

RESEARCH

Open Access



Optimal transmission of messages in computer networks – an optimal control problem involving control-dependent time-delayed arguments

Kar-hung Wong¹, Yu-chung Eugene Lee^{2*}  and Heung-wing Joseph Lee²

*Correspondence:

eugeneycle@yahoo.com

²Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
Full list of author information is available at the end of the article

Abstract

In this paper, we find the optimal transmission of messages in computer networks. This problem has been formulated as a nondelayed optimal control problem in several recent papers on TCP (transmission control protocol). Since the actual transmission of messages from origins to destinations should consist of both forward transmission delays of the buffers and latency of the links, we remodel the problem as a time-delayed optimal control problem consisting of both control-dependent time-delayed arguments and discrete time-delayed arguments. We then develop a modified control parameterization method for solving this time-delayed optimal control problem. The gradients of the new objective function and constraint functions generated by this modified control parametrization method are derived. A numerical example is solved by using the time-delayed version of the problem that we formulate, as well as the nondelayed version of the problem in the literature. Numerical results clearly illustrate the efficiency of the modified control parameterization method for solving both versions of this optimal transmission problem. Comparison of results of the two versions concerning the optimal transmission rates at the origins, the optimal output flow rates at the destination, and the queue sizes at the buffers are obtained. These comparison results clearly reflect how the optimal transmission of messages in computer networks in real life can be affected by both the forward transmission delays of the buffers and the latency of the links.

Keywords: Optimal transmission in computer networks; Buffer equations; Forward transmission delays of the buffers; Latency of the links; Optimal control problem with both control-dependent time-delayed arguments and discrete time-delayed arguments; Modified control parameterization method; Gradient formula

1 Introduction

In computer networks, queues build up in the buffer of the links when the input rates are larger than the available bandwidth. Long queues can cause congestion leading to packet losses and delays in transmission. Various versions of TCP (transmission control proto-

© The Author(s) 2022. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

col) have been developed. The existing congestion control techniques aim to adjust the transmission rates of the competing users in such a way that the network resources are efficiently shared.

Since the pioneer works [1] and [2], a lot of efforts ([3–14]) have been devoted to the modeling and design of the internet congestion control. These research works involve the establishment of a fluid model for finding the proper resource allocation of the network. Different resource allocation algorithms, such as primal, dual, and primal-dual algorithms have been developed, which enable the user to find the optimal transmission rates asymptotically using local feedback from the network.

Despite the progress in the analysis and synthesis of internet congestion control, an important modeling issue has been neglected for simplicity. Specifically, most existing fluid models of congestion control assume that the flow rate of a TCP flow at every link in its path is equal to the source rate at the first link. However, in practice, due to the queuing effect caused by more than one source using the same link, the rate of the TCP flow from each source in an intermediate link is usually (but not always) less than that of its preceding links. Thus the rate of a TCP flow should be decreasing most of the time as it travels from one link to another link from source to destination. In view of the above shortcoming of the existing models, [12] developed a closed-loop form for the buffer dynamic to address the instantaneous queuing rate of each TCP flow at the router of every link it passes through; consequently, the instantaneous flow rate of each TCP flow at a particular link can be obtained by subtracting all the queuing rates preceding this link from the input source rate. Hence a more accurate model has been developed for modeling the behavior of the network under the congestion controller. The stability of the congestion control algorithms was devised by imposing several simple conditions.

In this paper, we consider a computer network that has two origins, N sources, three links, and one destination such that each of the TCP flows from the sources uses two links to reach the destination. We then modify the model of [12] further by imposing the time-delay arguments into the buffer equation (i.e., the queuing equation at the buffer of a link) and the conservation of the flow equation (i.e., the equation governing the flow rate at one link and its successive link).

As mentioned in [4], there always exists a time difference between the time that a packet first arrives at the router of a link and the time that the router serializes it onto the link. This is called the forward transmission delay, which is equal to packet size (bits)/transmission rate (bits per second). Furthermore, there also exists a time difference between the time that the packet leaves the router of one link (i.e., immediately after the router has serialized the packet into the link) and the time that it arrives at the receiving end of the link (i.e., the time that the packet reaches the router of the preceding link during its trip from source to destination). This is called the latency of the link, which is a constant for each link (independent of the packet size and transmission rate). Thus by inserting the forward transmission delay arguments into the buffer equations and the latency into the conservation of flow equations we obtain even a more accurate behavior of the network under the congestion controller than that of [12].

Since each link in this “optimal transmitting of messages in computer networks” model may be shared by more than one source, competition for flow rate exists at some or all the given links. Thus the objective of this problem should involve allocating the flow rate of each link to all the competing sources in such a way that the weighted sum of all the

messages sent by the sources to the destination in each time interval is maximized, subject to the conditions that both the buffer capacity and the link capacity (which is usually called the bandwidth of the link) are not violated.

Using the above objective function, we first formulate this problem as a time-delayed optimal control problem consisting of both control-dependent time-delayed arguments and discrete-time-delayed arguments. The control-dependent time-delayed arguments arise from the transmission delays of the buffers and the discrete-time-delayed arguments arise from the latency of the links. We then develop a modified control parameterization method for solving this optimal control problem. The gradients of the new objective functions and the constraint functions generated by this modified control parameterization method are derived for the first time in the literature. Finally, a numerical example is solved by using both the nondelayed version and the time-delayed version of the above optimal transmission problem.

Numerical results clearly illustrate the efficiency of this modified control parameterization method for solving time-delayed optimal control problems with control-dependent arguments.

Comparison of results of the two versions concerning the optimal transmission flow rates at the origin, the output flow rates at the destination, and the queue sizes at each buffer are obtained. These comparison results clearly reflect how the optimal transmission of messages in computer networks in real life can be affected by both the forward transmission delays of the buffers and the latency of the links.

The classical control parameterization method has been efficiently used for solving lots of optimization problems or optimal control problems, such as constrained optimization problems [15], nondelayed optimal control problems with lumped parameter systems [16–21], nondelayed optimal control problems with distributed parameter systems [22], optimal control problems with constant time delays ([17–19, 23–25]), optimal control problems with given time-varying delays [26], and optimal control problems with unknown time-varying delays [27]. However, the extension of this method to solving optimal control problems with control-dependent time-delayed arguments is much more complicated due to the reasons given in the next two paragraphs.

To solve nondelayed optimal control problems, or optimal control problems involving constant time delays h_1, \dots, h_n , or given time-varying delays $\alpha_1(t), \dots, \alpha_n(t)$ ([16–26]) by the classical control parametrization method, we simply need to express each control function $u(t)$ as

$$u(t) = \sum_{i=1}^p \sigma^i \chi_i(t), \quad j = 1, \dots, n, t \in [0, T], \quad (1)$$

where $\sigma^1, \dots, \sigma^p$ are the control parameters, p is the number of partitions in the time interval $[0, T]$, and $\chi_i(t)$ is the characteristic function. Then we can easily calculate the gradient of the objective function and the constraint functions with respect to the control parameters $\sigma^1, \dots, \sigma^p$, as well as solve the state equations and the costate equations by any numerical scheme, such as the fourth-order Runge–Kutta method. Although solving nondelayed optimal control problems involving unknown time-varying delays $\alpha_1(t), \dots, \alpha_n(t)$ [27] is slightly more difficult, we can still employ the same technique given above to solve

the problem by the classical control parameterization, except that we need to add additional control vectors $v_1(t), \dots, v_n(t)$ for the n unknown time delays via

$$v_j(t) = \sum_{i=1}^p \sigma_j^i \chi_i(t), \quad j = 1, \dots, n, t \in [0, T], \quad (2)$$

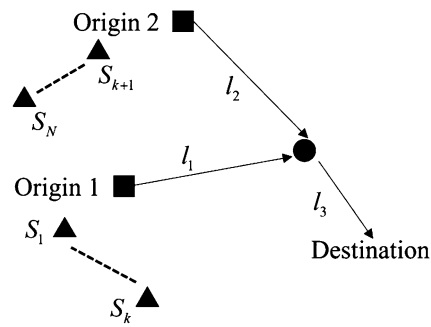
where σ_j^i ($i = 1, \dots, p, j = 1, \dots, n$) are the control parameters.

However, the optimal control problem in this paper involves a control-delayed vector of the form $(u_1(t - \alpha_1(u_1)(t)), \dots, u_m(t - \alpha_m(u_m)(t)))^T$. Thus, to efficiently solve the problem, besides expressing the control vector as vector-valued functions of the control parameters $\sigma^1, \dots, \sigma^p$, we also need to express each component of the control-delayed vector $(u_1(t - \alpha_1(u_1)(t)), \dots, u_m(t - \alpha_m(u_m)(t)))^T$ as a function of the control parameters $\sigma^1, \dots, \sigma^p$, which is a very difficult task. ($\alpha_i(u_i)(t)$, $i = 1, \dots, m$, are given functions of $u_i(t)$.) Since each i th component of the control-delayed vector is a convolution of the function $u_i(t)$, finding the gradient of the objective function or constraint functions of this time-delayed optimal control problem with control-dependent arguments is much harder than finding the optimal control problems with constant time delays or time-varying delays given in [16–27]. Thus, we need to devise a concrete method for solving this difficult optimal control problem with control-dependent time-delayed arguments.

Thus the contribution of this paper is twofold. From the practical point of view, this is a pioneering paper, which provides an open-loop control for the time-delayed optimal transmission problem of the computer network; in this way, we can allocate the optimal flow rate to each link of the real-life computer network so that the maximum amount of messages can be sent from N origins to one destination in a given period. From the mathematical point of view, this pioneering paper develops a modified control parameterization method for solving time-delayed problems with both control-dependent time-delayed arguments and discrete time-delayed arguments. Thus this pioneering paper also provides a concrete method for solving the time-delayed optimal transmission problem of the computer network. Numerical results clearly illustrate the efficiency of this modified control parameterization method for solving real-life optimal control problems with control-dependent time-delayed arguments.

In Sect. 2, we provide a formulation of the time-delayed optimal transmission problem of the computer network. In Sect. 3, we convert the time-delayed optimal control problem into a canonical form so that it can be solved by a modified control parameterization method. In Sect. 4, we describe the modified control parameterization method for solving the time-delayed control problem discussed in Sects. 2 and 3. In Sect. 5, we derive the gradient formulae for solving the time-delayed optimal control problem by the modified control parameterization method. In Sect. 6, we solve a numerical example consisting of four sources, three links, two origins, and one destination; each source uses two links to travel from its origin to its destination. We solve this example using both the nondelayed version and the time-delayed version of the optimal transmission problem. The concluding remarks and suggestions for further study are given in Sect. 7.

Figure 1 The structure of a computer network that has 2 origins, N sources, 3 links, and 1 destination



2 Formulation of the time-delayed optimal transmission problem

2.1 Description of the computer network

We consider a computer network that has two origins, N sources, three links, and one destination, such that sources s_1, \dots, s_K use links l_1 and l_3 to reach the destination and sources s_{K+1}, \dots, s_N use links l_2 and l_3 to reach the destination. (See Fig. 1 for details.)

Let $\tilde{u}_1(t), \dots, \tilde{u}_K(t)$ be the input rates (bits per second) associated with sources s_1, \dots, s_K at the buffer of link l_1 at time t . Let $\tilde{u}_{K+1}(t), \dots, \tilde{u}_N(t)$ be the input rates (bits per second) associated with sources s_{K+1}, \dots, s_N at the buffer of link l_2 at time t . Let $\hat{u}_1(t), \dots, \hat{u}_N(t)$ be the input rates (bits per second) associated with sources s_1, \dots, s_N at the buffer of link l_3 at time t . Then $\tilde{u}_1(t), \dots, \tilde{u}_N(t)$ and $\hat{u}_1(t), \dots, \hat{u}_N(t)$ are controls of this problem. (In fact, $\hat{u}_1(t), \dots, \hat{u}_N(t)$ are artificial controls only, because their values are completely determined as long as the input rates $\tilde{u}_1(t), \dots, \tilde{u}_N(t)$ and the queue sizes (to be defined later) at links l_1 and l_2 are given.)

For each $i = 1, \dots, K$, let $b_{i,1}(t)$ be the queue length at the buffer of link l_1 associated with source s_i at time t . For each $i = K + 1, \dots, N$, let $b_{i,2}(t)$ be the queue length at the buffer of link l_2 associated with source s_i at time t . For each $i = 1, \dots, N$, let $b_{i,3}(t)$ be the queue length at the buffer of link l_3 associated with source s_i at time t . Let $b_1(t) = \sum_{i=1}^K b_{i,1}(t)$, $b_2(t) = \sum_{i=K+1}^N b_{i,2}(t)$, and $b_3(t) = \sum_{i=1}^N b_{i,3}(t)$ be, respectively, the queue sizes associated with all the sources at links l_1 , l_2 , and l_3 at time t . Let c_j ($j = 1, 2, 3$) (bits per second) be the link capacity (bandwidth) of link l_j .

2.2 Preliminary formulation of the buffer (state) equation and the conservation of flow equation without considering the time-delayed arguments

In view of the computer network described in the previous section, we have the following buffer equation from [12]:

$$\dot{b}_1(t) = \left(\sum_{i=1}^K \tilde{u}_i(t) - c_1 \right)_{b_1(t)}^+, \quad (3)$$

$$\dot{b}_2(t) = \left(\sum_{i=K+1}^N \tilde{u}_i(t) - c_2 \right)_{b_2(t)}^+, \quad (4)$$

$$\dot{b}_3(t) = \left(\sum_{i=1}^N \hat{u}_i(t) - c_3 \right)_{b_3(t)}^+, \quad (5)$$

with the initial condition

$$b_j(0) = 0, \quad j = 1, \dots, 3, \quad (6)$$

where

$$(x)_y^+ = \begin{cases} x & \text{if } y > 0, \\ \max(x, 0) & \text{if } y \leq 0. \end{cases} \quad (7)$$

Remark 2.1 The explanation of Eqs. (3)–(5) is as follows: If the queue length (backlog) at time t is positive, then the instantaneous rate of change of the queue length is equal to the total incoming source rate minus the capacity of the link; if the queue length (backlog) at time t is zero, the rate of change should be equal to the maximum between “total incoming source rate minus capacity” and zero, because in real life the queue length cannot be negative.

Since the buffer of any link is shared by more than one source (i.e., the buffers of links l_1 , l_2 , and l_3 are shared by sources s_1, \dots, s_K , sources s_{K+1}, \dots, s_N , and sources of all the links s_1, \dots, s_N , respectively), we wish to specify at which rates the different flows leave the buffer. Let $\theta_{i,j}(t)$ be the ratio of the queue rate due to source s_i at the buffer of link l_j to the total queue rate at the buffer of link l_j . Then the equations for $b_{i,j}(t)$ are as follows:

$$\dot{b}_{i,1}(t) = \theta_{i,1}(t) \dot{b}_1(t), \quad i = 1, \dots, K, t \in [0, T], \quad (8)$$

$$\dot{b}_{i,2}(t) = \theta_{i,2}(t) \dot{b}_2(t), \quad i = K + 1, \dots, N, t \in [0, T], \quad (9)$$

$$\dot{b}_{i,3}(t) = \theta_{i,3}(t) \dot{b}_3(t), \quad i = 1, \dots, N, t \in [0, T], \quad (10)$$

with the initial conditions

$$b_{i,1}(0) = 0, \quad i = 1, \dots, K, \quad (11)$$

$$b_{i,2}(0) = 0, \quad i = K + 1, \dots, N, \quad (12)$$

$$b_{i,3}(0) = 0, \quad i = 1, \dots, N, \quad (13)$$

where

$$\theta_{i,1}(t) = \frac{\tilde{u}_i(t)}{\sum_{i=1}^K \tilde{u}_i(t)}, \quad i = 1, \dots, K, \quad (14)$$

$$\theta_{i,2}(t) = \frac{\tilde{u}_i(t)}{\sum_{i=K+1}^N \tilde{u}_i(t)}, \quad i = K + 1, \dots, N, \quad (15)$$

$$\theta_{i,3}(t) = \frac{\hat{u}_i(t)}{\sum_{i=1}^N \hat{u}_i(t)}, \quad i = 1, \dots, N \quad (16)$$

are the ratios assigned in accordance with the WFQ service discipline described in [12].

Moreover, we need to impose the following control bounds:

$$\bar{u} \geq \tilde{u}_i(t), \quad \hat{u}_i(t) \geq \underline{u}, \quad i = 1, \dots, N, t \in [0, T]. \quad (17)$$

Due to the conservation of TCP flow from the buffer of one link to that of the next link, we have

$$\hat{u}_i(t) = \begin{cases} \tilde{u}_i(t) - \dot{b}_{i,1}(t), & i = 1, \dots, K, \\ \tilde{u}_i(t) - \dot{b}_{i,2}(t), & i = K + 1, \dots, N. \end{cases} \quad (18)$$

The above equation is called the conservation of flow equations.

Remark 2.2 By Remark 2.1, it is possible that the input rate associated with source s_i at the buffer of one link is larger than that at the buffer of the previous link. (In other words, $\hat{u}_i(t)$ can be larger than $\tilde{u}_i(t)$.)

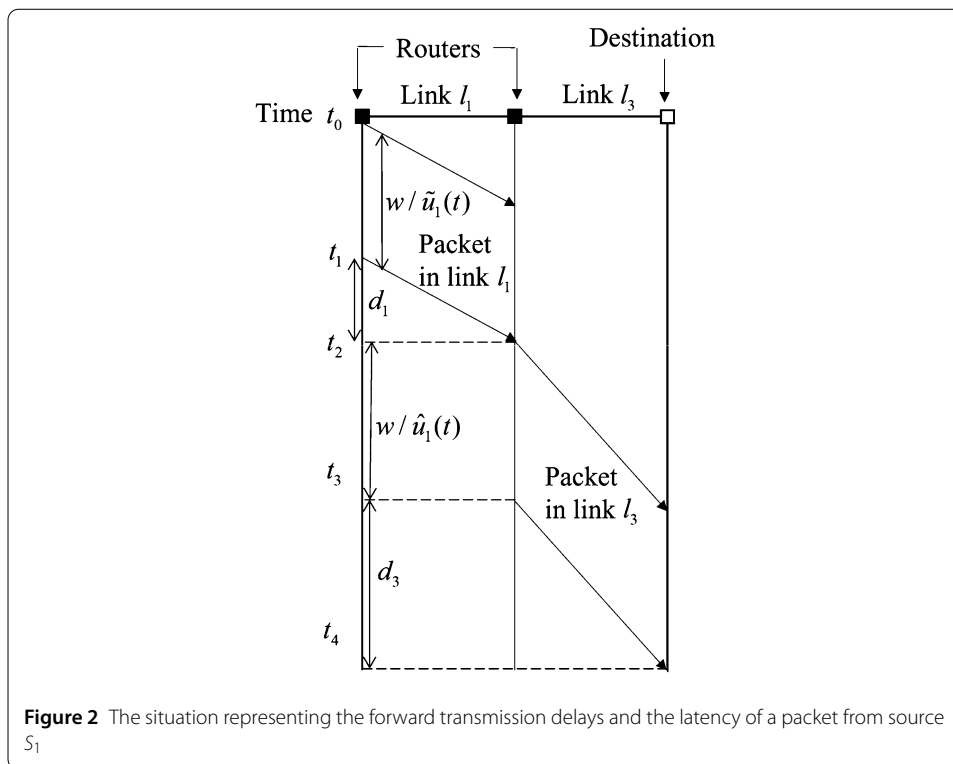
2.3 Formulation of the time-delayed optimal control problem when forward transmission delays and latency of the links are taken into consideration

As mentioned in the introduction, the buffer equation and conservation of flow equation described in Sect. 2.1 ignore the existence of time delays during the transmission from the routers to links. Moreover, the conservation of the TCP flow equations ignores the existence of the latency of the link, which is equal to the difference between the time that a packet leaves the router of one link and the time that it arrives at the receiving end of the link, i.e., the time that the packet arrives at the router of the preceding link during its trip from source to destination. Thus (2)–(4) do not reflect the evolution of the buffer size accurately, and (17) does not reflect the conservation of TCP flow accurately. Because of this weakness, we follow the approach of [4] to change the buffer equations and the conservation of the TCP flow equations so that they do not ignore the time-delayed arguments. The situation representing the forward transmission delays and the latency is depicted in Fig. 2.

In Fig. 2, we use the following notations:

- t_0 : The time that the packet from sources s_1 is being transmitted from its origin
- t_1 : The time that the packet serializes the packet onto link l_1 from its router
- t_2 : The time that the packet arrives at the receiving end of link l_1 (i.e., the time it takes for the packet to arrive at the router of link l_3 , waiting to be serialized into link l_3)
- t_3 : The time that the packet serializes the packet onto l_3 from its router
- t_4 : The time that the packet arrives at the receiving end of link l_3 (i.e., the time it takes for the packet to arrive at the destination)
- \tilde{u}_1 : The transmission rate from source s_1 at the time it reaches the router of link l_1
- \hat{u}_1 : The transmission rate from source s_1 at the time it reaches the router of link l_3
- $w/\tilde{u}_1(t)$: Forward transmission delay of a packet with size w using the transmission rate $\tilde{u}_1(t)$.
- $(w/\hat{u}_1(t))$ can be defined in a similar way.)
- d_1 : Latency of link l_1 , which is the difference between the time that the packet leaves link l_1 and the time that it reaches the receiving end of link l_1 .
- (d_3) can be defined in a similar way.)

Suppose, on average, the router transmits a packet of constant size w_1 (bits) at any time t to link l_1 using the transmitting rate $\tilde{u}_1(t)$ (bits per second). Then the forward transmission delay is given by $w_1/\tilde{u}_1(t)$ (seconds) (i.e., it is directly proportional to the size of the packet but inversely proportional to the packet transmission rate.) In other words, the packet has



to queue at the router for $w_1/\tilde{u}_1(t)$ (seconds) before it can be serialized into link l_1 . For simplicity, we assume that the packet size inputted to each of the links l_1 , l_2 , and l_3 at any time t is w (bits).

The phenomenon of the forward transmission delay and latency in a computer network can be explained as follows: Suppose that we wish to send goods from point A to point B. There is a large conveyor belt for transferring goods from point A to point B, and any item placed on the conveyor belt will move at a constant speed of u items per second. The distance between A and B is y meters. Using both hands, we can transfer the goods from the carton to the belt at a constant speed of w items per second. Then the forward transmission delay for the transportation of goods is (w/u) meters/second, and the latency is $d = y/u$ seconds. Thus in a computer network the forward transmission delay is the time difference between the time that the packet first arrives at the router of a link (waiting for the router to put it into the belt) and the time that the router serializes it onto the link.

Thus by inserting the transmission delay into the buffer Eqs. (3)–(6) we modify the buffer equations as follows:

$$\dot{b}_1(t) = \left(\sum_{i=1}^K \tilde{u}_i(t - w/\tilde{u}_i(t)) - c_1 \right)_{b_1(t)}^+, \quad t \in [0, T], \quad (19)$$

$$\dot{b}_2(t) = \left(\sum_{i=K+1}^N \tilde{u}_i(t - w/\tilde{u}_i(t)) - c_2 \right)_{b_2(t)}^+, \quad t \in [0, T], \quad (20)$$

$$\dot{b}_3(t) = \left(\sum_{i=1}^N \hat{u}_i(t - w/\hat{u}_i(t)) - c_3 \right)_{b_3(t)}^+, \quad t \in [0, T], \quad (21)$$

with the following initial conditions on the queue size and the input transmission rates:

$$b_j(t) = 0, \quad j = 1, 2, 3, \quad (22)$$

$$\tilde{u}_i(t - w/\tilde{u}_i(t)) = 0 \quad \text{whenever } t - w/\tilde{u}_i(t) < 0, i = 1, \dots, N, t \in [0, T], \quad (23)$$

$$\hat{u}_i(t - w/\hat{u}_i(t)) = 0 \quad \text{whenever } t - w/\hat{u}_i(t) < 0, i = 1, \dots, N, t \in [0, T], \quad (24)$$

and

$$\dot{b}_{i,1}(t) = \frac{\tilde{u}_i(t - w/\tilde{u}_i(t))}{\sum_{j=1}^K \tilde{u}_j(t - w/\tilde{u}_j(t)) + \varepsilon} \dot{b}_1(t), \quad i = 1, \dots, K, t \in [0, T], \quad (25)$$

$$\dot{b}_{i,2}(t) = \frac{\tilde{u}_i(t - w/\tilde{u}_i(t))}{\sum_{j=K+1}^N \tilde{u}_j(t - w/\tilde{u}_j(t)) + \varepsilon} \dot{b}_2(t), \quad i = K + 1, \dots, N, t \in [0, T], \quad (26)$$

$$\dot{b}_{i,3}(t) = \frac{\hat{u}_i(t - w/\hat{u}_i(t))}{\sum_{j=1}^N \hat{u}_j(t - w/\hat{u}_j(t)) + \varepsilon} \dot{b}_3(t), \quad i = 1, \dots, N, t \in [0, T], \quad (27)$$

$$b_{i,1}(0) = 0, \quad i = 1, \dots, K, \quad (28)$$

$$b_{i,2}(0) = 0, \quad i = k + 1, \dots, N, \quad (29)$$

$$b_{i,3}(0) = 0, \quad i = 1, \dots, N, \quad (30)$$

together with initial conditions on the input transmission rates given by (22)–(24), where ε is a very small number, and the notation $(x)_y^+$ is defined in (7).

To ensure that the state Eqs. (19)–(21) and (25)–(27) are well defined, we need to impose the bounds on the input transmission rates $\tilde{u}_i(t)$, $\hat{u}_i(t)$:

$$\bar{u} \geq \tilde{u}_i(t), \quad \hat{u}_i(t) \geq \underline{u}, \quad i = 1, \dots, N, t \in [0, T]. \quad (31)$$

Moreover, we need to impose the buffer capacity constraints:

$$0 \leq b_j(t) \leq B_j, \quad j = 1, 2, 3. \quad (32)$$

Remark 2.3 Eqs. (19)–(21) state that the change of queue size at the buffer of each link occurs at the moment that the packet is being serialized into the link, instead of the time that it first arrives at the router of the link. Similarly, Eqs. (25)–(27) state that the allocation of queue size at the buffer of each link actually occurs when the packet is being serialized into the link, instead of the time that it first arrives at the router of the link. The reason for adding ε to the denominator of (25)–(27) is to ensure that the denominators of the terms of these equations is nonzero when each of the terms $\tilde{u}_j(t - w/\tilde{u}_j(t))$ or $\hat{u}_j(t - w/\hat{u}_j(t))$ is zero.

Moreover, as mentioned in [4], there also exists a time difference between the time that the packet leaves the router of one link (i.e., immediately after the router has serialized the packet into the link) and the time that it arrives at the receiving end of the link. This is called the latency of the link, which is time-independent. For simplicity, we assume that the latency of each link is the same. Let d (seconds) be the latency of the link. Thus

by inserting the latency of the links into the conversation of the TCP flow equations we modify the conservation of TCP flow Eq. (16) as follows:

$$\hat{u}_i(t) = \begin{cases} \tilde{u}_i(t-d) - \dot{b}_{i,1}(t-d), & i = 1, \dots, K, t \in [0, T], \\ \tilde{u}_i(t-d) - \dot{b}_{i,2}(t-d), & i = K+1, \dots, N, t \in [0, T], \end{cases} \quad (33)$$

where

$$\tilde{u}_i(t-d) = 0 \quad \text{whenever } t-d \leq 0, i = 1, \dots, N, t \in [0, T]. \quad (34)$$

Remark 2.4 Eq. (33) states that the input transmission rate at the router of link 3 at time t due to any source is equal to the output transmission rate at the router of the previous link (i.e., either link 1 or link 2) due to this source at time $t-d$.

Thus, in view of (33), and (25)–(27), we can express the input transmission rate at the router of link 3 as a function of the delayed input transmission rates due to the forward transmission delay at the buffer of the previous link (i.e., either link 1 or link 2) and the latency of the previous link (i.e., either link 1 or link 2) as follows:

$$\hat{u}_i(t) = \begin{cases} \tilde{u}_i(t-d) - \left(\frac{\tilde{u}_i(t-d-w/\tilde{u}_i(t))}{\sum_{j=1}^K \tilde{u}_j(t-d-w/\tilde{u}_j(t)) + \varepsilon} \right) \\ \quad \times \sum_{j=1}^K (\tilde{u}_j(t-d-w/\tilde{u}_j(t)) - c_1)_{b_1(t)}^+, \\ \quad i = 1, \dots, K, t \in [0, T], \\ \tilde{u}_i(t-d) - \left(\frac{\tilde{u}_i(t-d-w/\tilde{u}_i(t))}{\sum_{j=K+1}^N \tilde{u}_j(t-d-w/\tilde{u}_j(t)) + \varepsilon} \right) \\ \quad \times \sum_{j=K+1}^N (\tilde{u}_j(t-d-w/\tilde{u}_j(t)) - c_2)_{b_2(t)}^+, \\ \quad i = K+1, \dots, N, t \in [0, T], \end{cases} \quad (35)$$

where

$$\tilde{u}_i(t-d) = 0 \quad \text{whenever } t-d \leq 0, i = 1, \dots, N, t \in [0, T], \quad (36)$$

$$\tilde{u}_i(t-d-w/\tilde{u}_i(t)) = 0 \quad \text{whenever } t-d-w/\tilde{u}_i(t) \leq 0, i = 1, \dots, N, t \in [0, T]. \quad (37)$$

Let $U(u)$ be the utility function of our problem, which represents the benefit of the entire network when the source s_i , $i = 1, \dots, N$, is transmitting data at the rate $\tilde{u}_i(t)$. Now the output flow rate of source s_i , $i = 1, \dots, N$, at the receiving end of link l_3 is given by $\hat{u}_i(t-d) - \dot{b}_{i,3}(t-d)$. Thus by maximizing the weighted source output in accordance with the relative importance of the message sent by each source we obtain

$$\begin{aligned} \max U(u) &= \sum_{i=1}^N M_i \int_0^T [\hat{u}_i(t-d) - \dot{b}_{i,3}(t-d)] dt \\ &= \sum_{i=1}^N M_i \int_0^T \left[\hat{u}_i(t-d) - \left(\frac{\hat{u}_i(t-d-w/\hat{u}_i(t))}{\sum_{j=1}^N \hat{u}_j(t-d-w/\hat{u}_j(t)) + \varepsilon} \right) \right. \\ &\quad \left. \times \sum_{j=1}^N (\hat{u}_j(t-d-w/\hat{u}_j(t)) - c_3)_{b_3(t)}^+ \right] dt, \end{aligned} \quad (38)$$

where

$$\hat{u}_i(t-d) = 0 \quad \text{whenever } t-d \leq 0, i = 1, \dots, N, t \in [0, T], \quad (39)$$

$$\hat{u}_i(t-d-w/\hat{u}_i(t)) = 0 \quad \text{whenever } t-d-w/\hat{u}_i(t) \leq 0, i = 1, \dots, N, t \in [0, T], \quad (40)$$

and M_i represents the relative importance of the messages sent from source s_i to the destination.

The time-delayed optimal control problem can be stated as follows.

Problem (P1) Subject to the buffer (state) Eqs. (19)–(21) with initial conditions (22)–(24), the conservation of TCP flow Eq. (35) with initial conditions (34), (36), and (37), the bounds on the input transmission rate (31) and the buffer capacity constraints (32), we want to find piecewise continuous controls $\tilde{u}_i(t)$, $i = 1, \dots, N$, that maximize the objective function (38).

3 Obtaining a new time-delayed optimal control problem with smooth functions and smooth canonical constraints

The time-delayed optimal control problem described in Sect. 2 involves nonsmooth functions and nonsmooth constraints, which cannot be easily solved by computational methods. In this section, we describe a method for converting all the nonsmooth functions and nonsmooth constraints into smooth functions and smooth canonical constraints. The details are given in the next subsections.

3.1 Converting the nonsmooth functions in the buffer equations and the objective function into smooth functions

We first need to use an approximation method to convert the nonsmooth function in the right-hand side of the buffer Eqs. (19)–(21), the conservation of the TCP flow Eq. (35), and the objective function (38) into sufficiently smooth functions.

Noting that all these nonsmooth functions have the form $(z(t))_{y(t)}^+$, we can approximate $(z(t))_{y(t)}^+$ by a sufficiently smooth function $\xi_\delta(y(t), z(t)) \in C^1(R^2)$ as follows:

$$\xi_\delta(y(t), z(t)) = \begin{cases} z(t) & \text{if } y(t) > 0, \\ I_\delta(y(t))z(t) + (1 - I_\delta(y(t)))\max_\delta(z(t)) & \text{if } -\delta \leq y(t) \leq 0, \\ \max_\delta(z(t)) & \text{if } y(t) < -\delta, \end{cases} \quad (41)$$

where $\max_\delta(z(t))$ is the function used for smoothing $\max\{z(t), 0\}$ defined by

$$\max_\delta(z(t)) = \begin{cases} 0 & \text{if } z(t) \leq -\delta, \\ \frac{(z(t)+\delta)^2}{4\delta} & \text{if } -\delta \leq z(t) \leq \delta, \\ z(t) & \text{if } z(t) > \delta, \end{cases} \quad (42)$$

and

$$I_\delta(y(t)) = -2\left(\frac{y(t)}{\delta}\right)^3 + 3\left(\frac{y(t)}{\delta}\right)^2 + 1 \quad \text{if } -\delta \leq y(t) \leq 0 \quad (43)$$

is a real number between 0 and 1, and δ is a small number.

Using this approximation method, from (19)–(21) we obtain the new state equations

$$\dot{b}_1(t) = \xi_\delta \left(b_1(t), \sum_{j=1}^K \tilde{u}_j(t - w/\tilde{u}_j(t)) - c_1 \right), \quad t \in [0, T], \quad (44)$$

$$\dot{b}_2(t) = \xi_\delta \left(b_2(t), \sum_{j=K+1}^N \tilde{u}_j(t - w/\tilde{u}_j(t)) - c_2 \right), \quad t \in [0, T], \quad (45)$$

$$\dot{b}_3(t) = \xi_\delta \left(b_3(t), \sum_{j=1}^N \hat{u}_j(t - w/\hat{u}_j(t)) - c_3 \right), \quad t \in [0, T]. \quad (46)$$

Similarly, by using this approximation method from (38) we obtain the new objective function

$$\begin{aligned} \max U(u) = & \sum_{i=1}^N M_i \int_0^T \left[\hat{u}_i(t - d) - \left(\frac{\hat{u}_i(t - d - w/\hat{u}_i(t))}{\sum_{j=1}^N \hat{u}_j(t - d - w/\hat{u}_j(t)) + \varepsilon} \right) \right. \\ & \left. \times \xi_\varepsilon \left(b_3(t), \sum_{i=1}^N \hat{u}_i(t - d - w/\hat{u}_i(t)) - c_3 \right) \right] dt. \end{aligned} \quad (47)$$

3.2 Converting the nonsmooth constraint functions into smooth canonical functions

Similarly, by using the above approximation method together with the constraint approximation method described in [19–21] we convert the new conservation of the TCP flow Eq. (35) into smooth canonical constraints

$$\begin{aligned} \int_0^T & \left[G_i(b_1(t), \hat{u}_i(t), \tilde{u}_i(t - d), \tilde{u}_1(t - d - w/\tilde{u}_1(t)), \dots, \tilde{u}_K(t - d - w/\tilde{u}_K(t))) \right]^2 dt \\ & = 0 \quad \text{for } i = 1, \dots, K, \end{aligned} \quad (48)$$

$$\begin{aligned} \int_0^T & \left[G_i(b_2(t), \hat{u}_i(t), \tilde{u}_i(t - d), \tilde{u}_{K+1}(t - d - w/\tilde{u}_{K+1}(t)), \dots, \tilde{u}_N(t - d - w/\tilde{u}_N(t))) \right]^2 dt \\ & = 0 \quad \text{for } i = K + 1, \dots, N, \end{aligned} \quad (49)$$

where

$$\begin{aligned} & G_i(b_1(t), \hat{u}_i(t), \tilde{u}_i(t - d), \tilde{u}_1(t - d - w/\tilde{u}_1(t)), \dots, \tilde{u}_K(t - d - w/\tilde{u}_K(t))) \\ & = \hat{u}_i(t) - \tilde{u}_i(t - d) + \left(\frac{\tilde{u}_i(t - d - w/\tilde{u}_i(t))}{\sum_{j=1}^K \tilde{u}_j(t - d - w/\tilde{u}_j(t)) + \varepsilon} \right) \\ & \quad \times \delta_\varepsilon \left(b_1(t), \sum_{j=1}^K \tilde{u}_j(t - d - w/\tilde{u}_j(t)) - c_1 \right), \quad i = 1, \dots, K, \end{aligned} \quad (50)$$

$$\begin{aligned} & G_i(b_2(t), \hat{u}_i(t), \tilde{u}_i(t - d), \tilde{u}_{K+1}(t - d - w/\tilde{u}_{K+1}(t)), \dots, \tilde{u}_N(t - d - w/\tilde{u}_N(t))) \\ & = \hat{u}_i(t) - \tilde{u}_i(t - d) + \left(\frac{\tilde{u}_i(t - d - w/\tilde{u}_i(t))}{\sum_{j=K+1}^N \tilde{u}_j(t - d - w/\tilde{u}_j(t)) + \varepsilon} \right) \end{aligned}$$

$$\times \delta_\varepsilon \left(b_2(t), \sum_{j=K+1}^N \tilde{u}_j(t-d-w/\tilde{u}_j(t)) - c_2 \right), \quad i = K+1, \dots, N. \quad (51)$$

Furthermore, by using the constraint approximation method described in [19–21] we convert the new buffer capacity constraints (32) into smooth canonical constraints

$$\int_0^T \bar{G}_i(b(t)) = 0, \quad i = 1, 2, 3, 4, 5, 6, \quad (52)$$

where

$$\bar{G}_i(b(t)) = \begin{cases} \hat{G}_i(b(t)), & \text{if } \hat{G}_i(b(t)) \leq -\varepsilon, \\ -\frac{(\hat{G}_i(b(t)) - \varepsilon)^2}{4\varepsilon}, & \text{if } -\varepsilon \leq \hat{G}_i(b(t)) \leq \varepsilon, \\ 0, & \text{if } \hat{G}_i(b(t)) \geq \varepsilon, \end{cases} \quad (53)$$

$$\hat{G}(b(t)) = (b_1(t), b_2(t), b_3(t), B_1 - b_1(t), B_2 - b_2(t), B_3 - b_3(t))^T. \quad (54)$$

3.3 Obtaining a smooth time-delayed optimal control problem

With the preparation work given in the previous subsections, we are able to convert problem (P1) into a new smooth time-delayed optimal control problem (P2), which can be solved by a modified control parameterization method given in the next section. Problem (P2) is as follows.

Problem (P2) Subject to the buffer Eqs. (44)–(46) with initial conditions (23)–(25), the conservation of the TCP flow constraints in canonical form (48)–(49) with initial conditions (34), (36), and (37), the bounds on the input transmission rate (32), and the buffer capacity constraints in canonical form (52), we want to find piecewise continuous controls $\tilde{u}_i(t)$, $i = 1, \dots, N$, which maximize the objective function (47).

4 Solving the optimal control problem with control and state-dependent time delays by the control parametrization technique

In this section, we use the control parameterization technique to solve problem (P2). In the past, the control parameterization method was used to solve optimal control problems without any time-delayed arguments or time-delayed optimal control problems with discrete time-delayed arguments only. In this paper, we extend the control parameterization method to solve a more difficult optimal control problem, namely, an optimal control problem with control-dependent time-delayed arguments. We first create an approximated problem for problem (P2) in such a way that it does not contain any control-dependent time-delayed argument. The details are as follows.

Define

$$U = \{v = (v_1, \dots, v_{2N})^T : \underline{u} \leq v_i \leq \bar{u}, i = 1, \dots, 2N\}. \quad (55)$$

A piecewise continuous function $u(t) = (\tilde{u}_1(t), \dots, \tilde{u}_N(t), \hat{u}_1(t), \dots, \hat{u}_N(t))^T$ from $[0, T]$ into R^{2N} is said to be an admissible control of problem (P2) if $u(t) \in U$ for all $t \in [0, T]$. Let \mathcal{U} be the set consisting of all such admissible controls.

We now partition the time horizon $[0, T]$ into p equal subintervals so that the component of the admissible control $u(t)$ is a constant over each subinterval. Let I^p be the partition of $[0, T]$ defined by

$$I^p = \{(k-1)\Delta, k\Delta) : k = 1, \dots, p\}, \quad (56)$$

where $\Delta = T/p$ is the length of each subinterval.

Let \mathcal{U}^p be the subset of \mathcal{U} consisting of all the piecewise constant controls consistent with the partition I^p . Let $u(t) \in \mathcal{U}^p$. Then each component of $u(t)$ assumes the form

$$\tilde{u}_i(t) = \begin{cases} \sum_{k=1}^p \tilde{\sigma}_i^k \chi_{I_k^p}(t), & i = 1, \dots, N, t \in [0, T], \\ 0, & i = 1, \dots, N, t \leq 0, \end{cases} \quad (57)$$

$$\hat{u}_i(t) = \begin{cases} \sum_{k=1}^p \hat{\sigma}_i^k \chi_{I_k^p}(t), & i = 1, \dots, N, t \in [0, T], \\ 0, & i = 1, \dots, N, t \leq 0, \end{cases} \quad (58)$$

where χ_I denotes the indicator function of I defined by

$$\chi_I(t) = \begin{cases} 1 & \text{if } t \in I, \\ 0 & \text{otherwise.} \end{cases} \quad (59)$$

Let

$$\sigma^k = [\tilde{\sigma}_1^k, \dots, \tilde{\sigma}_N^k, \hat{\sigma}_1^k, \dots, \hat{\sigma}_N^k]^T \quad (60)$$

and

$$\sigma = [\sigma^1, \dots, \sigma^p]^T \in \mathbb{R}^{2NP}. \quad (61)$$

Restricting to \mathcal{U}^p , the control constraints (i.e., the bounds on the input transmission rates), (31) becomes

$$\underline{u} \leq \tilde{\sigma}_i^j, \quad \hat{\sigma}_i^j \leq \bar{u}, \quad i = 1, \dots, N, j = 1, \dots, p. \quad (62)$$

Let Ξ^p be the set of all vectors σ that satisfy constraint (62). Then each control $u \in \mathcal{U}^p$ can be uniquely determined by a control parameter vector $\sigma \in \Xi^p$ and vice versa. Let $\hat{u}(t|\sigma) \in \mathcal{U}^p$ be the control completely specified by the vector $\sigma \in \Xi^p$.

Suppose that the length of each partition T/p is sufficiently large so that the lower bound \underline{u} of the control satisfies the inequality

$$0 \leq \frac{w}{\underline{u}} + d < \frac{T}{p}. \quad (63)$$

In other words, the duration from one switching time to the next switching time of the control $u \in \mathcal{U}^p$ should be greater than its largest time delay (i.e., the forward transmission

delay plus the latency) at any time $t \in [0, T]$. (By (31) the maximum possible value of “forward” transmission delay plus latency is $\frac{w}{u} + d$.) In view of (63) and the initial control conditions (23)–(24), by identifying the control $\hat{u} \in \mathcal{U}^p$ with parameter $\sigma \in \Xi^p$, we can convert the control-dependent time-delayed term $\tilde{u}_i(t - w/\tilde{u}_i(t))$ and $\hat{u}_i(t - w/\hat{u}_i(t))$, $i = 1, \dots, N$, in (44)–(46) into $\tilde{\tilde{u}}_i(t|\sigma)$ and $\hat{\hat{u}}_i(t|\sigma)$, respectively, where

$$\begin{aligned}\tilde{\tilde{u}}_i(t|\sigma) &= \tilde{u}_i(t - (w/\tilde{u}_i(t))) \\ &= \begin{cases} 0 & \text{if } t \in [0, w/\tilde{\sigma}_i^1), \\ \tilde{\sigma}_i^k & \text{if } t \in [w/\tilde{\sigma}_i^k + (k-1)T/p, w/\tilde{\sigma}_i^{k+1} + kT/p), k = 1, \dots, p-1, \\ \tilde{\sigma}_i^p & \text{if } t \in [w/\tilde{\sigma}_i^p + (p-1)T/p, T), \end{cases} \quad (64)\end{aligned}$$

$$\begin{aligned}\hat{\hat{u}}_i(t|\sigma) &= \hat{u}_i(t - (w/\hat{u}_i(t))) \\ &= \begin{cases} 0 & \text{if } t \in [0, w/\hat{\sigma}_i^1), \\ \hat{\sigma}_i^k & \text{if } t \in [w/\hat{\sigma}_i^k + (k-1)T/p, w/\hat{\sigma}_i^{k+1} + kT/p), k = 1, \dots, p-1, \\ \hat{\sigma}_i^p & \text{if } t \in [w/\hat{\sigma}_i^p + (p-1)T/p, T). \end{cases} \quad (65)\end{aligned}$$

Similarly, in view of (63) and the initial conditions (36)–(37), by identifying the control $\hat{u} \in \mathcal{U}^p$ with parameter $\sigma \in \Xi^p$ we can convert the control-dependent time-delayed term $\tilde{u}_i(t - d - (w/\tilde{u}_i(t)))$ and $\hat{u}_i(t - d - (w/\hat{u}_i(t)))$, $i = 1, \dots, N$, in (48)–(49) into $\tilde{\tilde{\tilde{u}}}_i(t|\sigma)$ and $\hat{\hat{\hat{u}}}_i(t|\sigma)$, respectively, where

$$\begin{aligned}\tilde{\tilde{\tilde{u}}}_i(t|\sigma) &= \tilde{u}_i(t - d - (w/\tilde{u}_i(t))) \\ &= \begin{cases} 0 & \text{if } t \in [0, w/\tilde{\sigma}_i^1 + d), \\ \tilde{\sigma}_i^k & \text{if } t \in [w/\tilde{\sigma}_i^k + d + (k-1)T/p, w/\tilde{\sigma}_i^{k+1} + d + kT/p), \\ & k = 1, \dots, p-1, \\ \tilde{\sigma}_i^p & \text{if } t \in [w/\tilde{\sigma}_i^p + d + (p-1)T/p, T), \end{cases} \quad (66)\end{aligned}$$

$$\begin{aligned}\hat{\hat{\hat{u}}}_i(t|\sigma) &= \hat{u}_i(t - d - (w/\hat{u}_i(t))) \\ &= \begin{cases} 0 & \text{if } t \in [0, w/\hat{\sigma}_i^1 + d), \\ \hat{\sigma}_i^k & \text{if } t \in [w/\hat{\sigma}_i^k + d + (k-1)T/p, w/\hat{\sigma}_i^{k+1} + d + kT/p), \\ & k = 1, \dots, p-1, \\ \hat{\sigma}_i^p & \text{if } t \in [w/\hat{\sigma}_i^p + d + (p-1)T/p, T). \end{cases} \quad (67)\end{aligned}$$

Let $b(t|\sigma) \in \mathbb{R}^3$ be the solution of the state (buffer) Eqs. (44)–(46) and the initial conditions (22)–(24) corresponding to $\sigma \in \Xi^p$. When we identify the control $u(t) \in \mathcal{U}^p$ with the parameter $\sigma \in \Xi^p$, we obtain the new state equations as follows:

$$\dot{b}_1(t|\sigma) = F_1(b_1(t|\sigma), \tilde{\tilde{u}}_1(t|\sigma), \dots, \tilde{\tilde{u}}_N(t|\sigma)), \quad t \in [0, T], \quad (68)$$

$$\dot{b}_2(t|\sigma) = F_2(b_2(t|\sigma), \tilde{\tilde{u}}_{K+1}(t|\sigma), \dots, \tilde{\tilde{u}}_N(t|\sigma)), \quad t \in [0, T], \quad (69)$$

$$\dot{b}_3(t|\sigma) = F_3(b_3(t|\sigma), \hat{\hat{u}}_1(t|\sigma), \dots, \hat{\hat{u}}_N(t|\sigma), b_3(t|\sigma)), \quad t \in [0, T], \quad (70)$$

$$b_i(0|\sigma) = 0, \quad i = 1, 2, 3, \quad (71)$$

where $F_1(b_1(t|\sigma), \tilde{u}_1(t|\sigma), \dots, \tilde{u}_N(t|\sigma))$, $F_2(b_2(t|\sigma), \tilde{u}_1(t|\sigma), \dots, \tilde{u}_N(t|\sigma))$, and $F_3(b_3(t|\sigma), \hat{u}_1(t|\sigma), \dots, \hat{u}_N(t|\sigma))$ can be obtained from the functions in the right-hand side of (44), (45), and (46) in an obvious manner.

Similarly, we obtain from (66) and (67) the new canonical constraints concerning the conversation of flow equation as follows:

$$\int_0^T \bar{G}_i(b_1(t|\sigma), \hat{u}_i(t|\sigma), \tilde{u}_i(t-d|\sigma), \tilde{u}_1(t|\sigma), \dots, \tilde{u}_K(t|\sigma)) = 0, \quad i = 1, \dots, K, \quad (72)$$

$$\int_0^T \bar{G}_i(b_2(t|\sigma), \hat{u}_i(t|\sigma), \tilde{u}_i(t-d|\sigma), \tilde{u}_{K+1}(t|\sigma), \dots, \tilde{u}_N(t|\sigma)) = 0, \\ i = K + 1, \dots, N, \quad (73)$$

where $\bar{G}_i(b_1(t|\sigma), \hat{u}_i(t|\sigma), \tilde{u}_i(t-d|\sigma), \tilde{u}_1(t|\sigma), \dots, \tilde{u}_K(t|\sigma))$ and $\bar{G}_i(b_1(t|\sigma), \hat{u}_i(t|\sigma), \tilde{u}_i(t-d|\sigma), \tilde{u}_1(t|\sigma), \dots, \tilde{u}_K(t|\sigma))$ can be obtained from the functions on the left-hand side of (48) and (49) in an obvious manner.

Similarly, we obtain from (52) the new canonical state constraints

$$\int_0^T \hat{G}_i(b(t|\sigma)) = 0, \quad i = 1, 2, 3, 4, 5, 6. \quad (74)$$

Similarly, from (67) we obtain the new objective function

$$\max U(u) \\ = \sum_{i=1}^N \int_0^T J(b_3(t), \hat{u}_i(t-d|\sigma), \hat{u}_1(t|\sigma), \dots, \hat{u}_N(t|\sigma)), \quad (75)$$

where $J(b_3(t), \hat{u}_i(t-d|\sigma), \hat{u}_1(t|\sigma))$ can be obtained from the right-hand side of the objective function (47) in an obvious manner.

Thus the new optimal control problem, problem (P3), can be stated as follows.

Problem (P3) Subject to the new state (buffer) Eqs. (68)–(70) with initial condition (71), the new conversation of TCP flow constraints in canonical form (72)–(73), the bounds on the input transmission rate (32), and the buffer capacity constraints in canonical form (74), we wish to find $\sigma \in \Xi^p$ that minimizes the new objective function (75).

Remark 4.1 Using the modified control parameterization method, we have translated problem (P2) into the new problem (P3), which contains only discrete time-delayed arguments (i.e., does not contain any control-dependent time-delayed arguments) with no explicit initial condition on the control functions when the time t is nonpositive. (The discrete time-delayed arguments are $\tilde{u}_i(t-d|\sigma)$ and $\hat{u}_i(t-d|\sigma)$ in (72), (73), and (75), respectively.) Thus problem (P3) can be solved more easily by any optimization method than problem (P2).

5 Gradient formulae for solving the control-dependent time-delayed optimal control problem by the control parametrization method

To solve problem (P3) as a mathematical programming problem, we require the gradient formulae for the objective function and all constraints. We will derive the gradient formula

for the canonical constraints because the gradient formulae for the objective function (79) and the canonical continuous state constraints (75) can be obtained in a similar manner. For simplicity, let the superscript $i = 1$ in (72) in the derivation of the gradient formula. Thus we consider

$$g_1(\sigma) = \int_0^T \bar{G}_1(b_1(t|\sigma), \hat{u}_1(t|\sigma), \tilde{u}_1(t-d|\sigma), \tilde{\tilde{u}}_1(t|\sigma), \dots, \tilde{\tilde{u}}_K(t|\sigma)) \quad (76)$$

subject to

$$\frac{d(b_1(t|\sigma))}{dt} = F_1(b_1(t|\sigma), \tilde{u}_1(t|\sigma), \dots, \tilde{\tilde{u}}_K(t|\sigma)). \quad (77)$$

Before we derive the gradient formulae for $g_1(\sigma)$, we need the following analytic considerations.

Remark 5.1 From the control bounds (31) and the initial conditions of controls (23) and (36), we observe that $\tilde{u}_i(t - w/\tilde{u}_i(t))$ and $\tilde{\tilde{u}}_i(t - d - w/\tilde{u}_i(t))$ are continuously differentiable with respect to $\tilde{u}_i(t)$ for all $u \in \mathcal{U}$. Thus from (64), (66), (48), (50), and the fact that $\xi_\delta(b_1(t), \sum_{i=1}^K \tilde{u}_i(t - w/\tilde{u}_i(t)) - c_1)$ and $\xi_\delta(b_1(t), \sum_{i=1}^K \tilde{\tilde{u}}_i(t - d - w/\tilde{u}_i(t)) - c_1)$ are continuously differentiable with respect to each of their arguments, it is clear that all $F_1(b_1(t|\sigma), \tilde{\tilde{u}}_1(t|\sigma), \dots, \tilde{\tilde{u}}_K(t|\sigma))$ in (68) and $\bar{G}_1(b_1(t|\sigma), \hat{u}_1(t|\sigma), \tilde{u}_1(t-d|\sigma), \tilde{\tilde{u}}_1(t|\sigma), \dots, \tilde{\tilde{u}}_K(t|\sigma))$ in (72) are continuously differentiable with respect to each of their arguments. However, by (64) $\tilde{u}_i(t|\sigma)$ is not continuous with respect to σ at the discrete times $t = \frac{w}{\sigma_i^k} + \frac{(k-1)T}{p}$, $k = 1, p-1$, and by (66) $\tilde{\tilde{u}}_i(t|\sigma)$ is not continuous with respect to σ at the discrete times $t = \frac{w}{\sigma_i^k} + d + \frac{(k-1)T}{p}$, $k = 1, p-1$. Thus $F_1(b_1(t|\sigma), \tilde{\tilde{u}}_1(t|\sigma), \dots, \tilde{\tilde{u}}_K(t|\sigma))$ and $\bar{G}_1(b_1(t|\sigma), \hat{u}_1(t|\sigma), \tilde{u}_1(t-d|\sigma), \tilde{\tilde{u}}_1(t|\sigma), \dots, \tilde{\tilde{u}}_K(t|\sigma))$ are both continuously differentiable with respect to σ , except at the discrete times mentioned above.

Let $\sigma = [\tilde{\sigma}_1^1, \dots, \tilde{\sigma}_N^1, \hat{\sigma}_1^1, \dots, \hat{\sigma}_N^1, \dots, \tilde{\sigma}_1^p, \dots, \tilde{\sigma}_N^p, \hat{\sigma}_1^p, \dots, \hat{\sigma}_N^p]^T \in R^{2Np}$. Then from (76) and (77) for all $k = 1, \dots, p$, we have

$$\frac{\partial g_1(\sigma)}{\partial \tilde{\sigma}_i^k} = 0, \quad i = K+1, \dots, N, \quad (78)$$

and

$$\frac{\partial g_1(\sigma)}{\partial \hat{\sigma}_i^k} = 0, \quad i = 2, \dots, N. \quad (79)$$

Moreover, the method of finding the gradients of $g_1(\sigma)$ with respect to $\tilde{\sigma}_i^k$ and $\hat{\sigma}_1^k$, $i = 1, \dots, K$ are the same. Thus we only consider the method of finding the gradients of $g_1(\sigma)$ with respect to $\tilde{\sigma}_i^k$, $i = 1, \dots, K$. To derive the gradient of $g_1(\sigma)$ with respect to σ_i^k , $i = 1, \dots, K$, we define the following adjoint system defined on $[0, T]$:

$$\begin{aligned} & (\dot{\lambda}_i(t|\sigma))^T \\ &= - \frac{\partial H(b_1(t|\sigma), \hat{b}_1(t|\sigma), \hat{u}_1(t|\sigma), \hat{u}_1(t|\sigma), \tilde{u}_1(t|\sigma), \tilde{\tilde{u}}_1(t|\sigma), \dots, \tilde{\tilde{u}}_K(t|\sigma), \tilde{\tilde{u}}_1(t|\sigma), \dots, \tilde{\tilde{u}}_K(t|\sigma), \lambda_1(t|\sigma))}{\partial (b_1(t|\sigma))}, \end{aligned}$$

$$t \in [0, T], \quad (80)$$

$$(\lambda_1(T|\sigma))^T = 0, \quad (81)$$

where

$$\bar{u}_1(t|\sigma) = \tilde{u}_1(t-d|\sigma), \quad (82)$$

$$\widehat{u}_1(t|\sigma) = \hat{u}_1(t+d|\sigma), \quad (83)$$

$$\widehat{b}_1(t|\sigma) = \tilde{b}_1(t+d|\sigma), \quad (84)$$

and $H : R^{2K+4} \rightarrow R$ is the Hamiltonian function defined by

$$\begin{aligned} H(b_1, \widehat{b}_1, \hat{u}_1, \bar{u}_1, \tilde{u}_1, \tilde{\tilde{u}}_1, \dots, \tilde{\tilde{u}}_K, \tilde{\tilde{u}}_1, \dots, \tilde{\tilde{u}}_K, \lambda_1) \\ = \bar{G}_1(b_1, \hat{u}_1, \bar{u}_1, \tilde{\tilde{u}}_1, \dots, \tilde{\tilde{u}}_K) + \bar{G}_1(\widehat{b}_1, \widehat{u}_1, \tilde{u}_1, \tilde{\tilde{u}}_1, \dots, \tilde{\tilde{u}}_K)e(T-d-t) \\ + \lambda_1 F_1(b_1, \tilde{\tilde{u}}_1, \dots, \tilde{\tilde{u}}_K), \end{aligned} \quad (85)$$

where $e(t)$ is the unit step function defined by

$$e(t) = \begin{cases} 1 & \text{if } t > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (86)$$

Furthermore, we set

$$\bar{u}_1(t|\sigma) = 0, \quad t \in (T-d, T]. \quad (87)$$

For simplicity, we replace the notation $t|\sigma$ by t in the discussion of the remainder of this section, except for the statement of Theorem 5.1 and the last part of its proof. Moreover, we set

$$\tilde{F}_1 = F_1(b_1(t|\sigma), \tilde{\tilde{u}}_1(t|\sigma), \dots, \tilde{\tilde{u}}_K(t|\sigma)), \quad (88)$$

$$\tilde{G}_1 = \bar{G}_1(b_1(t|\sigma), \hat{u}_1(t|\sigma), \bar{u}_1(t|\sigma), \tilde{\tilde{u}}_1(t|\sigma), \dots, \tilde{\tilde{u}}_K(t|\sigma)), \quad (89)$$

$$\tilde{\tilde{G}}_1 = \bar{G}_1(b_1(t+d|\sigma), \hat{u}_1(t+d|\sigma), \tilde{u}_1(t|\sigma), \tilde{\tilde{u}}_1(t|\sigma), \dots, \tilde{\tilde{u}}_K(t|\sigma)), \quad (90)$$

and

$$\begin{aligned} \tilde{H} = H(b_1(t|\sigma), \widehat{b}_1(t|\sigma), \hat{u}_1(t|\sigma), \widehat{u}_1(t|\sigma), \bar{u}_1(t|\sigma), \tilde{u}_1(t|\sigma), \tilde{\tilde{u}}_1(t|\sigma), \dots, \\ \tilde{\tilde{u}}_K(t|\sigma), \tilde{\tilde{u}}_1(t|\sigma), \dots, \tilde{\tilde{u}}_K(t|\sigma), \lambda_1(t|\sigma)). \end{aligned} \quad (91)$$

In view of Remark 5.1, we have the following theorem. (Note that the technique used in proving the following theorem is modified from those used in the proofs of Theorems 5.3.1 and 12.3.1 of [19].)

Theorem 5.1 Consider problem (P3). The gradient of the functional $g_1(\sigma)$ in (76) with respect to the component $\tilde{\sigma}_i^k, i = 1, \dots, K, k = 1, \dots, p$, is given by

$$\begin{aligned} & \frac{\partial g_1(\sigma)}{\partial \tilde{\sigma}_i^k} \\ &= \int_0^T \left[\bar{\bar{G}}_1(b_1(t+0), \hat{u}_1(t+0), \bar{u}_1(t+0), \tilde{\tilde{u}}_1(t+0), \dots, \tilde{\tilde{u}}_K(t+0)) \right. \\ & \quad \left. - \bar{\bar{G}}_1(b_1(t-0), \hat{u}_1(t-0), \bar{u}_1(t-0), \tilde{\tilde{u}}_1(t-0), \dots, \tilde{\tilde{u}}_K(t-0)) \right] \\ & \quad \times \delta \left(t - \frac{w}{\sigma_i^k} - \frac{(k-1)T}{p} - d \right) \times \frac{w}{(\sigma_i^k)^2} dt \\ & \quad + \int_0^T \left[\lambda^T(t) \times [\bar{F}(b_1(t+0), \tilde{\tilde{u}}_1(t+0), \dots, \tilde{\tilde{u}}_K(t+0)) \right. \\ & \quad \left. - \bar{F}(b_1(t-0), \tilde{\tilde{u}}_1(t-0), \dots, \tilde{\tilde{u}}_K(t-0))] \right. \\ & \quad \left. \times \delta \left(t - \frac{w}{\sigma_i^k} - \frac{(k-1)T}{p} \right) \times \frac{w}{(\sigma_i^k)^2} \right] dt + \int_0^T \bar{f}_i^k(t) dt, \end{aligned} \quad (92)$$

where

$$\bar{f}_i^k(t) = \begin{cases} \frac{\partial \bar{H}}{\partial \bar{u}_1} \times \chi_{[(k-1)T/p, kT/p)}(t) + \frac{\partial \bar{H}}{\partial \tilde{u}_1} \times \chi_{[w/\tilde{\sigma}_1^k + (k-1)T/p, w/\tilde{\sigma}_1^{k+1} + kT/p)}(t), \\ \quad \text{for } i = 1, k = 2, \dots, p-1, \\ \frac{\partial \bar{H}}{\partial \bar{u}_1} \times \chi_{[(k-1)T/p, kT/p)}(t) + \frac{\partial \bar{H}}{\partial \tilde{u}_1} \times \chi_{[w/\tilde{\sigma}_1^k + (k-1)T/p, T)}(t) \quad \text{for } i = 1, k = p, \\ \frac{\partial \bar{H}}{\partial \tilde{u}_i} \times \chi_{[w/\tilde{\sigma}_i^k + (k-1)T/p, w/\tilde{\sigma}_i^{k+1} + kT/p)}(t) \quad \text{for } i = 2, \dots, K, k = 1, \dots, p-1, \\ \frac{\partial \bar{H}}{\partial \tilde{u}_i} \times \chi_{[w/\tilde{\sigma}_i^k + (k-1)T/p, T)}(t) \quad \text{for } i = 2, \dots, K, k = p. \end{cases} \quad (93)$$

and for any analytic function $f(t)$, we have

$$f(t+0) = \lim_{\tau \downarrow t} f(\tau), \quad (94)$$

and

$$f(t-0) = \lim_{\tau \uparrow t} f(\tau), \quad (95)$$

Proof Let $\sigma = [\tilde{\sigma}_1^1, \dots, \tilde{\sigma}_N^1, \hat{\sigma}_1^1, \dots, \hat{\sigma}_N^1, \dots, \tilde{\sigma}_1^p, \dots, \tilde{\sigma}_N^p, \hat{\sigma}_1^p, \dots, \hat{\sigma}_N^p]^T \in R^{2Np}$ be given. Let us perturb each component $\tilde{\sigma}_i^k, i = 1, \dots, K, k = 1, \dots, p$ by an arbitrary small number ε . Let

$$\sigma(\varepsilon) = \sigma + e(\tilde{\rho}_i^k)\varepsilon, \quad (96)$$

where $e(\tilde{\rho}_i^k) \in R^{2Np}$ is a column vector whose $[(2k-1)+i]$ th component is 1, and for other components, the value is 0. Thus the first-order variation of $g_1(\sigma)$ can be written as

$$\Delta g_1(\sigma) = \frac{dg_1(\sigma(\varepsilon))}{d\varepsilon} \bigg|_{\varepsilon=0} = \sum_{i=1}^K \sum_{k=1}^p \frac{\partial g_1(\sigma)}{\partial \tilde{\sigma}_i^k} \tilde{\rho}_i^k. \quad (97)$$

Let $\Delta b_1(t), \Delta \bar{u}_1(t), \Delta \tilde{u}_1(t), \Delta \tilde{\tilde{u}}_1(t), \dots, \Delta \tilde{u}_K(t), \Delta \tilde{\tilde{u}}_1(t), \dots, \Delta \tilde{\tilde{u}}_K(t)$ be, respectively, the first-order variation of $b_1(t), \bar{u}_1(t), \tilde{u}_1(t), \tilde{\tilde{u}}_1(t), \dots, \tilde{u}_K(t), \tilde{\tilde{u}}_1(t), \dots, \tilde{\tilde{u}}_K(t)$, respectively, when σ changes to $\sigma(\varepsilon)$. Thus by using a similar approach as that used in the proof of Theorem 5.3.1 of [19] we get

$$\begin{aligned} \Delta g_1(\sigma) &= \int_0^T \sum_{i=1}^K \sum_{k=1}^p [\bar{G}_1(\tilde{b}_1(t+0), \hat{u}_1(t+0), \bar{u}_1(t+0), \tilde{\tilde{u}}_1(t+0), \dots, \tilde{\tilde{u}}_K(t+0)) \\ &\quad - \bar{G}_1(\tilde{b}_1(t-0), \hat{u}_1(t-0), \bar{u}_1(t-0), \tilde{\tilde{u}}_1(t-0), \dots, \tilde{\tilde{u}}_K(t-0))] \\ &\quad \times \delta \left(t - \frac{w}{\sigma_i^k} - \frac{(k-1)T}{p} - d \right) \times \frac{w}{(\sigma_i^k)^2} \times \tilde{\rho}_i^k dt \\ &\quad + \int_0^T \left[\frac{\partial \bar{G}_1}{\partial b_1} \Delta b_1(t) + \frac{\partial \bar{G}_1}{\partial \bar{u}_1} \Delta \bar{u}_1(t) + \sum_{k=1}^K \frac{\partial \bar{G}_1}{\partial \tilde{\tilde{u}}_i} \Delta \tilde{\tilde{u}}_i(t) \right] dt. \end{aligned}$$

Then the proof of this theorem can be obtained by using a standard method used in obtaining the gradient of an analytical function in optimal control theory, such as that given in Theorem 12.3.1 of [19]. \square

6 Numerical example

Example 6.1 In this numerical example, we consider a particular problem that consists of four sources, three links, two origins, and a single destination, such that sources s_1 and s_2 use links l_1 and l_3 to reach the destination, whereas sources s_3 and s_4 use links l_2 and l_3 to reach the destination.

The structure of the computer network is the same as that shown in Fig. 1.

The parameters of this example are as follows:

The final time of the problem (second): 3.

The capacity of the three links, i.e., the bandwidth of the three links (bits per second): $c_1 = 1, c_2 = 0.5, c_3 = 0.4$.

The packet size transmitted by the router of the three links (bits): $w_1 = 0.1, w_2 = 0.15, w_3 = 0.11$.

The capacity of the three routers (bits per second): $B_1 = 1, B_2 = 1.8, B_3 = 2.7$.

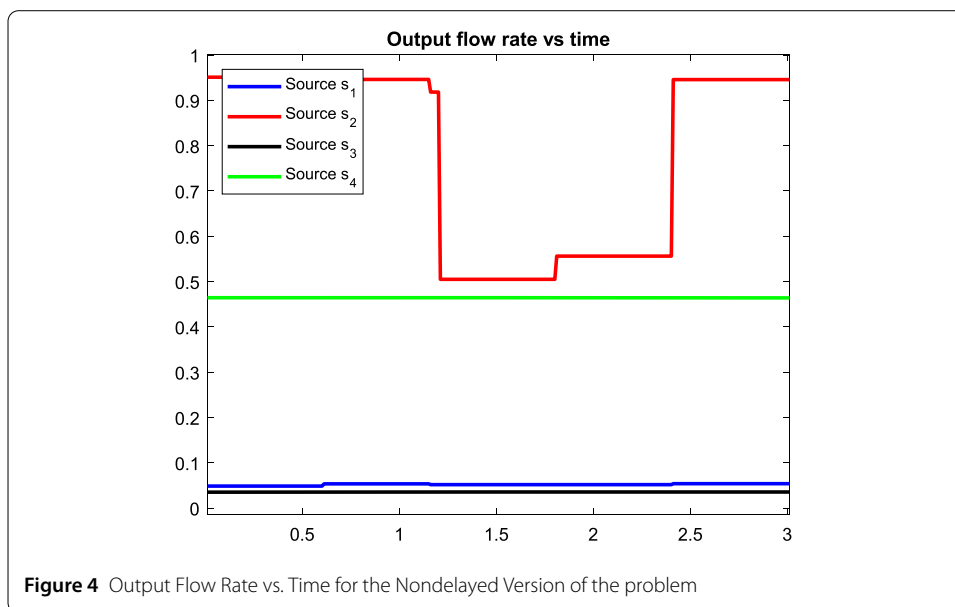
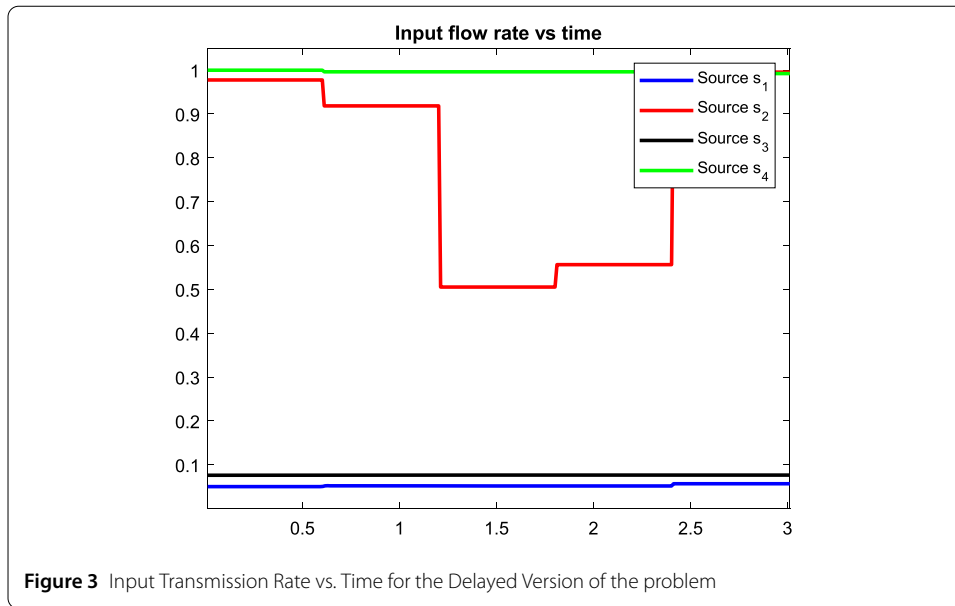
The relative importance of the messages sent from the four sources to the destination: $M_1 = 1.5, M_2 = 2.5, M_3 = 1.7, M_4 = 2.7$.

The upper bound of the input source rate from each of the sources s_i ($i = 1, 2, 3, 4$) (bits per second) at the origin: (1.05, 1.00, 0.98, 1.00).

The lower bound of the input source rate from each of the sources s_i ($i = 1, 2, 3, 4$) (bits per second) at the origin: (0.05, 0.06, 0.07, 0.05).

The latency of each link d_i ($i = 1, 2, 3$) = 0.3, 0.5, 0.5.

Problem (P3) has been formulated and solved by the modified control parameterization method described in Sect. 5 using the number of partitions of the control in the time interval $[0, 3]$ equal to 5. For comparison, the nondelayed version of problem (P3) has also been solved using the same number of partitions in the time interval $[0, 3]$, where the nondelayed version of problem (P3) is obtained by deleting both the forward transmission

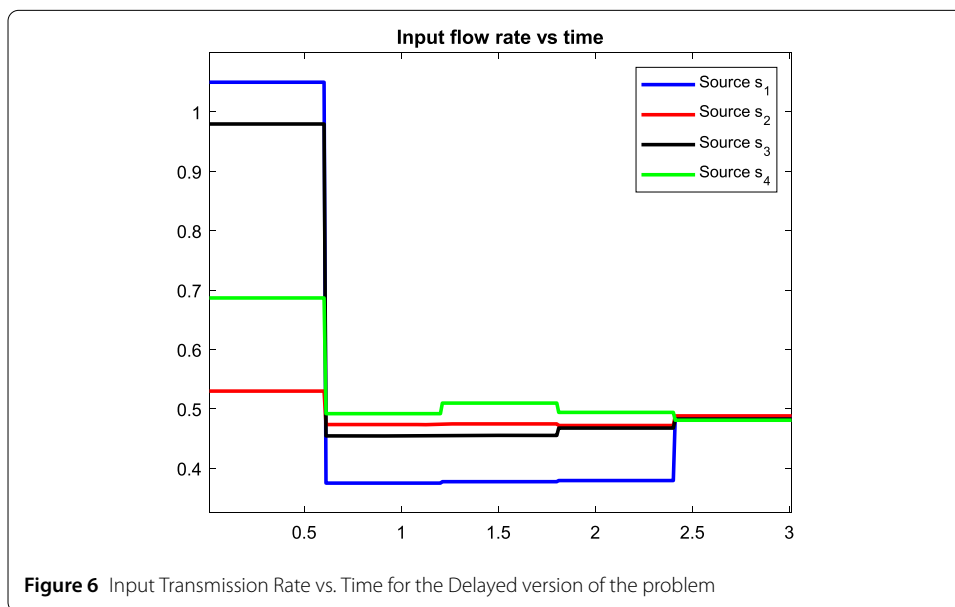
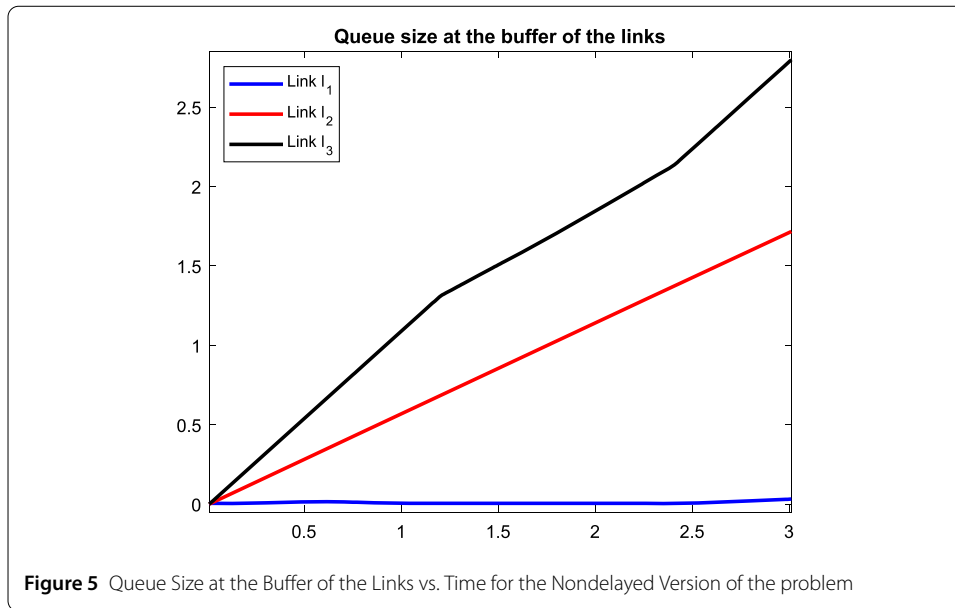


delays and the latencies in the buffer equations and the conservation of flow equation of problem (P3).

The graphical results for the nondelayed version of the problem are plotted in Figs. 3, 4, and 5. The graphical results for the delayed version of the problem are plotted in Figs. 6, 7, and 8.

Comparison of the optimal input transmission rates of the two versions of the problem

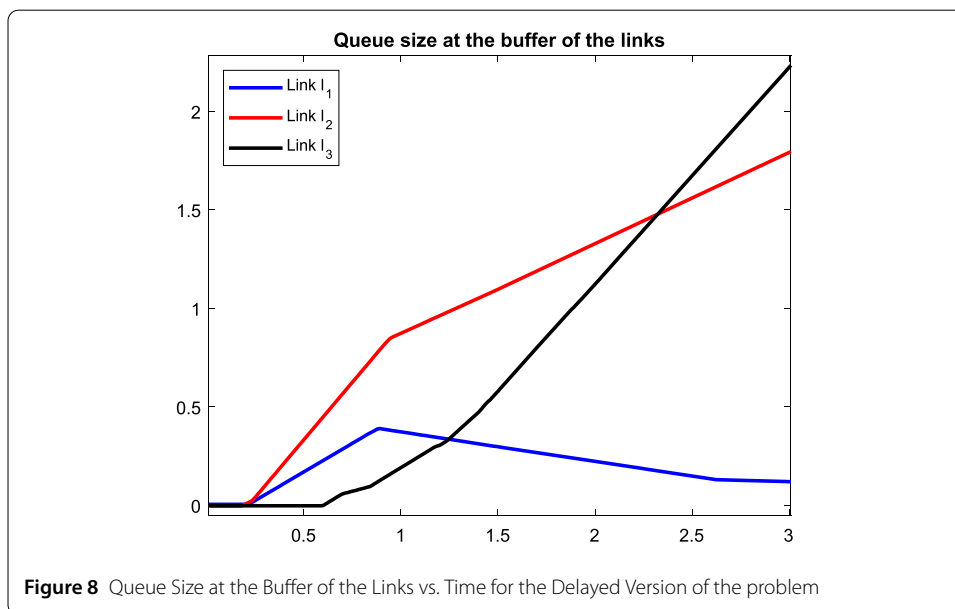
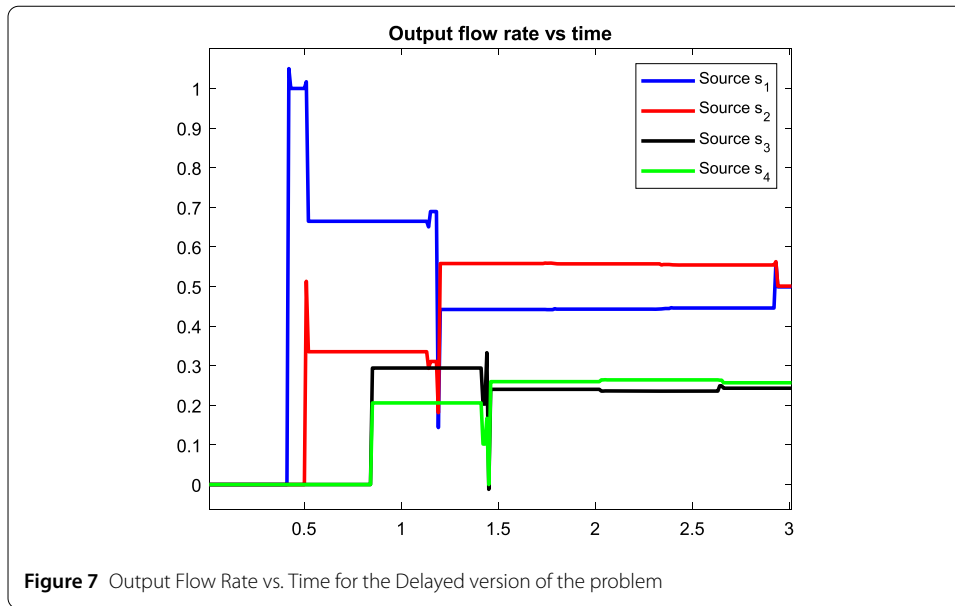
From Fig. 3, we observe that at the optimal solution of the nondelayed version of the problem, source s_4 , whose input transmission rate lies in the upper bound (i.e., 1.0) of its control constraint region for all $t \in [0, 3]$, has the largest average input transmission rate of 1.0; source s_2 has the second largest average transmission rate of about 0.784; source s_3 , whose input transmission rate lies in the lower bound (i.e., 0.07) of the constraint region



for all $t \in [0, 3]$, has the third largest average transmission rate of 0.07; source s_1 , whose input transmission rate lies in the lower bound (i.e., 0.05) of the constraint region for all $t \in [0, 3]$, has the smallest average input transmission rate of 0.05.

From Fig. 7, we observe that at the optimal solution of the delayed version of the problem, source s_3 has the largest average input transmission rate of about 0.572; source s_4 has the second largest average input transmission rate of about 0.540; source s_1 has the third average input transmission rate of about 0.538; and source s_2 has the smallest average input transmission rate of about 0.492.

The average input transmission rates of source s_1 , source s_2 , source s_3 , and source s_4 of the delayed version are higher than those of the nondelayed version by about 976.00%, -37.24%, 717.14%, and -46.00%, respectively. The overall input transmission rate of all



the four sources of the delayed version is higher than that of the nondelayed version by 12.50%. However, if we take the relative importance of the messages sent from each of the four sources to the destination also into consideration, then we obtain that the overall weighted input transmission rates of all the four sources of the delayed version is lower than those of the nondelayed version by 7.96%.

Comparison of the output flow rates of the two versions of the problem

From Fig. 4 we observe that at the optimal solution of the nondelayed version of the problem, source s_2 has the largest average output flow rate of 0.774; source s_4 , whose output flow rate is equal to 0.48 for all $t \in [0, 3]$, has the second largest average output flow rate of 0.48; source s_1 , whose output flow rate is equal to 0.06 for all $t \in [0, 3]$, has the third

largest average output flow rate of 0.06; source s_3 , whose output flow rate is equal to 0.04 for all $t \in [0, 3]$, has the smallest average output flow rate of 0.04.

From Fig. 7 we first observe that the output flow rates of all the sources of the delayed version of the problem are equal to zero during short periods at the beginning of their transmission. This phenomenon confirms the fact that due to the forward transmission delays and the latencies, all the sources require certain times to send their messages from the origin to the destination. From Fig. 8 we observe that at the optimal solution of the delayed version of problem (P3), source s_1 , whose output flow rate is equal to zero for all $t \in [0, 0.4]$, has the largest average output flow rate of 0.5777; source s_2 , whose output flow rate is equal to zero for all $t \in [0, 0.5]$, has the second largest average output flow rate of 0.4870; source s_3 , whose output flow rate is equal to zero for all $t \in [0, 0.75]$, has the third largest average output flow rate of 0.1780; source s_4 , whose output flow rate is equal to zero for all $t \in [0, 0.75]$, has the smallest average output flow rate of 0.1692.

Thus the average output flow rates of source s_1 , source s_2 , source s_3 , and source s_4 of the delayed version are higher than those of the nondelayed version by 862.83%, -37.36% , 345.00%, and -64.75% , respectively. The overall output flow rate of all the four sources of the delayed version is higher than that of the nondelayed version by 4.02%. However, if we take the relative importance of the messages sent from each of the four sources to the destination also into consideration, we obtain that the overall weighted output flow rates of all the four sources of the delayed version is lower than that of the nondelayed version by 16.31%.

Comparison of the queue sizes of the two different versions of the problem

From Fig. 5, we observe that at the optimal solution of the nondelayed version of the problem, the queue size at the buffer of link l_1 is zero for all $t \in [0, 3]$; the queue size at the buffer of link l_2 is increasing with a constant speed of about 0.53; the queue size at the buffer of link l_3 is increasing with an average speed of about 0.9. Thus the buffer capacity constraints of links l_1 and l_2 are both nonbinding for all $t \in [0, 3]$, but the buffer capacity constraint of link l_3 is binding at the final time $t = 3$ (i.e., the buffer capacity B_3 of link $l_3 = 2.7$ at $t = 3$.) This means that the buffer of link l_3 should not handle more queues at the final time $t = 3$; otherwise, some messages will be lost.

Similar to the behavior of the output flow rates of the delayed version of the problem, we observe from Fig. 8 that the queue sizes at all the buffers of the delayed version are equal to zero during short periods at the beginning of the sources' transmission. This phenomenon confirms the fact that due to the forward transmission delays and/or the latency, all the routers at the beginning of the sources' transmission must wait for certain times before they can serialize the first packet of messages into the links. From Fig. 8 we observe that at the optimal solution of the delayed version of the problem, the queue size at the buffer of link l_1 is equal to zero for all $t \in [0, 0.2]$, but increases with a constant speed of about 0.52 for all $t \in [0.2, 0.9]$ and decreases with an almost constant speed of about 0.17 for all $t \in [0.9, 3.0]$; the queue size at the buffer of link l_2 is also equal to zero for all $t \in [0, 0.2]$, but increases with a speed of about 1.4 for all $t \in [0.2, 0.92]$ and increases with a speed of about 0.74 for all $t \in [0.92, 3]$; the queue size at the buffer of link l_3 is equal to zero for all $t \in [0, 0.6]$, but increases with an average speed of about 0.92 for all $t \in [0.6, 3]$. Thus the buffer capacity constraint of each of the links l_1 , l_2 , and l_3 is nonbinding for all $t \in [0, 3]$.

Thus, unlike the queue sizes of the optimal solution of the nondelayed version of the problem, the queue sizes of one of the links (i.e., link l_1) of the optimal solution of the

delayed version show both increasing and decreasing trends. Moreover, the average queue sizes at the buffers of links 1, 2, and 3 of the delayed version are higher than those of the nondelayed version by a percentage of infinity (because the queue size at the buffer of link 1 is zero), 51.57%, and -2.208% , respectively. (The main reason that the average queue size at the buffer of link l_3 of the delayed version is slightly less than that of the nondelayed version is mainly due to the fact that at the beginning of the transmission, the router of link l_3 of the delayed version has to wait for 0.6 units of time before they can serialize the first packet into the link.)

Comparison of the objective function values of the two versions of the problem

The optimal objective function values for the nondelayed and time-delayed versions of the problem are 10.0313 and 7.4805, respectively. Thus the objective function value of the delayed version is less than that of the nondelayed version by 38.415.

From the above discussions, we can see that the amount of weighted sources' input flows at the origins and the amount of the weighted sources' output flows at the destination of the delayed version of the problem are both less than those of the nondelayed version of the problem by margins of about 8% and 16%, respectively, but the queue sizes at the buffers of the delayed version of the problem are very much greater than those of the nondelayed version of the problem. Thus this example clearly reflects how the optimal transmission of messages in computer networks in real life can be affected by both the forward transmission delays in the buffers and the latencies in the links.

7 Conclusion and suggestions for further study

In this paper, we first formulate the optimal transmission of messages in a computer network as a nondelayed optimal control problem. We then develop a modified control parameterization method for solving this time-delayed optimal control problem. The gradients of the new objective function and the new constraint functions generated by this modified control parameterization method are derived. In this way, we can provide an open-loop control for the time-delayed optimal transmission problem of the computer network; by allocating the optimal flow rate to each link of a real-life computer network, we maximize the number of messages sent from N origins to one destination in a given period.

A numerical example involving four sources, three links, two origins, and one destination has been solved by using both the nondelayed and the time-delayed versions of the above optimal transmission problem. The results show that the amount of weighted sources' input flows at the origins and the amount of weighted sources' output flows at the destination of the delayed version of the transmission problem are both less than those of the nondelayed version of the problem by margins of about 8% and 16%, respectively, but the queue sizes at the buffers of the delayed version of the problem are much greater than those of the nondelayed version of the problem. Thus this example clearly reflects how the optimal transmission of messages in computer networks in real life can be affected by both the forward transmission delays in the buffers and the latencies in the links.

For further research direction, we would like to extend the modified control parameterization method to solve more difficult optimal control problems, which involve both state-dependent and control-dependent time-delayed arguments.

Acknowledgements

The authors would like to thank the anonymous referees for the valuable comments and suggestions.

Funding

No funding is acquired for this research work.

Availability of data and materials

There is no data set used in this paper.

Declarations**Competing interests**

The authors declare that they have no competing interests.

Authors' contributions

Each author contributes equally to the production of this manuscript. KH focuses on the theory of the research; both KH and YCE work on the modeling of the problem and focus on the computational work, and HW focuses on the background of the posed problem. All authors read and approved the final manuscript.

Author details

¹School of Computer Science and Applied Mathematics, University of the Witwatersrand, Johannesburg, South Africa.

²Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 12 January 2022 Accepted: 9 June 2022 Published online: 07 July 2022

References

1. Kelly, F.P.: Charging and rate control for elastic traffic. *Eur. Trans. Telecommun.* **8**(1), 33–37 (1997)
2. Kelly, F.P., Maullo, A., Tan, D.: Rate control in communication networks: shadow prices, proportional fairness and stability. *J. Oper. Res. Soc.* **49**, 237–252 (1998)
3. Abolfazli, Shah-Mansouri, V.: Dynamic adjustment of queue levels in TCP legas-based networks. *Electron. Lett.* **52**, 361–363 (2016)
4. Lai, K., Baker, M.: Measuring link bandwidths using a deterministic model of packet delay. *Comput. Commun. Rev.* **30**, 283–294 (2000)
5. Lavaei, J., Doyle, J.C., Low, S.H.: Utility functionals associated with available congestion control algorithms. In: *Proceeding of IEEE International Conference Computation Communication*, pp. 1–9 (2010)
6. Lestas, I., Vinnicombe, G.: How good are deterministic models for analyzing congestion control in delayed stochastic networks? In: *Proceedings IEEE Conference Decision Control*, pp. 4984–4989 (2004)
7. Misra, V., Gong, W.B., Towsley, D.: Fluid-based analysis of a network of AQM routers supporting TCP flows with an application to RED. In: *Proceedings of ACM SIGCOMM*, pp. 151–160 (2000)
8. Mo, J., Walrand, J.: Fair end-to-end window-based congestion control. *IEEE/ACM Trans. Netw.* **8**(5), 556–567 (2000)
9. Raina, G., Manjunath, S., Prasad, Giridhar, K.: Stability and performance analysis of compound TCP with rem and drop-tail queue management. *IEEE/ACM Trans. Netw.* **24**, 1961–1973 (2016)
10. Shakkottai, S., Srikant, R.: Network optimization and control. *Found. Trends Netw.* **2**(3), 271–379 (2008)
11. Sirisena, H., Hassan, M., Haider, A.: Optimal TCP Congestion Control, *Computer Communication* (2002)
12. Sojoudi, S., Low, S.H.: Buffering Dynamics and Stability of Internet Congestion Controllers. *IEEE/ACM Transactions on Networking* **22**(6) (2014)
13. Sojoudi, S., Low, S.H., Doyle, J.C.: Effect on buffers on stability of Internet congestion controllers. In: *Proceeding of IEEE International Conference Computer Communication Mini-Conference*, 471–475 (2011)
14. Zhu, J., Luo, T., Xie, Y.W., Dullerud, G.E.: An average queue-length-difference-based detection algorithm in TCP/AQM network. *Int. J. Adapt. Control Signal Process.* **32**, 742–752 (2018)
15. Yu, C., Teo, K.L., Zhang, L., Bai, Y.: A new exact penalty function method for continuous inequality constrained optimization problem. *J. Ind. Manag. Optim.* **6**, 895–910 (2010)
16. Li, B., Wang, Y., Zhang, K., Duan, G.R.: Constrained feedback control for spacecraft reorientation with an optimal gain. *IEEE Trans. Aerosp. Electron. Syst.* **57**, 3916–3926 (2021)
17. Li, B., Zhang, J., Dai, L., Teo, K.L., Wang, S.: A hybrid offline optimization method for reconfiguration of multi-UAV formations. *IEEE Trans. Aerosp. Electron. Syst.* **57**, 506–520 (2021)
18. Lin, Q., Loxton, R., Teo, K.L.: The control parametrization method for nonlinear optimal control: a survey. *J. Ind. Manag. Optim.* **10**, 275–309 (2014)
19. Teo, K.L., Goh, C.J., Wong, K.H.: *A Unified Computational Approach to Optimal Control Problems*. Pitman Monographs and Surveys in Pure and Applied Mathematics. London (1990)
20. Teo, K.L., Li, B., Yu, C., Rehbock, V.: *Applied and Computational Optimal Control: A Control Parametrization Approach*. Springer, Berlin (2021)
21. Yang, F., Teo, K.L., Loxton, R., Rehbock, V., Li, B., Yu, C., Jennings, L.: VISUAL MISER: an efficient user-friendly visual program for solving optimal control problems. *J. Ind. Manag. Optim.* **12**, 781–810 (2016)
22. Wong, K.H., Lee, H.W.J., Chan, C.K., Myburgh, C.: Control parametrization and finite element method for controlling multi-species reactive transport in an underground channel. *J. Optim. Theory Appl.* **151**, 168–187 (2013)
23. Liu, C., Loxton, R., Lin, Q., Teo, K.L.: Dynamic optimization for switched time-delay systems with state-dependent switching conditions. *SIAM J. Control Optim.* **56**, 3499–3523 (2018)
24. Wu, D., Bai, Y., Yu, C.: A new computational approach for optimal control problems with multiple time-delay. *Automatica* **101**, 388–395 (2019)

25. Yu, C., Lin, Q., Loxton, R., Teo, K.L., Wang, G.: A hybrid time-scaling transformation for time-delay optimal control problems. *J. Optim. Theory Appl.* **169**, 876–901 (2016)
26. Wu, D., Bai, Y., Xie, F.: Time-scaling transformation for optimal control problem with time-varying delay. *Discrete Contin. Dyn. Syst.* **13**(6), 1683–1695 (2020)
27. Liu, C., Loxton, R., Teo, K.L., Wang, S.: Optimal State-Delay Control in Nonlinear Dynamic Systems. *Automatica* **135** (2022). <https://doi.org/10.1016/j.automatica.2021.109981>

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)