

The Access Control Problem on Capacitated FIFO Networks  
with Unique O-D Paths is Hard

Alan L. Erera<sup>1</sup>, Carlos F. Daganzo<sup>2</sup>, David J. Lovell<sup>3</sup>

January 24, 2001

<sup>1</sup>School of Industrial and Systems Engineering, Georgia Institute of Technology

<sup>2</sup>Department of Civil and Environmental Engineering, University of California, Berkeley

<sup>3</sup>Department of Civil and Environmental Engineering, University of Maryland

## Abstract

This paper is concerned with the performance of multi-commodity capacitated networks in a deterministic but time-dependent environment. For a given time-dependent origin-destination table, the paper asks if it is easy to find a way of regulating the input flows into the network so as to avoid queues from growing internally; i.e., to avoid capacity violations. Problems of this type are conventionally approached in the traffic/transportation field with variational methods such as control theory (if time is continuous) and with mathematical programming (if time is discrete). However, these approaches can only be expected to work well if the set of feasible solutions is convex.

Unfortunately, it is found in this paper that this is not the case. It is shown that continuous-time versions of the problem satisfying the smoothness conditions of control theory can have a finite but very large number of feasible solutions. The same happens for the discrete-time case. These difficulties arise even with the simplest versions of the problem (with unique origin-destination paths, perfect information and deterministic travel times).

The paper also shows that the continuous-time feasibility problem is NP-hard, and that if we restrict our attention to (practical) problems whose data can be described with a finite number of bits (e.g., in discrete time), then the problem is NP-complete. These results are established by showing that the problem instances of interest can be related to the Directed Hamiltonian Path problem by a polynomial transformation.

# 1 Background

This paper examines the problem of regulating access to a capacitated, multi-commodity network with time-dependent demand so as to avoid internal queues. The network is “capacitated” in the sense that some links, called bottlenecks, have a finite capacity. A queue develops at a bottleneck whenever the bottleneck flow temporarily exceeds its capacity. These capacities are exogenously specified and may be time-dependent. The goal is to serve every unit of flow with as little delay as possible, while confining all the queues to the input points (externally to the network) to prevent them from interfering with the network flows. Problems of this type arise in a variety of contexts, including telecommunication networks, air space control (airport to airport traffic), freeway networks (ramp metering), generalized polling systems, etc. Of particular interest are applications where the flows moving through the network are so large that they can be modeled by continuous variables, and where the input queues obey a FIFO (first in first out) discipline. It will be shown that even when the networks have a very simple structure with unique paths between each origin and destination (no route choice), the problem of determining whether there is a control strategy that satisfies the capacity constraints is *NP*-hard. The difficulty is caused by the combination of FIFO and time-dependence, for it disappears if either one of these conditions is relaxed.

Network access control problems have been studied extensively in the context of freeway systems with a literature that dates back to the early 60’s; see examples of early work in Wattleworth (1963) and May (1964), and a more comprehensive review in Lovell (1997). It was soon found that this was rather easy in the steady-state case for networks with unique origin-destination (O-D) paths because the problem could then be cast as a linear program (Wattleworth, 1967). With the advent of faster computing, real-time control of large time-dependent systems became a possibility, and a variety of optimization methods for real-time control were proposed to address the problem; see e.g., Yuan and Kreer (1968), Stephanedes and Kwon (1993) and Papageorgiou (1995), among many others. However, the difficulties introduced by the FIFO discipline have not been noted until recently (Lovell, 1997; Lovell and

Daganzo, 1999). These two references identified special cases that could be solved easily such as networks with unique O-D paths that contained either a single origin, a single destination or a single bottleneck. The references also noted, however, that removing the route choice element from the problem was not sufficient to eliminate the non-linearity.

It was further stated in Lovell and Daganzo (1999) that the general FIFO capacitated network access control problem without route choice could be formulated as a standard (non-linear) problem in the theory of optimal control but that the FIFO non-linearity would preclude the variational methods of control theory from being generally able to identify the global optimum. Because control theory is a commonly used tool for addressing problems of this type (freeway access problems in particular), and because the FIFO condition is not explicitly recognized in the optimal control treatment of transportation networks (dating back to the original formulation in Gazis (1974)), this paper pursues this idea further. It will be shown that even if the problem satisfies strict smoothness conditions, some instances of the problem are of a combinatorial nature for which variational methods cannot be expected to yield global optima. These combinatorial problems are shown to be “hard.” It appears thus that the best practical methodologies to solve general FIFO network access problems will be approximations with heuristic components that should exploit problem-specific features.

The problem in question consists of a network, which is characterized by finite sets of origins,  $O = \{o_i\}$ , destinations,  $E = \{e_j\}$  and bottlenecks,  $B = \{b_k\}$ , where  $I = |O|$ ,  $J = |E|$ , and  $K = |B|$ ; and a time interval  $T = [t_B, t_E]$ . For every O-D pair there is a unique (used) path which includes a subset of  $B$ . These data are summarized with 0-1 indicators,  $\gamma_{ijk}$ , that are 1 if bottleneck  $b_k$  is on the path from  $o_i$  to  $e_j$  and zero otherwise. Given for each origin-destination pair is a *cumulative arrival function*  $A_{ij} : R \rightarrow R_0^+$ , which is continuous and non-decreasing with bounded and piecewise-continuous derivatives, and set to zero (arbitrarily) at the beginning of the interval,  $A_{ij}(t_B) = 0$ . As is conventional in fluid queueing approximations (see Newell, 1982), this function denotes the number of vehicles with destination  $e_j$  that would be ready to depart  $o_i$  by time  $t$  if unrestricted. (It is assumed that all the vehicles present at  $o_i$  are

embedded in the same queue, and that the queue is FIFO.) Given for each bottleneck is a non-negative *capacity function*  $c_k : R \rightarrow R_0^+$ , which is bounded and piecewise-continuous<sup>1</sup>. Finally, given for each  $(i, k)$ <sup>2</sup> is a non-negative travel time from  $o_i$  to  $b_k$ ,  $\tau_{ik}$ , along the unique path; this is the traffic flow model of the problem.

It should be noted that practical freeway control problems are often complicated by uncertainty in the time-dependent O-D flow data, which are rarely known with much accuracy, and also by the need to ensure that freeway access is not restricted so severely that the access queues spill back onto the local streets. Although methods that take these and other complications into consideration have been proposed (see Lovell (1997), and Lovell and Daganzo (1999)), these real world concerns will be ignored here since our goal is establishing the complexity of the simplest possible formulation of the problem.

The solution for this problem is specified in terms of piecewise-differentiable control functions,  $d_i(t)$ , that give the delay imparted to the vehicles released at time  $t$  from origin  $o_i$ . [It is assumed that the set of points where the derivative of  $d_i(t)$  does not exist is countable, and that elsewhere the derivative is bounded.] The delay function is simply the difference between the departure and arrival times, indexed by departure time; i.e., if a vehicle arrives at  $o_i$  at time  $t'$  and departs at time  $t$ , then its delay is  $d_i(t) = t - t'$ . Note that the delay function is independent of destination, since all the vehicles are entrapped in the same queue. The FIFO queuing discipline implies that the cumulative number of vehicles with destination  $j$  that arrived prior to  $t'$  must also have departed by time  $t$ . (This must be true because otherwise late departing vehicles would have been overtaken by the departure at  $t$ .) This condition is sufficient to determine non-decreasing cumulative departure curves  $D_{ij}(t)$  by O-D pair, as follows:

$$D_{ij}(t) = A_{ij}(t - d_i(t)) \quad \forall \quad i, j, t \tag{1}$$

This is the FIFO condition. Since the departure functions should be non-decreasing and should not exceed the arrival functions, we also require:

$$\dot{D}_{ij}(t) \geq 0 \text{ and } D_{ij}(t) \leq A_{ij}(t) \quad \forall \quad i, j, t \quad (2)$$

where an overdot is used to denote the derivative with respect to time. Here, and in Equation 3 below, the constraints are not considered for values of  $(i, j, t)$  where  $\dot{D}_{ij}(t)$  does not exist<sup>3</sup>. The capacity condition is:

$$\sum_{i,j} \gamma_{ijk} \dot{D}_{ij}(t - \tau_{ik}) \leq c_k(t) \quad \forall \quad k, t \quad (3)$$

The argument of  $\dot{D}_{ij}$  expresses the fact that the flows at bottleneck  $b_k$  had to be released from  $o_i$  a trip time earlier. This constraint implies that the number of vehicles passing a bottleneck in any time interval cannot exceed the the integral of the capacity function over the interval, whether or not the interval includes points of discontinuity. Finally, we require all queues to be cleared outside a time interval of interest:

$$D_{ij}(t) = A_{ij}(t); \quad \forall i, j, \quad t \notin (t_B, t_E) \quad (4)$$

Equations (1)–(4) specify the feasibility of a control. Given a feasible solution to these constraints, the ramp metering strategy is given by  $\dot{D}_i(t) = \sum_j \dot{D}_{ij}(t)$ , which defines the time-dependent rate at which vehicles should be released from each origin  $o_i$ .

Equations (1)–(4) can be cast in the standard form of control theory including a differential equation of state dynamics. This can be done in a variety of ways; e.g. by using  $\{d_i(t), D_{ij}(t)\}$  as the “state” and letting the derivative of the FIFO condition be part of the dynamic equation with  $\dot{d}_i(t)$  as controls. Formulation (1)–(4) is retained, however, because it is better suited for our purposes. An optimal control theory formulation would be completed by defining an objective function. This is not done here, however, because our goal is examining the nature of the feasible region. Section 3 shows that even in cases where the input data are smooth, there are instances where the feasible region defined by (1)–(4) consists of a finite, but large

number of “points.” This negates the usefulness of variational methods. Section 4 shows that identifying whether one such point exists is an *NP*-hard problem, and Section 5 that the version of (1)–(4) formulated in discrete time is *NP*-complete.

## 2 A ramp metering feasibility problem

Consider the following decision problem which is a feasibility version of the access control optimization problem. The problem as specified includes smoothness conditions on the input and output functions ( $n$ -differentiability, for any fixed  $n \geq 1$ ) that are sufficient to allow the use of any variational solution method. These conditions are not required for the proofs (they actually complicate the proofs somewhat), but they are imposed to show that the ramp metering feasibility problem is hard even if the data are of the type that would suggest the application of well-established variational methodologies. Establishing the results under these conditions is useful because it shows that variational methods are not well suited to solve general access control problems with FIFO.

### Capacitated Network Access Control Problem with FIFO (CNAP)

*Instance:* Given are finite sets  $O = \{o_i\}$ ,  $E = \{e_j\}$ , and  $B = \{b_k\}$ ; and a real interval  $T = [t_B, t_E]$ . Given for each  $(o_i, e_j) \in O \times E$  is a monotonic non-decreasing,  $(n + 1)$ -differentiable function  $A_{ij} : R \rightarrow R_0^+$  where  $A_{ij}(t_B) = 0$ ; and for each  $b_k \in B$  is a  $n$ -differentiable function  $c_k : R \rightarrow R_0^+$ . Given for each  $(o_i, e_j, b_k) \in O \times E \times B$  is a binary indicator  $\gamma_{ijk} \in \{0, 1\}$ ; and for each  $(o_i, b_k) \in O \times B$  is  $\tau_{ik} \in R_0^+$ .

*Question:* Does there exist piecewise 2-differentiable functions  $D_{ij} : R \rightarrow R_0^+$  for all  $(o_i, e_j) \in O \times E$  and  $d_i : R \rightarrow R_0^+$  for  $o_i \in O$  satisfying conditions (1)–(4)?

Thus, CNAP determines whether or not a feasible dynamic access control strategy exists for a given system. It will now be shown that there is a subset of CNAP whose instances only admit a finite number of feasible solutions.

### 3 CNAP instances with sequential release solutions

Certain instances of CNAP can be shown to have only *sequential release* solutions, i.e. solutions that serve each origin in sequence, releasing all vehicles. This section describes a class of instances, denoted  $U_{SR}$ , whose members have this property. Instances of type  $U_{SR}$  have periodic demand and capacity.

The network structure for problems of type  $U_{SR}$  is depicted in Figure 1<sup>4</sup>. Instances contain  $I$  origins,  $O = \{o_1, o_2, \dots, o_I\}$ , each connected to two destinations,  $E = \{e_1, e_2\}$ . Each destination is also associated with a bottleneck,  $B = \{b_1, b_2\}$ . Mathematically, the network connectivity is given by

$$\gamma_{ijk} = \begin{cases} 1, & j = k \\ 0, & \text{otherwise} \end{cases}, \quad \forall i, j, k$$

All travel times between origins and bottlenecks are assumed to be zero,  $\tau_{ik} = 0 \quad \forall i, k$ , and the time interval  $T$  is defined by  $t_B = -\delta$ ,  $t_E = 2I + \delta$ , where  $\delta$  is a small positive constant that remains fixed throughout this paper.

The arrival and capacity functions for instances of type  $U_{SR}$  will be described using *smooth pulse* functions; see Figure 2 and the following definition:

**Definition 3.1 (Smooth Pulse Function)** *A smooth pulse function with parameter  $\Delta$  (where  $\Delta \geq 2\delta$ ),  $p_\Delta : R \rightarrow [0, 1]$ , is defined as follows:*

$$p_\Delta(t) \text{ } n\text{-differentiable} \tag{5}$$

$$p_\Delta(t) = \dot{p}_\Delta(t) = 0, \quad t \leq -\delta, \quad t \geq \Delta + \delta \tag{6}$$

$$p_\Delta(t) = 1, \quad t \in [\delta, \Delta - \delta] \tag{7}$$

$$p_\Delta(t) \text{ monotonic increasing, } t \in (-\delta, \delta) \tag{8}$$

$$p_\Delta(t) = 1 - p_\Delta(-t), \quad t \in (-\delta, \delta) \tag{9}$$

$$p_\Delta(t) = p_\Delta(\Delta - t), \quad \forall t \tag{10}$$

□

Note that as  $\delta \rightarrow 0$ ,  $p_\Delta$  approximates a rectangular pulse of length  $\Delta$ . Condition (9) implies that  $p_\Delta$  is symmetric about the point  $(0, 0.5)$  on the interval  $(-\delta, \delta)$ . Condition (10) further implies that  $p_\Delta$  is symmetric about the line  $\frac{\Delta}{2}$ . Taken together, these conditions imply a certain decomposition property, namely that  $p_\Delta(t) + p_{\Delta'}(t - \Delta) = p_{\Delta + \Delta'}(t)$  for any two parameter values  $\Delta$  and  $\Delta'$ . Also, these conditions imply that  $\int_R p_\Delta(t) dt = \Delta$ . In the descriptions to follow, assume that  $\delta \ll 1$  and let the unit pulse function  $p_1(t)$  be abbreviated by  $p(t)$ .

The cumulative arrival functions for each origin  $o_i$  are identical and specified in terms of the dynamic arrival rates:

$$\dot{A}_{i1}(t) = p(t) \quad \forall i \quad (11)$$

$$\dot{A}_{i2}(t) = p(t - 1) \quad \forall i \quad (12)$$

with  $A_{ij}(t_B) = 0 \quad \forall i, j$ . Thus, vehicle arrivals at each origin consist of a smooth pulse of arrivals for destination (bottleneck)  $e_1$  ( $b_1$ ), followed by a pulse for destination (bottleneck)  $e_2$  ( $b_2$ ) with some mixing in the interval  $[1 - \delta, 1 + \delta]$  (see Figure 3(a)). Figure 3(b) depicts the cumulative arrival curves.

The bottleneck capacity functions for  $b_1$  and  $b_2$  are defined to be two periodic series of pulses, identical except for a time shift:

$$c_1(t) = \sum_{m=1}^I p(t - 2(m - 1)) \quad (13)$$

$$c_2(t) = c_1(t - 1) \quad (14)$$

As shown in Figure 3(c), there are exactly  $I$  capacity pulses for each bottleneck; the maximum number of vehicles that can be served by each is  $\int_R c_1(t) dt = \int_R c_2(t) dt = I$ .

It should be clear that all instances of type  $U_{SR}$  have feasible solutions and are thus “yes” instances. One such solution is to serve origin 1 first without metering beginning at  $t = -\delta$ ,

holding all other origins until  $t = 2 - \delta$ . At  $t = 2 - \delta$ , origin 2 is also released while origins  $o_i$ ,  $i > 2$  are held, and so on. Mathematically, this is equivalent to setting  $d_i(t) = 2(i - 1) \forall i, t$  so that the departure curves for each O-D pair consistent with (1) are just the arrival curves translated in time by a non-negative, origin-specific delay. Clearly then, (2) is satisfied and since the maximum delay (shift) is  $2(I - 1)$ , we see from Figure 3(b) that  $D_{ij}(t) = A_{ij}(t) = 1 \forall i, j, t > t_E = 2I + \delta$  and thus (4) is satisfied as well. To verify that the capacity condition is also satisfied, note that the specified  $\tau_{ik}$  and  $\gamma_{ijk}$ , and the constant time shifts imply that the LHS of (3) is:

$$\sum_{i,j} \gamma_{ijk} \dot{D}_{ij}(t) = \sum_i \dot{A}_{ik}(t - d_i(t)) = \sum_i \dot{A}_{ik}(t - 2(i - 1)) \quad \forall k \quad (15)$$

Substitution of (11) or (12) in (15) reduces it to (13) or (14) depending on  $k$ , i.e.:

$$\sum_{i,j} \gamma_{ijk} \dot{D}_{ij}(t - \tau_{ik}) = c_k(t), \quad \forall t, \quad k = 1, 2 \quad (16)$$

Thus, (3) is satisfied as an equality  $\forall t$ . In this case, bottlenecks  $b_1$  and  $b_2$  are “saturated” from  $t = -\infty$  to  $+\infty$ . This concept is formalized below:

**Definition 3.2 (Bottleneck Saturation/Undersaturation in an Interval)** *Bottleneck  $b_{k_0}$  is said to be saturated in an interval  $(t_0, t_1]$  if (3) is satisfied (in the interval) for  $k = k_0$ , and if in addition<sup>5</sup>:*

$$\int_{t_0}^{t_1} \sum_{i,j} \gamma_{ijk_0} \dot{D}_{ij}(t - \tau_{ik_0}) dt = \int_{t_0}^{t_1} c_{k_0}(t) dt \quad (17)$$

*If (3) is satisfied but (17) is a strict inequality the bottleneck will be said to be undersaturated in the interval; the difference between the two sides of the inequality will then be called the spare capacity in the interval.  $\square$*

Solutions such as the feasible solution just identified, where origins are released in sequence and then remain unmetered will be called *sequential release* solutions. They are formally defined below:

**Definition 3.3 (Sequential Release Solutions)** Let  $\langle v_i \rangle$  be a permutation of the origin indices  $i = 1, 2, \dots, I$ . Then, solution  $[d_i(t), D_{ij}(t)]$  is a sequential release if the  $D_{ij}(t)$  arise from (1) with  $d_{v_i}(t) = 2(i - 1) \quad \forall i, t$ .  $\square$

For the discussion that follows, it will be convenient to define a function  $C_k : R \rightarrow R_0^+$  for the cumulative capacity,  $C_k(t) = \int_{-\infty}^t c_k(x)dx$ , and the notation  $A_{.j}$  ( $D_{.j}$ ) for  $\sum_i A_{ij}$  ( $\sum_i D_{ij}$ ).

Consider now the following proposition:

**Proposition 3.1** A solution to an instance of type  $U_{SR}$  is feasible if, and only if, it is a sequential release.

*Proof.* The sufficiency of the condition clearly follows from symmetry; since each origin has an identical arrival function, the arguments that were given earlier to show the feasibility of the sequential release  $\langle v_i = i \rangle$  also apply to any permutation.

To show necessity, first note from (11)–(14) that:

$$I = A_{.j}(t_E) = C_j(t_E) \quad j = 1, 2 \tag{18}$$

Further, it is claimed that:

$$D_{.j}(t) = C_j(t) \quad \forall t, j = 1, 2. \tag{19}$$

The proof of (19) is by contradiction. First note that insofar as  $\tau_{ij} = 0 \quad \forall i, j$ , integration of condition (3) ensures both that:

$$D_{.j}(t) \leq C_j(t) \quad \forall t, j = 1, 2 \tag{20}$$

and that:

$$D_{.j}(t_E) - D_{.j}(t) \leq C_j(t_E) - C_j(t) \quad \forall t, j = 1, 2 \tag{21}$$

Therefore, if (19) were false,  $D_{.j}(t) < C_j(t)$  for some  $j, t$ . This would imply that  $D_{.j}(t_E) - D_{.j}(t) > D_{.j}(t_E) - C_j(t)$ , which in turn would imply  $D_{.j}(t_E) - D_{.j}(t) > C_j(t_E) - C_j(t)$  since (4)

and (18) ensure  $D_{.j}(t_E) = A_{.j}(t_E) = C_j(t_E)$ . This is impossible, however, since it contradicts (21). Thus, (19) holds.

To continue the necessity proof, suppose now that it is feasible to release from  $t = -\delta$  to  $1 - \delta$  a positive number of vehicles from more than one origin such that (19) is satisfied in the interval. It is shown now that this would lead to a subsequent violation of (19) and therefore that a single origin must be released from  $-\delta$  to  $1 - \delta$ . Let  $o_\ell$  be an origin which has released the most vehicles by time  $1 - \delta$ , i.e.  $D_{\ell 1}(1 - \delta) \geq D_{i1}(1 - \delta) \quad \forall i$ , and define  $\epsilon \equiv C_1(1 - \delta) - D_{\ell 1}(1 - \delta) > 0$ . The FIFO condition, therefore, implies that origin  $o_\ell$  needs to discharge  $\epsilon > 0$  vehicles for bottleneck  $b_1$  (and all other origins need to discharge more) before any vehicles for bottleneck  $b_2$  could be released; such a discharge requires time  $t_\epsilon > 0$ . Thus,  $D_{.2}(1 - \delta + t_\epsilon) - D_{.2}(1 - \delta) = 0$ . Since (19) is satisfied at  $t = 1 - \delta$  and  $C_2(1 - \delta + t_\epsilon) - C_2(1 - \delta) > 0$ , condition (19) is violated for  $j = 2$  and  $t = 1 - \delta + t_\epsilon$ . It follows, therefore, that a single origin  $o_{v_1}$  must be released from  $-\delta$  to  $1 - \delta$ . Similar arguments are now used to show that  $o_{v_1}$  must remain fully released until  $t = 2 + \delta$ , thus saturating bottleneck  $b_2$  during the interval  $[1 - \delta, 2 + \delta]$ . If this were not true, there must exist a brief interval  $[t, t + t_{\epsilon'}]$  where for the first time origin  $o_{v_1}$  sends  $\epsilon' > 0$  vehicles less than the maximum to  $b_2$ . To satisfy (19), this deficit would have to be made up by other origins. This is impossible, however, because for sufficiently small  $t_{\epsilon'}$  origins other than  $o_{v_1}$  can only release vehicles to  $b_1$ . (Recall that each of these origins has  $C_1(1 - \delta)$  vehicles destined through  $b_1$  at the head of its queue.) Therefore,  $o_{v_1}$  must be fully released from  $-\delta$  to  $2 + \delta$  and  $D_{v_1 j}(t)$  will be the result of setting  $d_{v_i} = 2(i - 1)$  for  $i = 1$ .

Similar logic is now applied to the remaining origins. At time  $2 - \delta$ , bottleneck  $b_1$  begins another capacity pulse. Although origin  $o_{v_1}$  remains unrestricted from  $2 - \delta$  to  $2 + \delta$ , it discharges vehicles only to bottleneck  $b_2$  during this interval and releases no additional vehicles after this interval. Thus, the problem of determining a feasible metering strategy for the remaining origins  $O \setminus \{o_{v_1}\}$  from time  $2 - \delta$  onward is identical to the original with one fewer origin. By induction, therefore, we conclude that exactly one origin must be fully released at times  $2 - \delta, 4 - \delta, \dots, 2(I - 1) - \delta$ , and thus every solution is a sequential release.  $\square$

Proposition 3.1 shows that all problem instances of type  $U_{SR}$  have an arbitrarily large but finite number of feasible solutions and therefore that the associated optimization problems cannot be solved with variational methods. It will now be shown that CNAP is an  $NP$ -hard problem through a polynomial transformation of the Directed Hamiltonian Path problem.

## 4 The continuous-time CNAP is $NP$ -hard

A Hamiltonian path on a directed graph  $G = (V, A)$  with numbered vertices is a sequence  $\langle v_1, v_2, \dots, v_{|V|} \rangle$  of distinct vertices from  $V$  such that  $(v_{i-1}, v_i) \in A$  for  $1 < i \leq |V|$ . Karp (1972) showed that the problem of the existence of a Hamiltonian cycle in an undirected graph is  $NP$ -complete, and Garey and Johnson (1979) extended this result to show that determining the existence of a Hamiltonian path on a directed graph (problem DHP) is also  $NP$ -complete. In this section, it is shown that every DHP instance,  $G$ , can be polynomially-transformed to an instance of CNAP, denoted  $U_{HP}(G)$ .

Let  $G = (V, A)$  (where  $I = |V|$ ) be an instance of DHP, and let  $A'$  be the set of complement arcs to  $A$ ; i.e., if  $(i, j) \notin A$  then  $(i, j) \in A'$  and vice versa<sup>6</sup>. To create  $U_{HP}(G)$ , first a pulse function  $p_\Delta$  is specified with the properties given in Definition 3.1, and additionally  $p_\Delta$  is assumed to be piecewise-polynomial. For each node  $i \in V$ , a corresponding origin  $o_i$ , destination  $e_i$ , and bottleneck  $b_i$  are created. Two additional destinations and bottlenecks are also created. Thus,  $O = \{o_1, o_2, \dots, o_I\}$ ,  $E = \{e_1, e_2, \dots, e_I, e_{I+1}, e_{I+2}\}$  and  $B = \{b_1, b_2, \dots, b_I, b_{I+1}, b_{I+2}\}$ .

The network structure descriptors are specified as follows (see Figure 4 for a depiction):

$$\gamma_{ijk} = \begin{cases} 1, & \text{if } j = k \\ 0, & \text{otherwise} \end{cases}, \quad \forall i, j, k \quad (22)$$

$$\tau_{ik} = \begin{cases} 2, & k \in \{1, 2, \dots, I\}, i \neq k \\ 0, & \text{otherwise} \end{cases} \quad (23)$$

Thus, travel times are zero between all origins and bottlenecks  $b_{I+1}$  and  $b_{I+2}$ , as well as from origin  $o_i$  to bottleneck  $b_i$  for all  $i$ . The time interval  $T$  is given by  $t_B = -\delta$  and  $t_E = 2(I+1) + \delta$ .

Next, bottleneck capacities are specified as follows:

$$c_k(t) = \begin{cases} \sum_{m=1}^I p(t - 2(m - 1)), & k = I + 1 \\ c_{I+1}(t - 1), & k = I + 2 \\ 1, & \text{otherwise} \end{cases} \quad (24)$$

Thus, bottlenecks  $b_{I+1}$  and  $b_{I+2}$  have periodic capacities identical to those described for instances of type  $U_{SR}$ , while the other bottlenecks have constant capacity.

Finally, the arrival functions are specified as follows:

$$\dot{A}_{ij}(t) = \begin{cases} p(t), & j = I + 1 \\ p(t - 1), & j = I + 2 \\ p_2(t), & j = i \quad , \quad \forall i \\ p_2(t), & (i, j) \in A' \\ 0, & \text{otherwise} \end{cases} \quad (25)$$

with  $A_{ij}(t_B) = 0 \quad \forall i, j$ . Arrivals at each origin destined for  $e_{I+1}$  and  $e_{I+2}$  and their associated bottlenecks are identical to those in  $U_{SR}$ . Additionally, a double-length pulse of vehicles arrives at origin  $o_i$  for destination  $e_i$ . Finally, if  $(i, j) \in A'$ , a double-length pulse of vehicles arrives at  $o_i$  for  $e_j$ .

To show that  $U_{HP}(G)$  is a polynomial transformation of the Directed Hamiltonian Path instance  $G$ , first consider the following proposition:

**Proposition 4.1** *A solution to  $U_{HP}(G)$  is feasible only if it is a sequential release solution.*

*Proof.* This is true since the constraints defining  $U_{HP}(G)$  include a subset (pertaining to destinations  $e_{I+1}$  and  $e_{I+2}$ ) that defines an instance of type  $U_{SR}$  as a subproblem. Thus, the necessity claim of Proposition 3.1 holds.  $\square$

Proposition 4.1 guarantees that every solution to  $U_{HP}$  corresponds to an ordering of the origin releases. Next, the permutations that satisfy the remaining constraints are characterized. It is also easy to verify the following:

**Proposition 4.2** *The capacity constraints (3) for bottlenecks  $b_{I+1}$  and  $b_{I+2}$  are satisfied by all sequential release solutions to  $U_{HP}(G)$ .*

Now consider the following proposition:

**Proposition 4.3** *Let  $\langle v_i \rangle$  be a permutation of the origin indices corresponding to a sequential release solution to  $U_{HP}(G)$ . This solution is feasible if, and only if,  $(v_{i-1}, v_i) \notin A'$  for all  $i = 2, \dots, I$ .*

*Proof.* First, Proposition 4.2 ensures that (3) are satisfied for bottlenecks  $b_{I+1}$  and  $b_{I+2}$  by all sequential releases. Now consider the remaining bottlenecks. Note from (22)–(25) that the capacity of a bottleneck  $b_{v_i}$ ,  $1 < i \leq I$  will be violated if and only if flow from an origin other than  $o_{v_i}$  is arriving at  $b_{v_i}$  when origin  $o_{v_i}$  is discharging flow. That is, the capacity constraint of  $b_{v_i}$  is violated at some time if and only if the origin preceding  $o_{v_i}$  in the release sequence  $(o_{v_{i-1}})$  also sends flow to  $b_{v_i}$ . This occurs if and only if  $(v_{i-1}, v_i) \in A'$ . (Note that the constraint for bottleneck  $b_{v_1}$  is always feasible.) Clearly then, (3) is satisfied for  $i = 2, \dots, I$  if and only if  $(v_{i-1}, v_i) \notin A'$ .  $\square$

### Theorem 1

*Every instance of DHP is polynomially-transformable to CNAP.*

*Proof.* The instance of CNAP denoted  $U_{HP}(G)$  can clearly be generated in a period of time that is bounded by a polynomial function of the size of the DHP instance  $G = (V, A)$ . Now it is shown that  $G$  contains a directed Hamiltonian path if, and only if, there exists a feasible metering scheme for  $U_{HP}(G)$ . First, let  $P = \langle v_1, v_2, \dots, v_n \rangle$  be a directed Hamiltonian path in  $G$ . Then,  $d_{v_i}(t) = 2(i - 1)$ ,  $\forall i, t$  is a feasible metering scheme. To see this, note that  $P$  is a Hamiltonian path and therefore  $(v_{i-1}, v_i) \in A$ ,  $(v_{i-1}, v_i) \notin A'$  for  $i = 2, \dots, I$ . Thus,

$P$  is a permutation of origins satisfying the conditions of Proposition 4.3, and its associated sequential release solution,  $d_{v_i}(t) = 2(i - 1)$ ,  $\forall i, t$ , must be feasible. Second, suppose there exists a feasible metering scheme for  $U_{HP}(G)$ . Proposition 4.1 guarantees that all solutions to  $U_{HP}$  are sequential releases. With that in mind, let  $\langle v_i \rangle$  be the permutation of origin indices corresponding to a feasible metering scheme. Proposition 4.3 guarantees that  $(v_{i-1}, v_i) \notin A'$  for  $i = 2, \dots, I$  and therefore that  $(v_{i-1}, v_i) \in A$ . Thus,  $\langle v_i \rangle$  is a directed Hamiltonian path in  $G$ .  $\square$

It is possible to specify a number of additional constraints that would define a particular subset of the problem CNAP for which all of the requirements of the class  $NP$  would be satisfied. For example, the general problem CNAP is defined in terms of real variables; a rational requirement could instead be imposed. Moreover, the restrictions placed on functions  $A_{ij}$  and  $c_k$  are not sufficient to guarantee the existence of a reasonable encoding scheme for the inputs, although this could be done with more restrictions. The general CNAP problem without these extra constraints, however, does not belong to the problem class  $NP$ .

**Corollary 1.1**

*CNAP is NP-hard.*

*Proof.* Since a known  $NP$ -complete problem is polynomially-transformable to CNAP, but  $CNAP \notin NP$ , the result follows.  $\square$

In the next section, a stronger result is proved for a very important related discrete-time problem.

## 5 The discrete-time CNAP is NP-complete

To this point, this paper has shown that the network access control problem is difficult to solve by applying variational methods to a continuous formulation. Often, difficult control theory problems are attacked by formulating a tractable mathematical program to solve a discretized version of the control problem. This section completes the the discussion of the hardness of optimal access control by showing that the discrete-time version of CNAP is in the problem class  $NP$ -complete.

The natural way to discretize the problems described in Section 1 is to partition the study time period  $T$  into many small intervals of width  $\xi > 0$  and then assume that the arrival flows and bottleneck capacities are constant within each of these intervals. Additionally, it is necessary to approximate the vehicle delays and trip times using integer multiples of  $\xi$  in order to properly model the FIFO queues; all vehicles that arrive in an interval will then depart together and experience the same delay. Such an approximation can be refined arbitrarily by letting  $\xi \rightarrow 0$ .

These ideas are formalized below:

### Discretized Capacitated Network Access Control Problem with FIFO (DNAP)

*Instance:* Given are finite sets  $O = \{o_i\}$ ,  $E = \{e_j\}$ , and  $B = \{b_k\}$ ; rationals  $t_E \in Q_0^+$  and  $\xi \in Q^+$  such that  $t_E = N\xi$  for some  $N \in Z_0^+$ . Let  $T \equiv \{\xi, 2\xi, \dots, N\xi\}$ . Given for each  $(o_i, e_j, t) \in O \times E \times T$  is  $a_{ij}(t) \in Q_0^+$ , and let  $A_{ij}(t) \equiv \sum_{n=1}^{\frac{t}{\xi}} a_{ij}(n\xi)$ . Additionally, define  $A_{ij}(t) \equiv 0 \quad \forall t \leq 0$ . Given for each  $(b_k, t) \in B \times T$  is  $c_k(t) \in Q_0^+$ . Given for each  $(o_i, e_j, b_k) \in O \times E \times B$  is a binary indicator  $\gamma_{ijk} \in \{0, 1\}$ ; and for each  $(o_i, b_k) \in O \times B$  is  $\eta_{ik} \in Z_0^+$  where  $\tau_{ik} \equiv \eta_{ik}\xi$ .

*Question:* Do there exist integers  $n_i(t) \in Z_0^+$  for all  $(o_i, t) \in O \times T$  and rationals  $D_{ij}(t) \in Q_0^+$  for all  $(o_i, e_j, t) \in O \times E \times T$  satisfying:

$$d_i(t) = n_i(t)\xi \quad \forall \quad t \quad (26)$$

$$D_{ij}(t) = A_{ij}(t - d_i(t)) \quad \forall \quad i, j, t \quad (27)$$

$$D_{ij}(t) \geq D_{ij}(t - \xi) \text{ and } D_{ij}(t) \leq A_{ij}(t) \quad \forall \quad i, j, t \quad (28)$$

$$\sum_{i,j} \gamma_{ijk} (D_{ij}(t - \tau_{ik}) - D_{ij}(t - \tau_{ik} - \xi)) \leq c_k(t) \quad \forall \quad k, t \quad (29)$$

$$D_{ij}(N\xi) = A_{ij}(N\xi) \quad \forall \quad i, j \quad (30)$$

where  $D_{ij}(t) \equiv 0, t \leq 0$ .

**Theorem 2**

*DNAP is NP-complete.*

*Proof.* Verifying that  $\text{DNAP} \in \text{NP}$  is straightforward. The instance can be encoded with a length that is  $O(|O| \cdot |E| \cdot |B| \cdot N)$  since all data are integer or rational, and the functions are defined on finite sets. Furthermore, it is straightforward to see that a non-deterministic algorithm could verify whether a “guess” assignment of integers  $d_i(t)$  satisfies (26)–(30) in time  $O(|O| \cdot |E| \cdot |B| \cdot N)$ .

Parallel to the proof in Section 4, it is now shown that the Directed Hamiltonian Path problem (DHP) is polynomially-transformable to DNAP. Given an instance of DHP,  $G = (V, A)$ , again let  $A'$  be the set of complement arcs to  $A$ . To create an instance of DNAP,  $U_{HPD}(G)$ , that can solve instance  $G$ , specify sets  $O$ ,  $E$ , and  $B$  and network descriptors  $\gamma$  and  $\tau$  identical to those of CNAP instance  $U_{HP}(G)$  described in Section 4. Further, let  $\xi = 1$  and let  $t_E = N = 2(I+1)$ , recalling that  $I = |O|$ ; therefore,  $T = \{1, 2, \dots, 2(I+1)\}$ . Specify bottleneck capacities as follows:

$$c_k(t) = \begin{cases} 1, & k = I + 1, \quad t \in \{1, 3, \dots, (2I - 1)\} \\ 1, & k = I + 2, \quad t \in \{2, 4, \dots, 2I\} \\ 1, & k \leq I, \quad t \in T \\ 0, & \text{otherwise} \end{cases} \quad (31)$$

and arrival patterns:

$$a_{ij}(t) = \begin{cases} 1, & j = I + 1, \quad t = 1 \\ 1, & j = I + 2, \quad t = 2 \\ 1, & j = i, \quad t \in \{1, 2\} \\ 1, & (i, j) \in A', \quad t \in \{1, 2\} \\ 0, & \text{otherwise} \end{cases}, \quad \forall i \quad (32)$$

First, it is clear that  $U_{HPD}(G)$  can be generated in polynomial time with respect to the size of  $G$ . Reasoning parallel to that in Sections 3 and 4 shows that  $U_{HPD}(G)$  is a transformation of  $G$ . The minor differences in the logic are outlined below.

Note first that Definition 3.1 is not needed, and that Definition 3.2 needs to be modified for discrete intervals<sup>7</sup>. In the new definition, the two sides of (17) are replaced by sums of the two sides of (29). Definition 3.3 holds verbatim. It is now possible to show that Proposition 4.1 holds for instance  $U_{HPD}(G)$ . Informally, this can be seen since  $U_{HPD}(G)$  is equivalent to  $U_{HP}(G)$  as  $\delta \rightarrow 0$  and Proposition 3.1 holds for  $\delta \rightarrow 0$ ; it can also be proven using parallel (but simpler) logic to the proof of necessity of Proposition 3.1. As in the continuous case, it is easy to verify that Proposition 4.2 holds for  $U_{HPD}(G)$ . Finally, Proposition 4.3 and Theorem 1 (and their proofs) hold verbatim, with  $U_{HPD}(G)$  substituted for  $U_{HP}(G)$  and DNAP for CNAP. Thus, the  $NP$ -complete problem DHP is polynomially-transformable to DNAP, and since  $DNAP \in NP$ , DNAP is therefore  $NP$ -complete.  $\square$

## Notes

<sup>1</sup>The notation  $R_0^+$  indicates the set of non-negative real numbers

<sup>2</sup>In this paper, index variable  $i$  always refers to origins,  $j$  to destinations,  $k$  to bottlenecks, and  $t$  to time. Unless otherwise stated, the generic notation  $\forall i$  is equivalent to  $i = 1, 2, \dots, I$ ,  $\forall j \equiv j = 1, 2, \dots, J$ ,  $\forall k \equiv k = 1, 2, \dots, K$ , and  $\forall t \equiv \forall t \in R$ . Furthermore, when the symbols  $\sum$  and  $\int$  are subscripted by one or more variables without specifying a range, it should be understood that it is the full range for the variable in question; e.g.  $i = 1, \dots, I$ ,  $j = 1, \dots, J$ , etc.

<sup>3</sup>The symbol  $\forall$  will be used in this restricted way when applied to time. Furthermore, if  $t$  is a generic argument in an expression involving quantities that could be undefined for certain values of  $t$ , it should also be understood that the expression pertains only to the values of  $t$  where all members are well-defined.

<sup>4</sup>The chosen network is unrealistic, but simple. This simplifies the exposition. The results

could also be established with more complicated and realistic networks.

<sup>5</sup>The (Lebesgue) integrals exist since the integrands are well-defined and bounded, except on a set of measure zero where the integrand on the left is not defined.

<sup>6</sup> $A'$  contains no arcs  $(i, j)$  where  $i = j$

<sup>7</sup>Such that  $t_0$  and  $t_1$  are multiples of  $\xi$ .

## Acknowledgement

This work was supported in part by a grant from the California Partners for Advanced Transit and Highways (PATH) under contract MOU-3004.

## References

- Garey, M. and Johnson, D. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H. Freeman, New York.
- Gazis, D. (1974). Traffic control: Theory and application. In D. Gazis, editor, *Traffic Science*, chapter 3. Wiley, New York.
- Karp, R. (1972). Reducibility among combinatorial problems. In R. Miller and J. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103. Plenum Press, New York.
- Lovell, D. (1997). *Traffic Control on Metered Networks without Route Choice*. Ph.D. thesis, University of California, Berkeley.
- Lovell, D. and Daganzo, C. (1999). Access control on networks with unique origin-destination paths. *Transportation Research B*, **34**, 185–202.
- May, A. (1964). Experimentation with manual and automatic ramp control. *Highway Research Record*, **59**.

- Newell, G. (1982). *Applications of Queueing Theory*. Chapman and Hall, New York, 2 edition.
- Papageorgiou, M. (1995). An integrated control approach for traffic corridors. *Transportation Research C*, **3**(1), 19–30.
- Stephanedes, Y. and Kwon, E. (1993). Adaptive demand-diversion prediction for integrated control of freeway corridors. *Transportation Research C*, **1**(1), 23–42.
- Wattleworth, J. (1963). *Peak-period Control of a Freeway System—Some Theoretical Considerations*. Ph.D. thesis, Northwestern University.
- Wattleworth, J. (1967). Peak-period analysis and control of a freeway system. *Highway Research Record*, **157**.
- Yuan, L. and Kreer, J. (1968). An optimal control algorithm for ramp metering of urban freeways. In *IEEE 6th Annual Allerton Conference on Circuit and System Theory*, University of Illinois.

# Figures

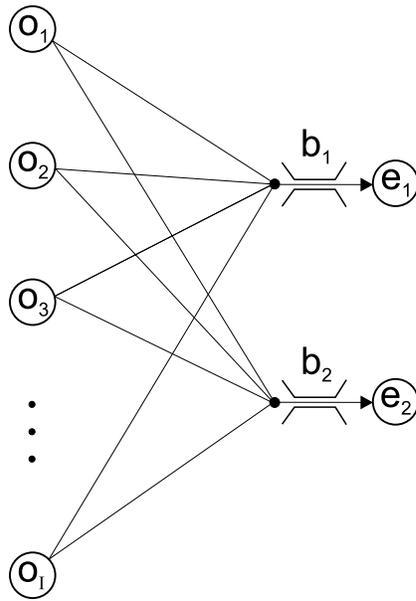


Figure 1: Network Structure for Instances of Type  $USR$



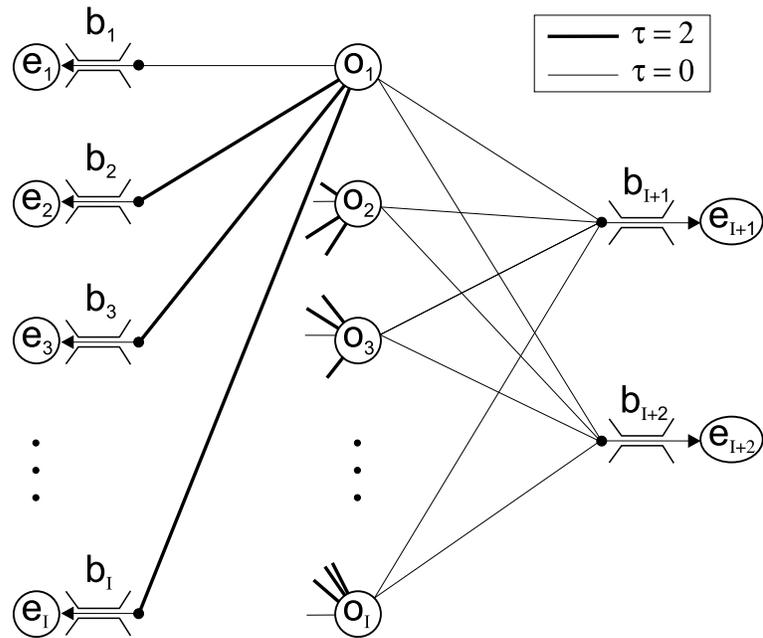


Figure 4: Network Structure for Instances of Type  $U_{HP}$