m_5}{m_2 + m_5} m_6 \right]^{-1} \dot{Q}$$

We immediately see from this solution that decrease in m_5 always reduces the value of $\dot{t}_d = \dot{t}_e = \dot{t}_f$. Therefore, the metering indeed reduces travel time for each destination except *c*, but always increases it for *c*; in other words, not every OD pair enjoys the reduction in travel time. This is the reason why we cannot definitely conclude whether the ramp metering reduces *total* travel time for queuing patterns in case (3); it depends on the particular values of $\dot{\mathbf{q}}(t)$ and $\mathbf{M}$ that obtain.

Fig. 10. The reduced networks for queueing patterns that fall into case (3)

For case (4), we note the following facts, First, the ramp metering in this network always lead to Pareto improvement for the OD pairs (*ie* the travel time for each OD pair is no worse after introduction of metering). Second, this case arises in all 21 queueing patterns in which link 2 is non-saturated as well as in 12 others. The reason why the metering when link 2 is non-saturated always reduces travel time can be explained as follows: metering on link 5 necessarily increases the travel time on path (a, d, e, b) provided the demand for this path remains fixed; on the other hand, in the DUE state, $\dot{t}_b$ should remain unchanged regardless of the metering on link 5 because $\dot{c}_2 = 0$ if link 2 remains non-saturated; together, these imply a decrease in flows on this path due to the metering, and this always reduces the travel time on links 4 and 7. This then causes a reduction in the travel time to nodes *d* and *e*, which cannot have negative effect to the travel time for nodes *c* and *f*. Thus, we see that the ramp metering in this case always leads to a Pareto improvement.

This mechanism can also be understood from the reduced networks. **Figure 11(a)** shows the reduced networks for queuing patterns 6, 14, 22 and 30, which are typical examples of the queueing patterns classified into case (4) in which link 2 is not saturated. Note here that, in all these networks, controlled link 5 forms a part of a particular type of loops that include all destinations and node *a*; this is almost the same as saying that *link 5 leaves destinations and arrive at node a* (*ie* flow on link 5 proceeds "*backward*" from destinations to an origin), which leads to emergence of the loop(s) because each destination should have entering links (*ie* the links with the direction opposite to link 5). For example, in queueing pattern 14, link 5 emanates from destinations *d*, *e* and *f*, and enters node *a*; this "*backward link*" gives rise to two loops (4, 5) and (3, 6, 5) that cover all the destinations and node *a*. This example also clearly shows the reason why this type of loops implies the effectiveness of the ramp metering on link 5. Considering flow conservation at a unified node (*d*,*e*,*f*) of the reduced network, we see that $\dot{t}_d = \dot{t}_e = \dot{t}_f$ decreases with the decrease in sum of outflows, X_{OUT}, from node (*d*,*e*,*f*), because sum of inflows, X_{IN}, to the node (which is identical to X_{OUT}) is determined as $(m_4 + m_6)\dot{t}_d$ at equilibrium. On the other hand, the fact that *link 5 emanates from destinations and enters node a* implies that X_{OUT} is decreased by the metering on link 5 (*ie* a decrease in m_5), because X_{OUT} is given by $(\dot{q}_d + \dot{q}_e + \dot{q}_f) + m_5\dot{t}_b$ and, $\dot{t}_a = \dot{t}_b$ is constant regardless of the metering for this queuing pattern. Hence, metering on link 5 always leads to the decrease in $\dot{t}_d = \dot{t}_e = \dot{t}_f$. Similarly, flow conservation at node *c* (*ie*. $m_3\dot{t}_c = \dot{q}_c + m_6\dot{t}_d$) reveals that $\dot{t}_c$ decreases with the decrease in $\dot{t}_d$, which is accomplished by the metering on link 5. These observations have shown that travel time for every destination is reduced by the ramp metering when this type of loop emerges.

Fig. 11 (a). The reduced networks for typical queueing patterns that fall into case (4) in which link 2 is not saturated.

In the remaining 12 queuing patterns in case (4), link 2 is saturated. The reason why the metering is always effective for these patterns (even if link 2 is saturated) can be understood by comparing the reduced networks for these patterns and those for case (3). Figure 11(b) shows the typical examples: we see that slight modifications of patterns 7, 23, 47 and 63 in case (4) would yields patterns 13, 29, 45 and 61 in case (3), respectively; the difference is that destination *c* is unified into node *b* in the latter (case (3)) whilst it is not unified (or unified into other destinations) in the former (case (4)). As we have shown in the remarks for case (3), the metering on link 5 increases travel time for destination *c* for the queuing patterns in case (3). By the same argument, travel time for node *b* (that is just a traversal node) in the queuing patterns of case (4) would be worsened by the metering. But, unlike the queuing patterns of case (3), destination *c* is separated from node *b* in the queuing patterns of case (4), which prevents the metering from increasing the travel time for destination *c*.

Fig.11(b). The reduced networks for typical queueing patterns fallen into case (4) in which link 2 is saturated.

4.4 Example Network and Possible Queuing Patterns

We now consider briefly the ramp-metering in a variant of the simple ladder network example that is shown in Figure 7 and analysed above. In this case, we consider the case where the direction of link 5 is reversed. Here, links 1, 2 and 3 could be taken to represent freeway sections with links 7 and 8 representing arterial streets, and links 4, 5, and 6 representing exit ramps from the freeway. Conversely, links 7 and 8 could be taken to represent freeway sections with links 1, 2 and 3 representing arterial streets, and links 4, 5, and 6 representing entrance ramps from to freeway. In either case, travellers to node f and beyond would be presented with a choice between the ramps represented by links 4, 5 and 6.

As before, we can classify each queuing pattern in this modified network into one of four cases according to the effect that metering on link 5 has on the total travel time. The resulting classification corresponds exactly to that in Table 2 except that classes 1 and 4 are interchanged. Thus the use of metering in this modified example most often leads to an increase in cost (33 cases out of 64). Furthermore, in all cases where link 2 is uncongested, metering on link 5 either always increases travel times (case 1) or at best has no influence on them (case 2). The same 9 queueing patterns can give rise to an improvement and so fall into case 3 as with the original network. That leaves just 3 queueing patterns (9, 33 and 41) for which metering on link 5 always results in a decrease in travel time (case 4).

The indeterminate case (3) in which metering on link 5 can reduce travel times, depending on demand, can be characterised as follows. In all of these queueing patterns, link 2 has a queue whilst link 3 does not, and two or more of links 4, 6, 7, and 8 has a queue. The two queueing patterns that conform to this specification that do not fall into case (3) are queueing pattern number 21, in which only links 2, 4 and 7 have queues and falls into case (2) (no influence), and queueing pattern number 41, in which only links 2, 6 and 8 have queues and falls into case (4) (always improves travel time).

Inspection of the three queueing patterns (9, 33 and 41) in case (4) for this network is illuminating. These are exactly the queueing patterns in which queues are present on both of links 2 and 5, and on at least one of links 6 and 8. In each of these cases, reducing the capacity of link 5 will cause traffic to transfer to the uncongested alternative route to node *e* that uses links 4 and 7, which provide unchanged travel time to node *e*: this will reduce the travel time on link 2 to the benefit of travellers to other destinations. We note with interest that the queueing pattern number 33 in this network corresponds closely to the Braess-like example shown in Figure 6 above, where in this network links 3 and 6 together correspond to link (2, 4) of the Braess example, whilst links 4 and 7 together correspond to link (1, 3).

5. CONCLUDING REMARKS

This paper has extended the theory presented in Akamatsu and Heydecker (2003) on the analysis and detection of "capacity paradoxes" under dynamic user equilibrium (DUE) assignment in networks that have a one-to-many OD pattern. The original theory was applicable directly only to saturated networks in which every link that carries any flow is overloaded. In the present paper, we have shown how networks in a broader class in which there are queues on some links but not on others can be reduced to an equivalent problem in a corresponding saturated network. This technique has been illustrated by constructing a "reduced network", which consists of only saturated (queuing) links of an original network. This technique implies that essential properties of a DUE flow pattern in a non-saturated network are fully expressed in the topological structure of the reduced network (*ie* spatial queuing pattern), and that we can detect the occurrence of capacity paradoxes using information from this alone.

As an application of this theory, we examined the effectiveness of a ramp metering/closing operation in two variants of a small ladder network. This application provides some interesting insights. First, it seems likely that metering or closing a freeway entrance ramp can be effective to reduce travel times not only on the freeway itself but also in the whole network including arterial streets. Second, effective metering operations can be implemented without information on detailed OD demands: this is because in many cases, the queuing pattern on the links provides sufficient information to judge the effectiveness of metering operations.

We note that these deductions depend on certain features of the analysis. First, we assumed flow patterns in a network to be a DUE state. An important topic for future research is to examine whether or not the implication obtained here holds in other route choice principles such as the reactive user optimal assignment analysed in Kuwahara and Akamatsu (2001). Second, the theory in this paper applies directly only for intervals throughout which the spatial queuing pattern (*ie* the set of queuing links) remains unchanged. For practical applications, we should take into account the evolution of spatial queuing patterns, which can be achieved by considering a series of intervals of the required kind. One of the simplest metering strategies considering this point would be to change the metering operation according to the current queuing pattern. Of course, such strategy is just a "myopic" control (with respect to the time horizon), and hence cannot be guaranteed to achieve global (*ex post fact*) optimality for the whole of a study period. The strategy, however, would be both practical and effective, and would be optimal among strategies *that do not require information on future OD demand patterns*. Exploring more detailed properties of such control strategy would be an interesting research topic.

Acknowledgement - The authors would like to thank Eiko Takahashi and Toshikazu Hayazaki for their research assistance on the calculations of several examples in this paper.

REFERENCES

T. AKAMATSU, "Some Methods for Detecting Capacity Paradoxes in Dynamic Traffic Assignment", *Department of Knowledge-based Information Engineering Research Report, TUTKIE-9901, Toyohashi University of Technology*, *Toyohashi, Aichi, Japan*, 1999.

T. AKAMATSU, "A Dynamic Traffic Assignment Paradox", *Transportation Research* **34B**, 515-531, 2000.

T. AKAMATSU, and B.G. HEYDECKER, "Detecting Dynamic Traffic Assignment Capacity Paradox in Saturated Networks", *Transportation Science (in press)*, 2003.

D. BRAESS, "Uber ein Paradoxen der Verkehrsplaning," *Unternehmensforschung* **12,** 258-268, 1968.

M. KUWAHARA, and T. AKAMATSU, "Dynamic User Optimal Assignment with Physical Queues for a Many-to-Many OD Pattern," *Transportation Research* **35B**, 461-479, 2001.

APPENDIX A

We will prove that the solution $\dot{t}(s)$ obtained by (13), (14) satisfies (5) and (6) for any patterns of $\mathbf{y}_F$. By the definition of the matrix **R**, we have the following identities: (a) $\dot{t}(s) = \mathbf{R}^T \dot{t}_R(s)$, ($b$) $\mathbf{R}\dot{\mathbf{Q}}(s) = \dot{\mathbf{Q}}_R(s)$, ($c$) $\mathbf{R}\mathbf{A}_Q = \mathbf{A}_R$, ($d$) $\mathbf{R}\mathbf{A}_{Q-} = \mathbf{A}_{R-}$, ($e$) $\mathbf{R}\mathbf{A}_F = \mathbf{0}$. From ($a$) and ($e$), $\dot{t}(s)$ obtained by (14) always satisfies (5). Note here that $[\mathbf{R}^T\mathbf{R}]$ and $[\mathbf{R}\,\mathbf{R}^T]$ are invertible. Hence, it follows from the identities that

$$\dot{t}_R(s) = [\mathbf{R}\,\mathbf{R}^T]^{-1}\mathbf{R}\,\dot{t}(s) \text{ ,} \tag{A1a}$$

$$\dot{\mathbf{Q}}(s) = [\mathbf{R}^T\mathbf{R}]^{-1}\mathbf{R}^T\,\dot{\mathbf{Q}}_R(s) \text{ ,} \tag{A1b}$$

$$\mathbf{A}_Q = [\mathbf{R}^T\mathbf{R}]^{-1}\mathbf{R}^T\mathbf{A}_R \text{,} \tag{A1c}$$

$$\mathbf{A}_{Q-}^T = \mathbf{A}_{R-}^T\mathbf{R}\,[\mathbf{R}^T\mathbf{R}]^{-1}. \tag{A1d}$$

Since $[\mathbf{R}^T\mathbf{R}]$ is a one-to-one mapping from N dimensional Euclid space to itself, the equation (6) is equivalent to the following equation:

$$[\mathbf{R}^T\mathbf{R}]\{\mathbf{A}_Q\mathbf{M}_Q\mathbf{A}_{Q-}^T\;\dot{t}(s) - \mathbf{A}_F\mathbf{y}_F\} = [\mathbf{R}^T\mathbf{R}]\,\dot{\mathbf{Q}}(s). \tag{A2}$$

From the identity (e), this reduces to

$$[\mathbf{R}^T\mathbf{R}]\{\mathbf{A}_Q\mathbf{M}_Q\mathbf{A}_{Q-}^T\;\dot{t}(s)\} = [\mathbf{R}^T\mathbf{R}]\,\dot{\mathbf{Q}}(s). \tag{A3}$$

Thus, it is enough for us to prove that (A3) holds if $\dot{t}(s)$ satisfies (13) and (14). Substitution of (A1c), (14) and the identity (d) into the L.H.S. of (A3) yields

$$\begin{aligned}[\mathbf{R}^T\mathbf{R}]\{\mathbf{A}_Q\mathbf{M}_Q\mathbf{A}_{Q-}^T\;\dot{t}(s)\} &= [\mathbf{R}^T\mathbf{R}]\{[\mathbf{R}^T\mathbf{R}]^{-1}\mathbf{R}^T\mathbf{A}_R\}\mathbf{M}_Q\mathbf{A}_{Q-}^T\,\{\mathbf{R}^T\;\dot{t}_R(s)\} \\ &= \mathbf{R}^T\mathbf{A}_R\mathbf{M}_R\mathbf{A}_{R-}^T\;\dot{t}_R(s).\end{aligned} \tag{A4}$$

For the $\dot{t}_R(s)$ satisfying (13), this further reduces to

$$[\mathbf{R}^T\mathbf{R}]\,\mathbf{A}_Q\mathbf{M}_Q\mathbf{A}_{Q-}^T\;\dot{t}(s) = \mathbf{R}^T\dot{\mathbf{Q}}_R(s). \tag{A5}$$

From (A1b), we see that $\mathbf{R}^T\dot{\mathbf{Q}}_R(s)$ is equal to $[\mathbf{R}^T\mathbf{R}]\,\dot{\mathbf{Q}}(s)$, which is precisely the same as the R.H.S of (A3). **QED.**

APPENDIX B

The final variables in the equations should be $\mathbf{Q}(s)$ rather than $\dot{\mathbf{Q}}(s)$ because the final variable $\mathbf{Q}(s)$ corresponds to a part of the bracket term (the *integral* term) of (20), $\int_0^s \{\partial \mathrm{t}(t)/\partial m_a\}dt$. To see this fact in a simpler form, let us consider the definitional equation of total travel time in the network during time period $[0,T]$:

$$C = \int_0^T \dot{\mathbf{Q}}(s)' \,\mathrm{t}(s)\,ds = \int_0^T \dot{\mathbf{Q}}(s)' \left\{ \mathrm{t}(0) + \int_0^s \dot{\mathrm{t}}(t)dt \right\} ds \tag{B1}$$

where superscript ' denotes transpose of vectors/ matrices. This can be represented by the variables defined in "*time t* (*instantaneous*) *reduced network*", which is constructed at *each instant t* in [*0*,*T*] by the same manner as "reduced networks" described in section **1.4**:

$$\begin{aligned} C &= \int_0^T \dot{\mathbf{Q}}(s)' \left\{ \mathrm{t}(0) + \int_0^s \mathbf{R}(t)' \,\dot{\mathrm{t}}_R(t)dt \right\} ds \\ &= \mathrm{t}(0)' \mathbf{Q}(T) + \int_0^T \dot{\mathbf{Q}}(s)' \left\{ \int_0^s \mathbf{R}(t)' \mathbf{V}_R^{-1}(t)\, \dot{\mathbf{Q}}_R(t)dt \right\} ds \end{aligned} \tag{B2}$$

where $\mathbf{V}_R(t) \equiv \mathbf{A}_R(t)\mathbf{M}_R(t)\mathbf{A}'_{R-}(t)$, and $\mathbf{A}_R(t)$ is an incidence matrix that represents the structure of the time *t* reduced network. Note that we can regard the first term of eq.(B2) a given constant in so far as we only aim to analyse the dynamic capacity paradox, and therefore, we omit the first term in the discussion below.

Let $t(n)$ be the time when new queuing pattern n emerges in the network, where n = 0,1,2,... is the number allocated to such events sequentially in the order of the event occurrence time from time 0. Then, the total travel time experienced by users during queuing pattern n is

$$C(n) \equiv \int_{t(n)}^{t(n+1)} \dot{\mathbf{Q}}(s)' \left\{ \int_0^s \mathbf{R}(t)' \mathbf{V}_R^{-1}(t)\, \dot{\mathbf{Q}}_R(t)dt \right\} ds . \tag{B3}$$

By definition of $t(n)$, queuing pattern n does not change during $[t(n), t(n+1))$, and $\mathbf{R}(t)' \mathbf{V}_R^{-1}(t) = \mathbf{R}(t(n))' \mathbf{V}_R^{-1}(t(n)) \quad \mathit{for} \quad \forall t \in [t(n), t(n+1))$. Hence, $C(n)$ reduces to

$$\begin{aligned} C(n) &= \int_{t(n)}^{t(n+1)} \dot{\mathbf{Q}}(s)' \left\{ \mathbf{R}(t(n))' \mathbf{V}_R^{-1}(t(n)) \int_0^s \dot{\mathbf{Q}}_R(t)dt \right\} ds \\ &= \int_{t(n)}^{t(n+1)} \dot{\mathbf{Q}}(s)' \mathbf{R}(t(n))' \mathbf{V}_R^{-1}(t(n))\, \mathbf{Q}_R(s)\, ds \end{aligned} \tag{B4}$$

This paper discusses the increase/decrease of only $C(n)$ (not C) with changes in some link capacities, given queuing pattern n (*ie.* $\mathbf{R}(t(n))$ and $\mathbf{A}_R(t(n))$ are given constant matrices).

Table 1: Queuing patterns for examining the paradox in link 5

Pattern Number	Link 8	7	6	4	3	2	Class.
0	0	0	0	0	0	0	(2)
1	0	0	0	0	0	1	(2)
2	0	0	0	0	1	0	(2)
3	0	0	0	0	1	1	(2)
4	0	0	0	1	0	0	(2)
5	0	0	0	1	0	1	(2)
6	0	0	0	1	1	0	(4)
7	0	0	0	1	1	1	(4)
8	0	0	1	0	0	0	(2)
9	0	0	1	0	0	1	(1)
10	0	0	1	0	1	0	(2)
11	0	0	1	0	1	1	(2)
12	0	0	1	1	0	0	(4)
13	0	0	1	1	0	1	(3)
14	0	0	1	1	1	0	(4)
15	0	0	1	1	1	1	(4)
16	0	1	0	0	0	0	(2)
17	0	1	0	0	0	1	(2)
18	0	1	0	0	1	0	(4)
19	0	1	0	0	1	1	(4)
20	0	1	0	1	0	0	(2)
21	0	1	0	1	0	1	(2)
22	0	1	0	1	1	0	(4)
23	0	1	0	1	1	1	(4)
24	0	1	1	0	0	0	(4)
25	0	1	1	0	0	1	(3)
26	0	1	1	0	1	0	(4)
27	0	1	1	0	1	1	(4)
28	0	1	1	1	0	0	(4)
29	0	1	1	1	0	1	(3)
30	0	1	1	1	1	0	(4)
31	0	1	1	1	1	1	(4)
32	1	0	0	0	0	0	(2)
33	1	0	0	0	0	1	(1)
34	1	0	0	0	1	0	(2)
35	1	0	0	0	1	1	(2)
36	1	0	0	1	0	0	(4)
37	1	0	0	1	0	1	(3)
38	1	0	0	1	1	0	(4)
39	1	0	0	1	1	1	(4)
40	1	0	1	0	0	0	(2)
41	1	0	1	0	0	1	(1)
42	1	0	1	0	1	0	(2)
43	1	0	1	0	1	1	(2)
44	1	0	1	1	0	0	(4)
45	1	0	1	1	0	1	(3)
46	1	0	1	1	1	0	(4)
47	1	0	1	1	1	1	(4)
48	1	1	0	0	0	0	(4)
49	1	1	0	0	0	1	(3)
50	1	1	0	0	1	0	(4)
51	1	1	0	0	1	1	(4)
52	1	1	0	1	0	0	(4)
53	1	1	0	1	0	1	(3)
54	1	1	0	1	1	0	(4)
55	1	1	0	1	1	1	(4)
56	1	1	1	0	0	0	(4)
57	1	1	1	0	0	1	(3)
58	1	1	1	0	1	0	(4)
59	1	1	1	0	1	1	(4)
60	1	1	1	1	0	0	(4)
61	1	1	1	1	0	1	(3)
62	1	1	1	1	1	0	(4)
63	1	1	1	1	1	1	(4)

Table 2: Classification of queuing patterns

Classification	Number of patterns	Pattern numbers
(1) always worsen	**3**	9, 33, 41
(2) no influence	**19**	0, 1, 2, 3, 4, 5, 8, 10, 11, 16, 17, 20, 21, 32, 34, 35, 40, 42, 43
(3) can improve	**9**	13, 25, 29, 37, 45, 49, 53, 57, 61
(4) always improve	**33**	6,7,12,14,15,18,19,22,23,24,26,27,28,30,31,36,38,38,39,44,46,47,48,50,51,52,54,55,56,58,59,60,62,63

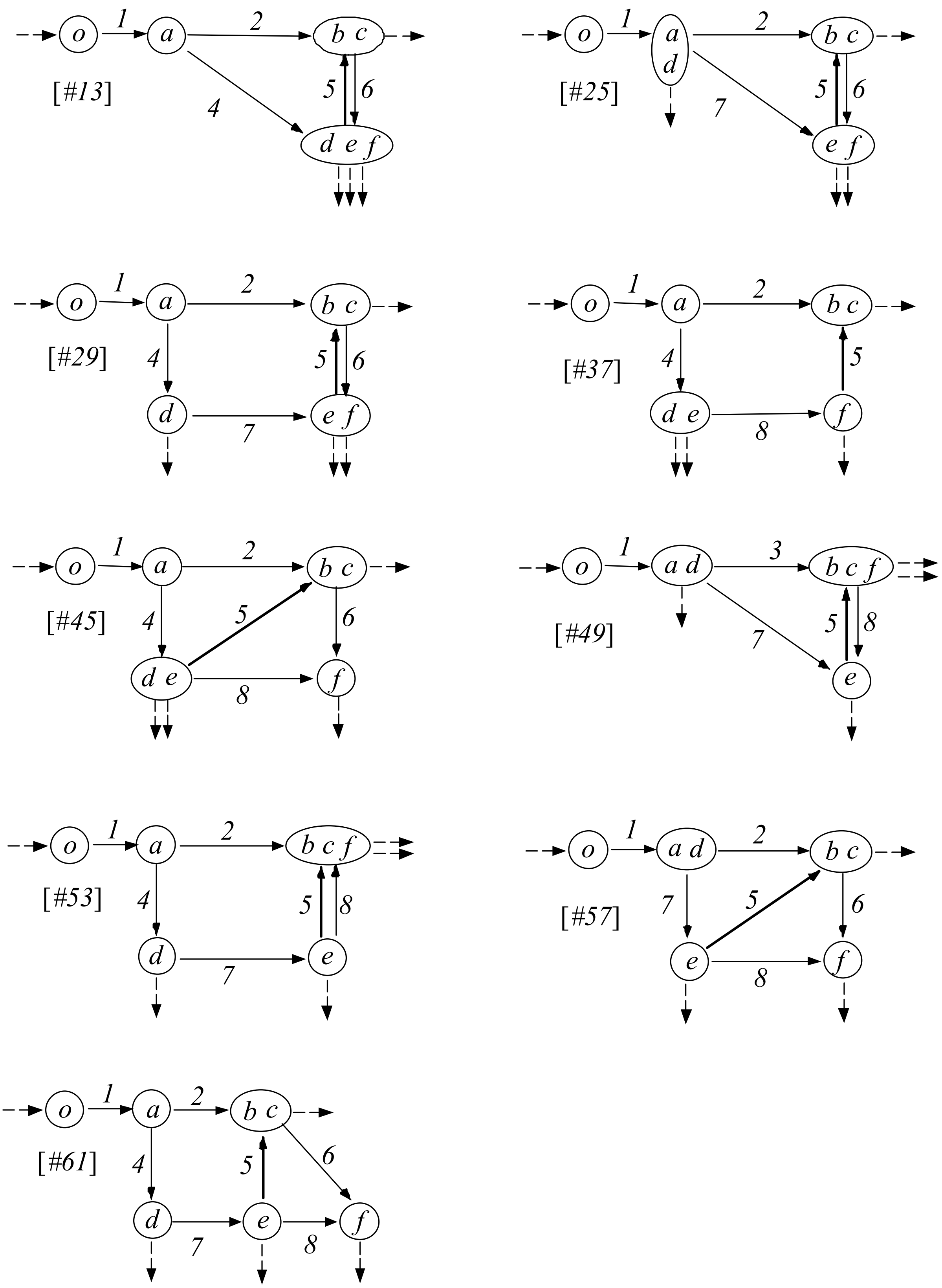
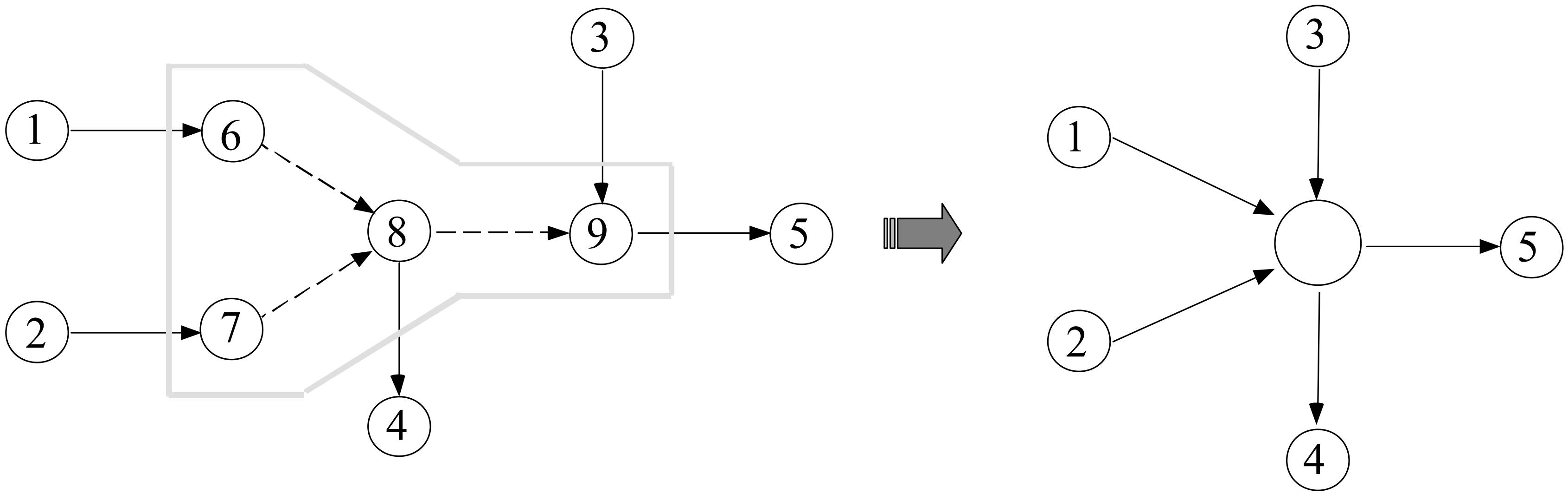


Figure 10. The reduced networks for the queueing patterns that fall into case 3.



(a) An example network

(b) The reduced network

Fig.1. Transformation of a non-saturated network to the corresponding reduced one

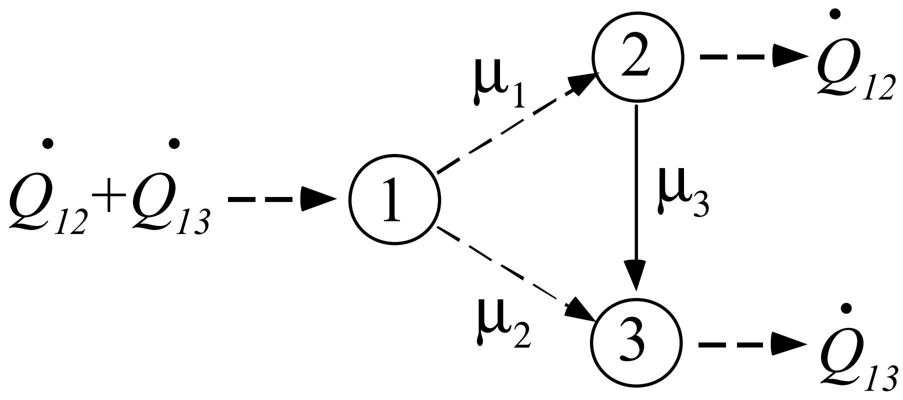


Fig. 2. An example of a queuing pattern that causes deletion of a saturated link

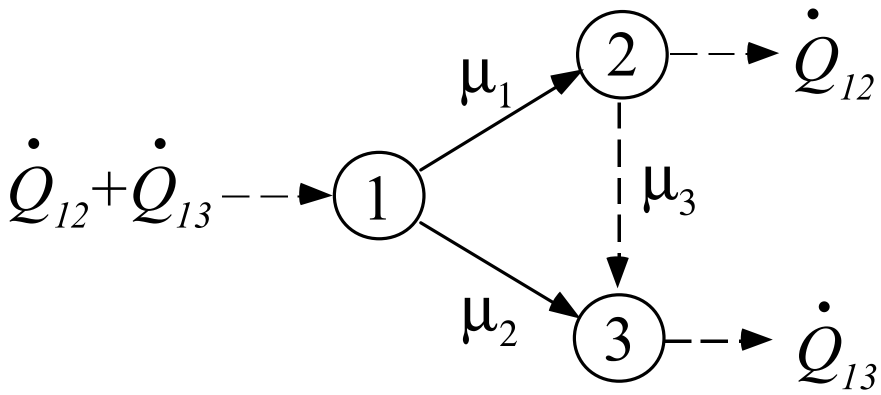


Fig. 3(a). An example network in which the matrix V is invertable

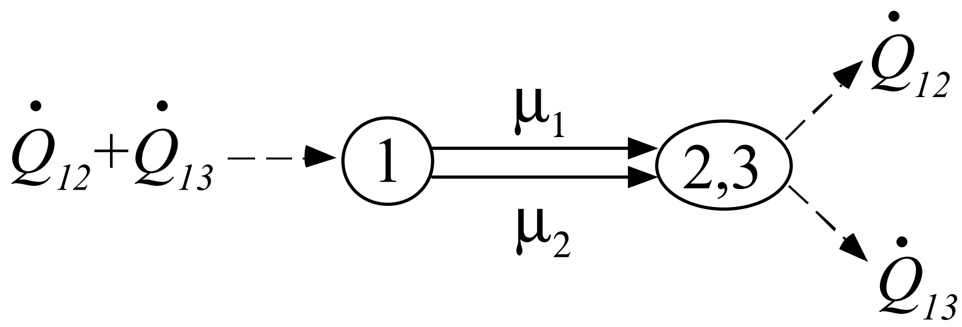


Fig. 3(b). The reduced network for the one in Fig. 3(a)

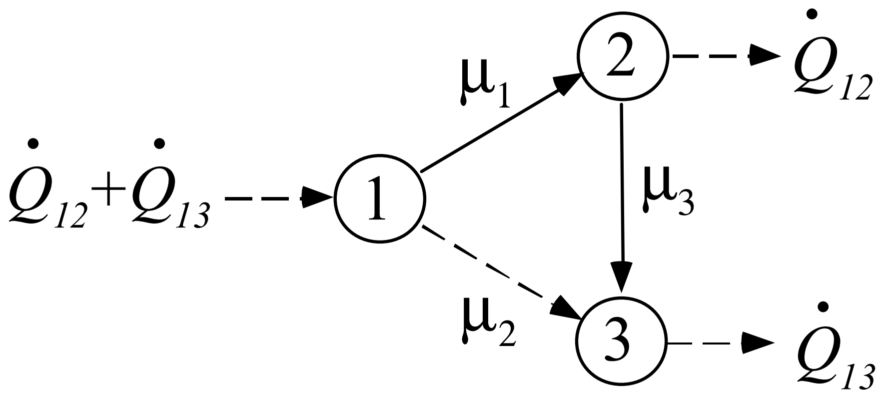


Fig. 4(a). A second example network in which the matrix V is invertable

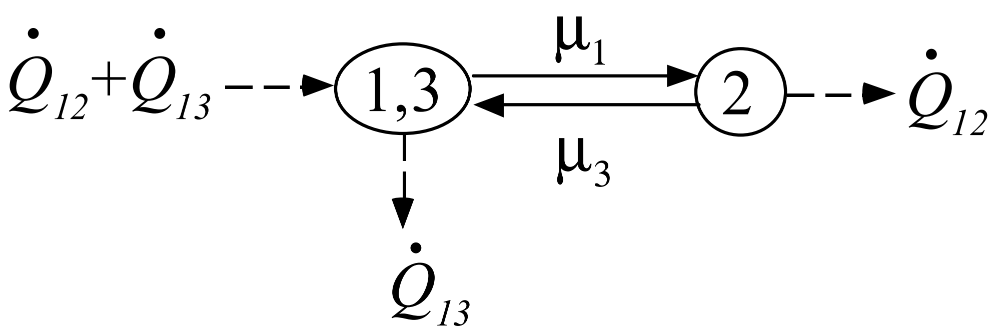


Fig. 4(b). The reduced network for the one in Fig. 4(a)

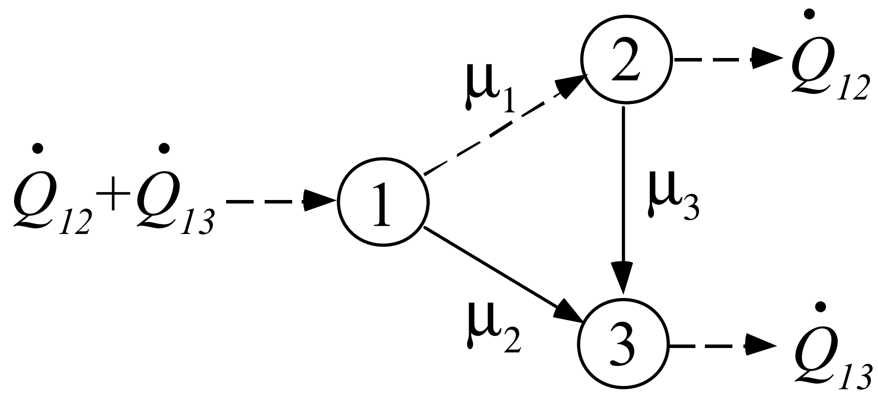


Fig. 5 (a). An example network in which the matrix V is non-invertable

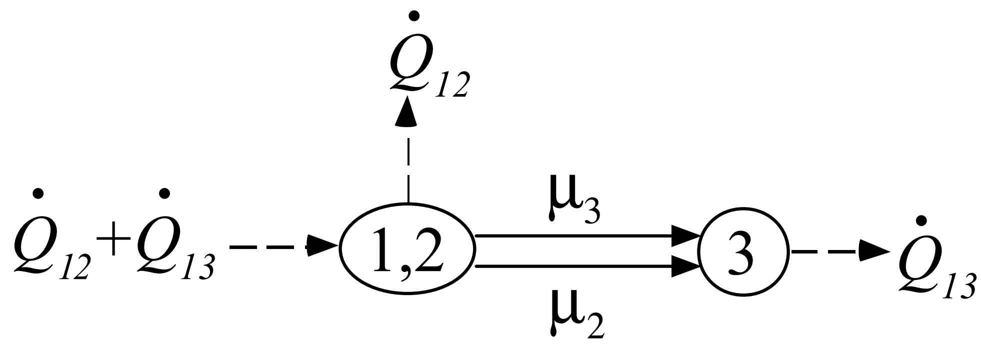


Fig. 5 (b). The reduced network for the one in Fig. 5(a)

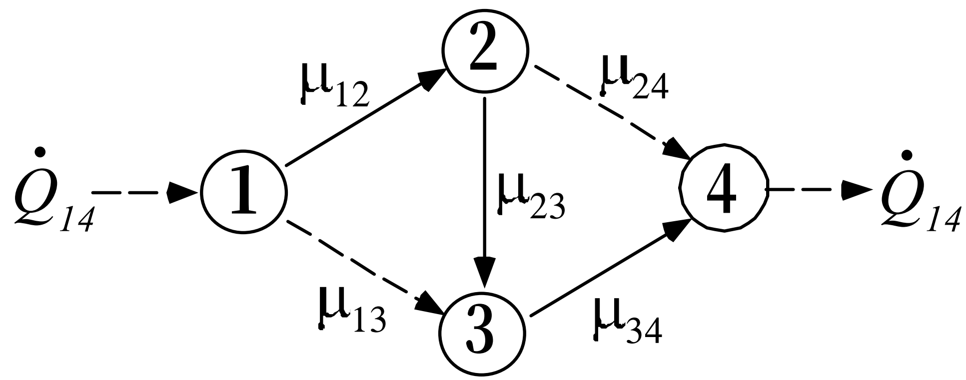


Fig. 6 (a). A second example network in which the matrix V is non-invertable

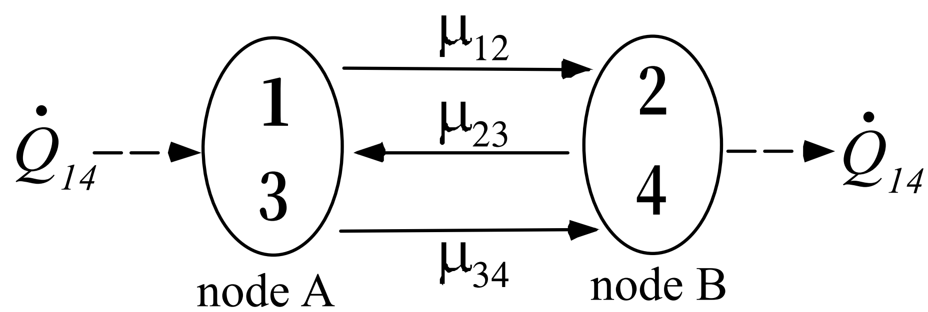


Fig. 6(b). The reduced network for the one in Fig. 6(a)

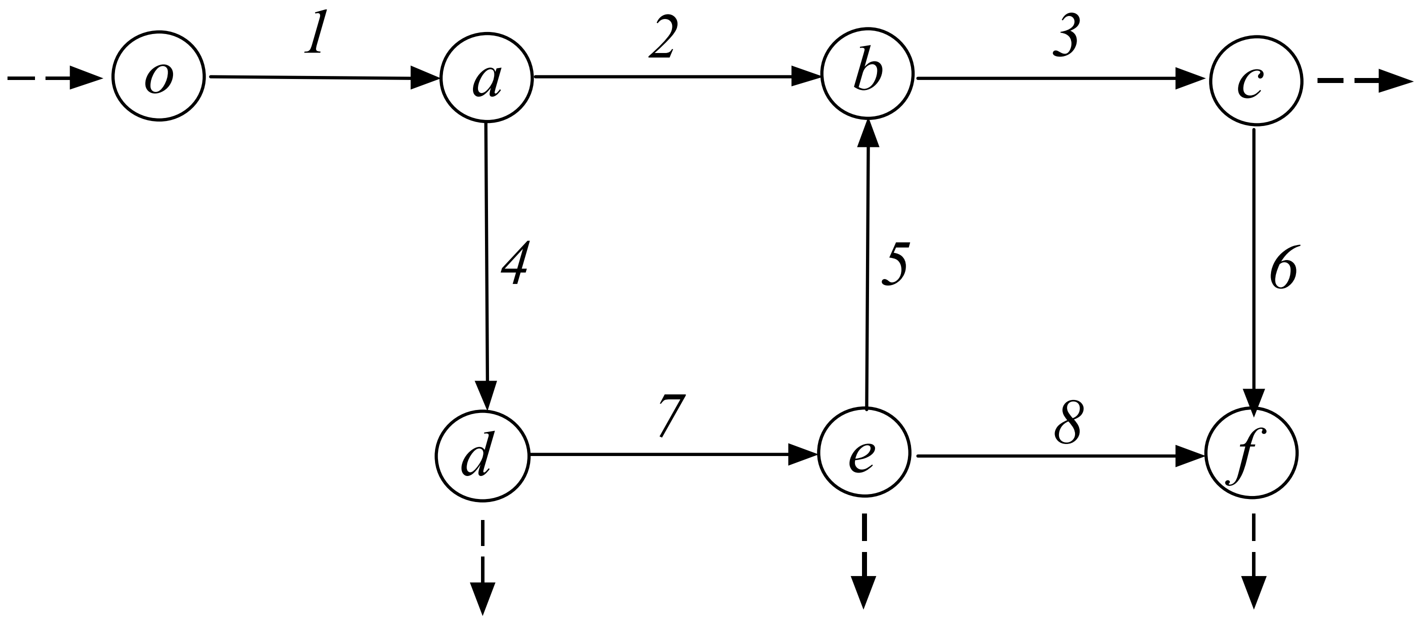


Figure 7. A simple ladder network

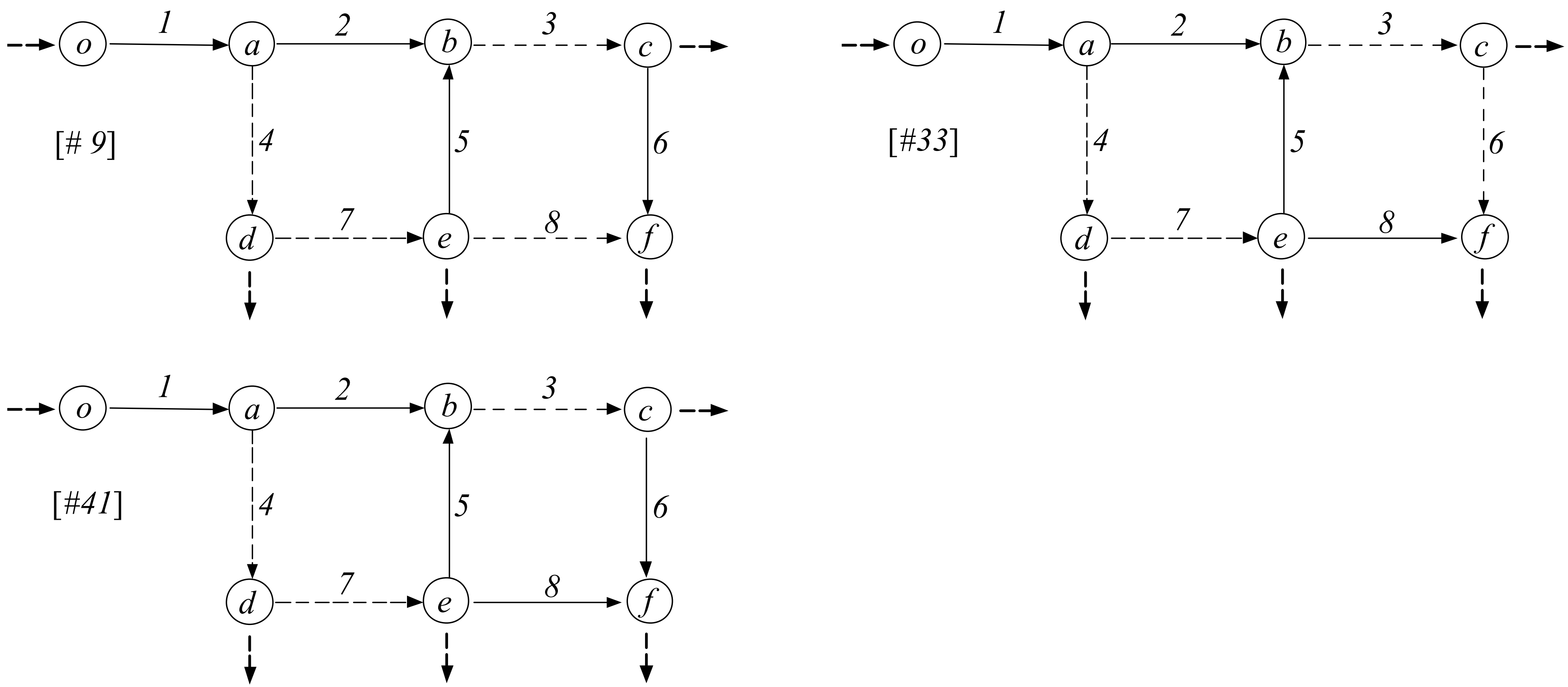


Fig. 8(a). Queueing patterns where metering on link 5 never reduces congestion.

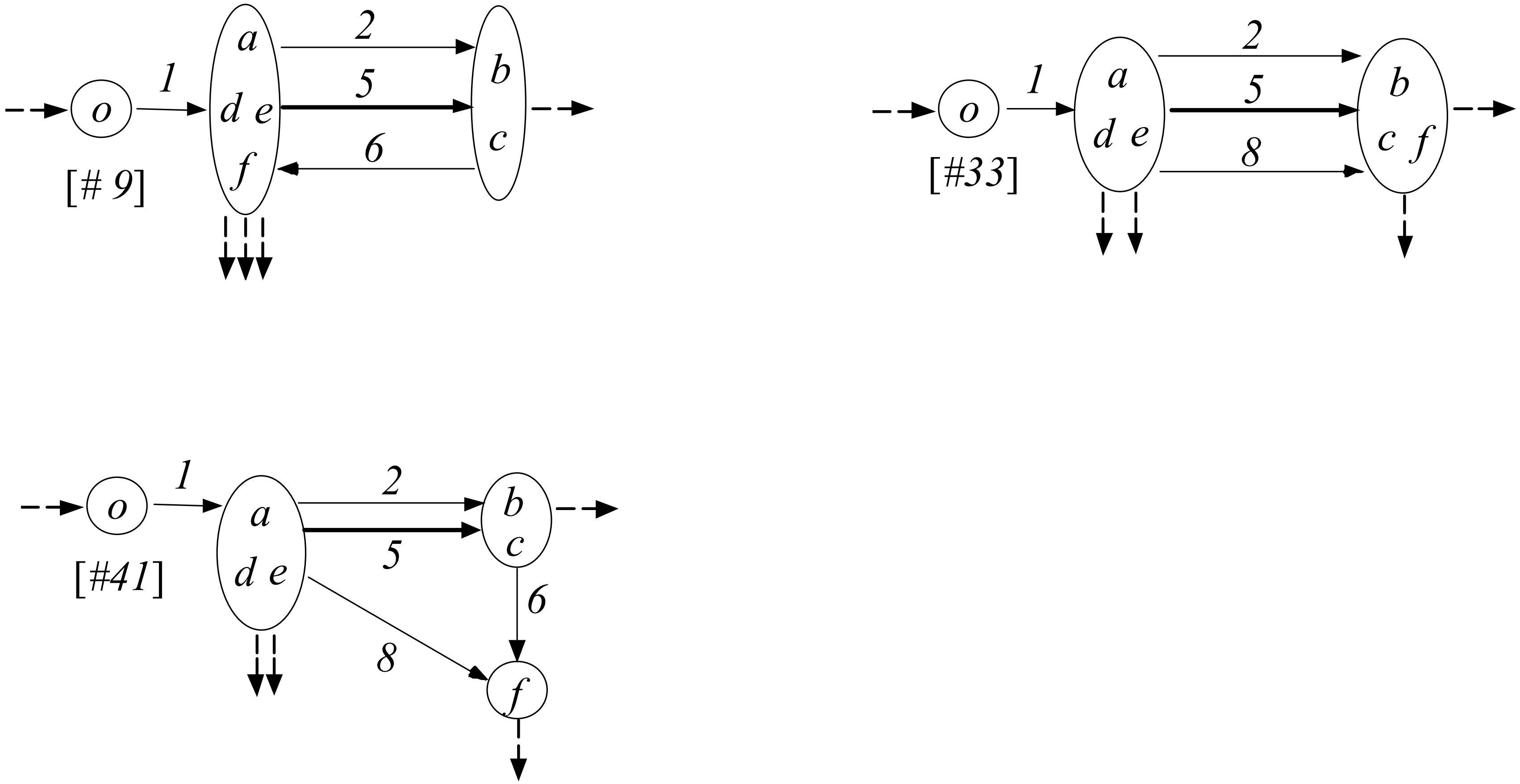


Figure 8(b). The reduced networks for the queueing patterns in Figure 8(a).

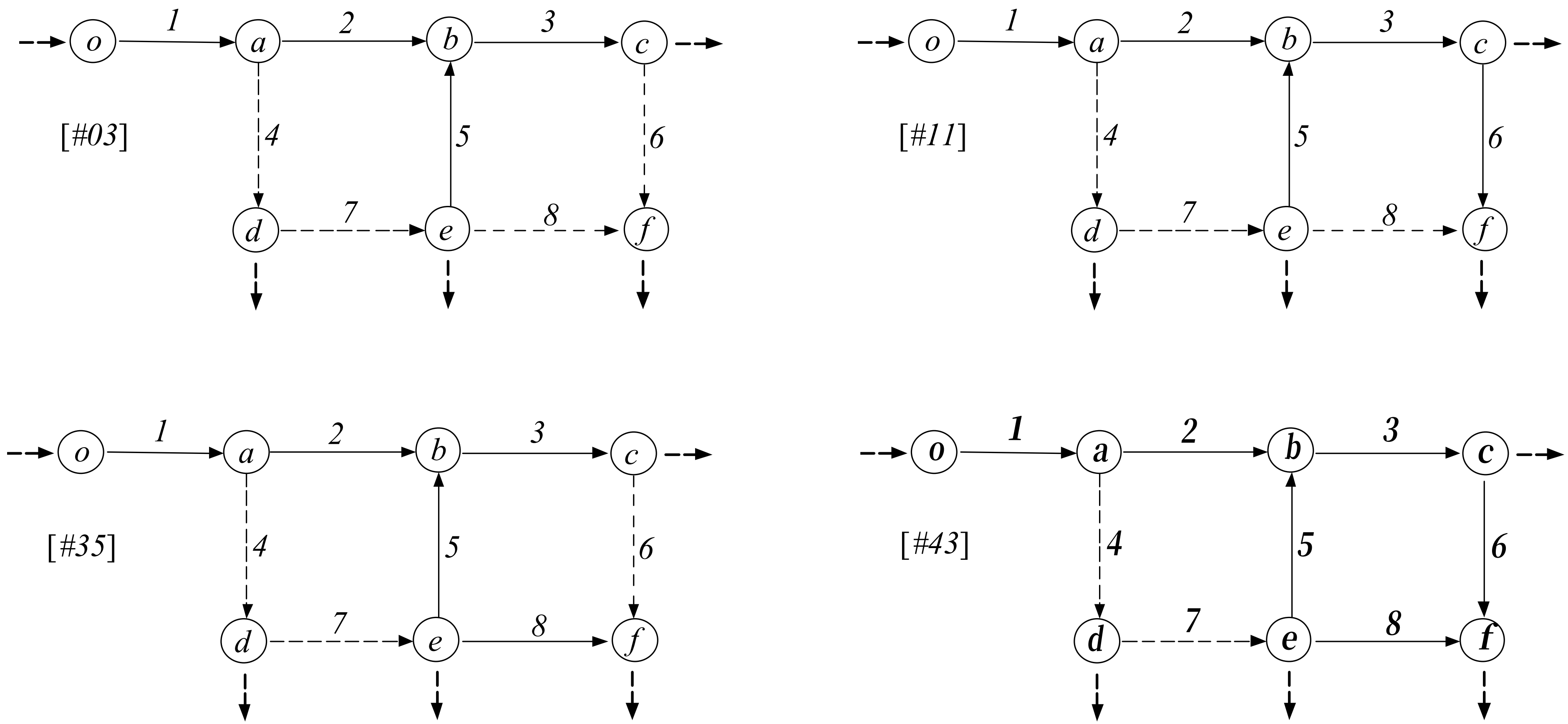


Figure 9(a). Queueing patterns where metering on link 5 dose not influence the total travel time.

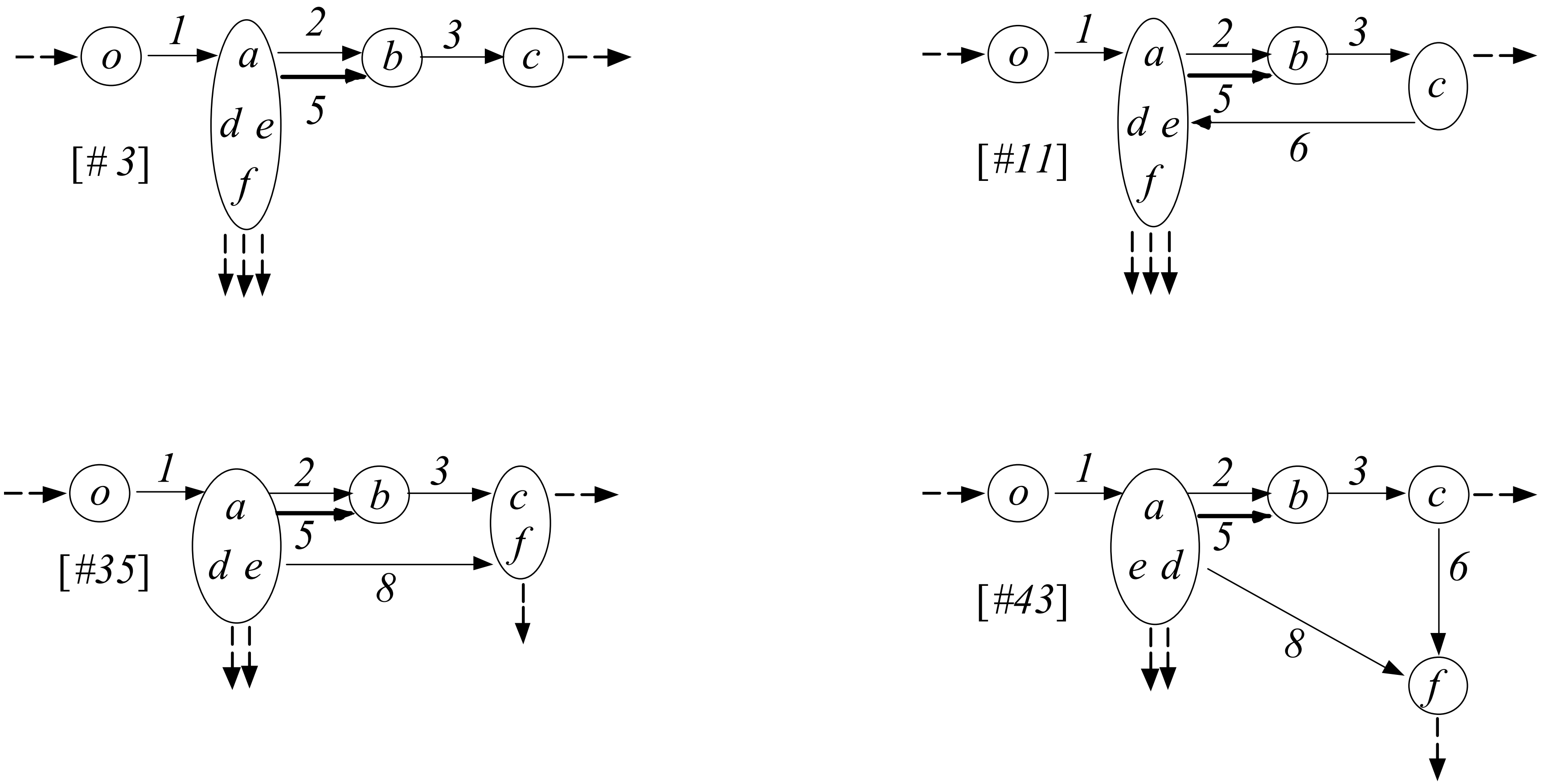


Figure 9(b). The reduced networks for the queueing patterns in Figure 9(a).

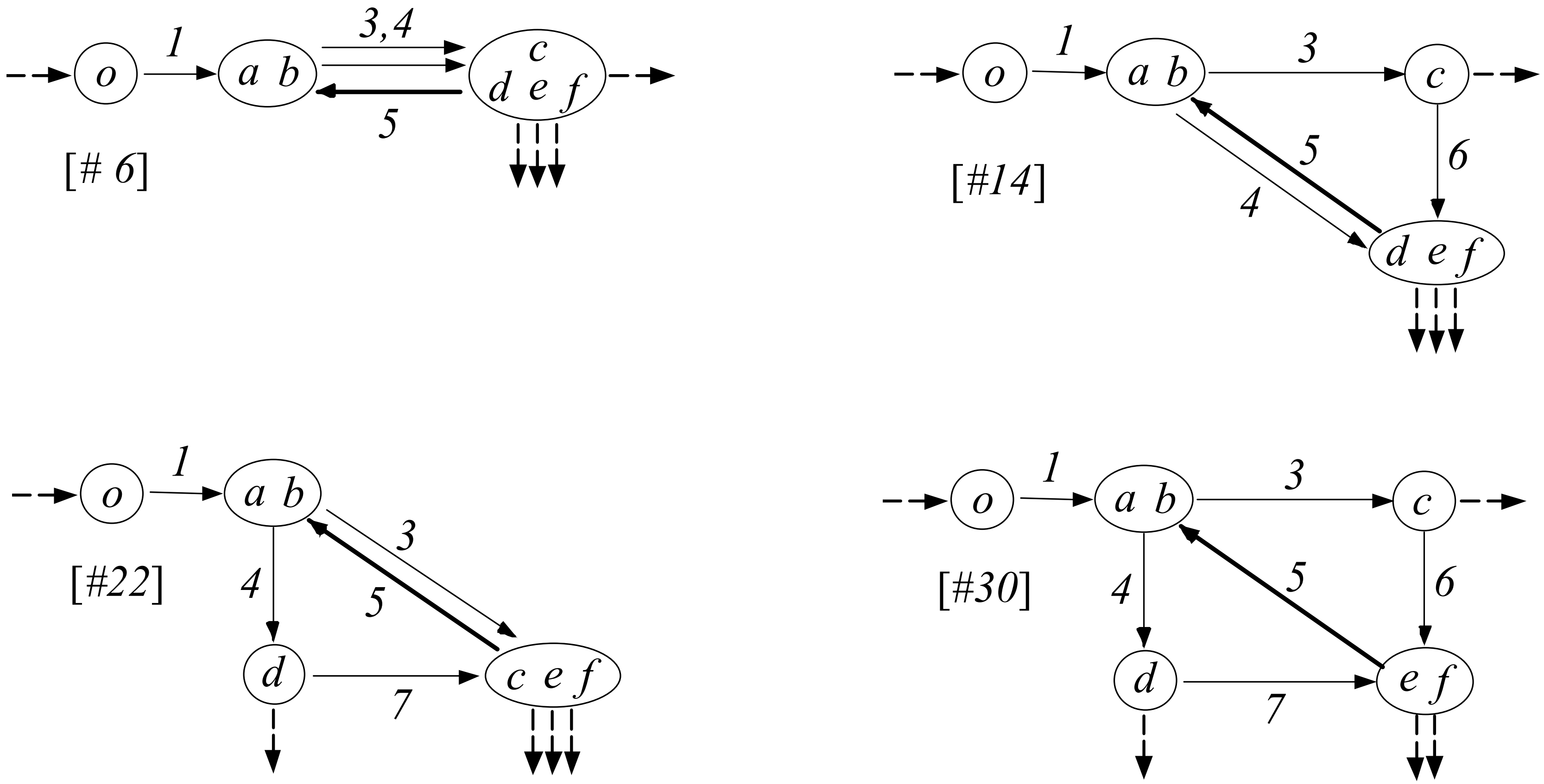


Figure 11(a). The reduced networks for typical queueing patterns that fall into case 4 in which link 2 is not saturated.

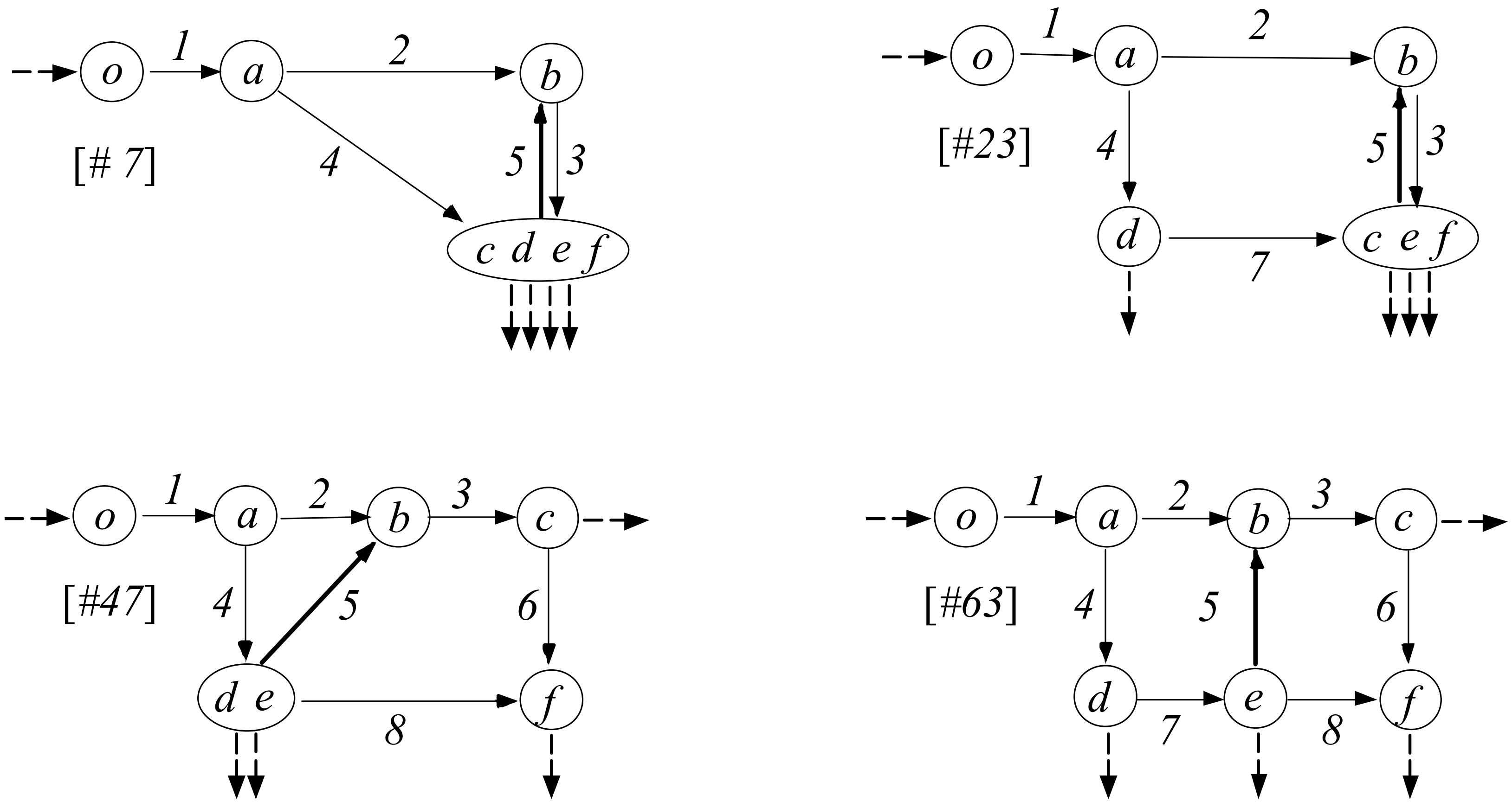


Figure 11(b). The reduced networks for the queueing patterns that fall into case 4 in which link 2 is saturated.