A simulative evaluation of a dynamic fluid flow model for designing Congestion Control using non-linear control theory

L. Rossides\textsuperscript{1}, A. Pitsillides\textsuperscript{1}, P. Ioannou\textsuperscript{2}

\textsuperscript{1}University of Cyprus, \textsuperscript{2}University of Southern California

Abstract:

Most of the congestion control algorithms used today are based on ad hoc techniques and intuition and their effectiveness is proven through extensive simulations. We are motivated by recent results in non-linear control theoretic techniques for the design of congestion controllers with proven analytical performance bounds, and seek suitable models. In this paper we present a well known fluid flow model and investigate using simulation its suitability for capturing essential dynamic behaviour. This model represents in an effective way the dynamics behind a modern high bandwidth network. We solve this model using MATLAB (a well known mathematical package) and OPNET (a well known event based network simulator) to show that the behaviour of our model is very close to the ‘real’ world. This result motivates its use for non-linear congestion control design.

1 Introduction

It is generally accepted that the problem of network congestion control remains a critical issue and a high priority, especially given the growing size, demand, and speed (bandwidth) of the increasingly integrated services network. One could argue that network congestion is a problem unlikely to disappear in the near future. Furthermore congestion may become unmanageable unless effective, robust, and efficient methods for congestion control are developed. This assertion is based on the fact that despite the vast research efforts, spanning a few decades, and the large number of different control schemes proposed, there are still no universally acceptable congestion control solutions. For example, the existing congestion control solutions deployed in the Internet Transport Control Protocol (TCP) \([1,2]\) are increasingly becoming ineffective, and it is generally accepted that these solutions cannot
easily scale up even with various proposed “fixes” [3,4,5]. Also, it is worth pointing out that the User Datagram Protocol (UDP), the other transport service offered by TCP/IP Internet, offers no congestion control. This service is increasingly used by more and more demanding users for the delivery of real time video and voice services. The newly developed (also largely ad-hock) strategies [6,7,8] are also not proven to be robust and effective. Asynchronous Transfer Mode (ATM) has also witnessed a similar approach, with various congestion control schemes proposed [9,10]. Most of the proposed schemes are developed using intuition and are mostly based on simple non-linear control designs. These simple schemes have been demonstrated to be robust in a variety of scenarios that have been simulated [10]. Since these schemes are designed with significant non-linearities (e.g. two-phase—slow start and congestion avoidance—dynamic windows, binary feedback, additive-increase multiplicative-decrease flow control etc) based mostly on intuition, the analysis of the closed loop behaviour is difficult if at all possible, even for single control loop networks. The interaction of additional non-linear feedback loops can produce unexpected and erratic behaviour [11].

Empirical evidence demonstrates the poor performance and cyclic behaviour of the controlled TCP/IP Internet [12] (also confirmed analytically [13]). This is exacerbated as the link speed increases to satisfy demand (hence the bandwidth-delay product, and thus feedback delay, increases), and also as the demand on the network for better quality of service increases. Note for WAN networks a multifractal behaviour has been observed [14], and it is suggested that this behaviour —cascade effect—may be related to existing network controls [15].

Despite the successful application of control theory to other complex systems (e.g. power, traffic, chemical plants, space structures, aerospace systems, etc.), the development of network congestion control based on control theoretic concepts is quite unexplored. This in spite of the significant demands placed on the network system over the recent years for the delivery of guaranteed performance (in terms of quality of service to the users). One may attribute this to the complexity of the control problem, coupled with the lack of collaboration between teletraffic engineers and control systems theorists. Recently several attempts have
been made to develop congestion controllers using optimal control theory [16]; linear control [11,17,18]; predictive adaptive control [19,20]; fuzzy and neural control [21,22,23,24,25]; and non-linear control [26,27,28,29,30]. Despite these efforts the design of congestion network controllers whose performance can be analytically established and demonstrated in practice is still a challenging unresolved problem.

It is well known that the network system is a nonlinear dynamical system. The development of effective congestion control algorithms derived using non-linear control theory remains quite unexplored. We argue that the richness of non-linear control theory developed during the recent years justifies its use now. Some of the most important recent developments in non-linear control theory include: feedback linearisation [31], passivity theory [32], control Lyapunov functions [33,34], backstepping and tuning functions [35], neural and fuzzy control systems [36,37,38], and robust adaptive control for linear and nonlinear systems [39,40,41].

Early attempts to use non-linear control theory for network congestion control include [26-30]. In these works we based the derivation of the control strategy on a simplistic nonlinear dynamic model of a network queue, derived from fluid flow considerations and from matching the M/M/1 queue behavior at equilibrium. It can be argued that this model cannot accurately predict the behavior of a network system. Even so, we have initiated a study to design congestion controllers using this simplistic model which captures the ‘dominant’ dynamics [39] of the network system, but neglects secondary effects and the noisy environment. Since feedback control is developed to be able to handle significant modeling errors and inaccuracies it is not surprising that our preliminary results based on this simple model were very successful and encouraging for pursuing this approach further.

In this paper we use simulative comparison between our model and ATM. In section 2 we present the model and in section 3 the results obtained from our simulative comparison. Finally in section 4 we present our conclusions and future work.
2. **Dynamic Network Models**

Most of the current congestion control methods are based on intuition and ad hoc control techniques together with extensive simulations to demonstrate their performance. The problem with this approach is that very little is known why these methods work and very little explanation can be given when they fail. The use of dynamic models could provide a better understanding of how the network operates and can be used to develop control techniques whose properties can be established analytically even when such techniques are based on intuition and ad hoc guesses. For control design purposes the model does not need to be accurate. It is because of the inability of modelling the real world accurately that feedback was invented and control theory is widely used. A good feedback control design should be able to deal with considerable uncertainties and inaccuracies that are not accounted for in the model. Robust adaptive control techniques for example can be used to control dynamical systems whose parameters are completely unknown and high frequency dynamics and disturbances are completely neglected in the control design [35,39]. A plethora of similar control techniques [31,35,39] and tools developed during the last decade offer a strong potential for solving complex congestion control problems in computer networks. Below we present a simple dynamic model that has been proposed in literature and used by a number of researchers for designing congestion controllers.

### 2.1 Fluid Flow Model

A dynamic model is sought, in a form suitable for a distributed control solution. The objective (from a controls point of view) is to find a model which captures the ‘essential’ dynamic behaviour, but has low order complexity relative to detailed probabilistic models such as the Chapman-Kolmogorov equations for determining the time-dependent state probability distribution for a Markovian queue [42]. Using the approximate fluid flow modelling approach proposed by Agnew [43], various dynamic models have been used by a number of researchers [42,44,45,46] to model a wide range of queueing and contention systems.
Using the flow conservation principle, for a single queue and assuming no losses, the rate of change of the average number of cells queued at the link buffer can be related to the rate of cell arrivals and departures by a differential equation of the form:

$$\dot{x}(t) = -f_{\text{out}}(t) + f_{\text{in}}(t)$$  \hspace{1cm} (1)

Where:

- $x(t)$ - state of the queue, given by the ensemble average of the number of cells $N(t)$ in the system (i.e. queue + server) at time $t$, i.e. $x(t)=E\{N(t)\}$
- $f_{\text{out}}(t)$ - ensemble average of cell flow out of the queue at time $t$
- $f_{\text{in}}(t)$ - ensemble average of cell flow into the queue at time $t$

The fluid flow equation is quite general and can model a wide range of queueing and contention systems as shown in the literature [42,45,44,46,47].

Assuming that the queue storage capacity is unlimited and the customers arrive at the queue with rate $\lambda(t)$, then $f_{\text{in}}(t)$ is just the offered load rate $\lambda(t)$ since no packets are dropped. The flow out of the system, $f_{\text{out}}(t)$, can be related to the ensemble average utilisation of the link $\rho(t)$ by $f_{\text{out}}(t)=C(t)\rho(t)$, where $C(t)$ is defined as the capacity of the queue server. We assume that $\rho(t)$ can be approximated by a function $G(x(t))$ which represents the ensemble average utilisation of the queue at time $t$ as a function of the state variable. Thus, the dynamics of the single queue can be represented by a non-linear differential equation of the form:

$$\dot{x}(t) = -G(x(t))C(t) + \lambda(t) , \quad x(0) = x_0$$  \hspace{1cm} (2)

Different approaches can be used to determine $G(x(t))$. A commonly used approach to determine $G(x)$ is to match the steady-state equilibrium point of (2) with that of an equivalent queueing theory model where the meaning of "equivalent" depends on the queueing discipline assumed. This method has been validated with simulation by a number of researchers, for different queueing models [42,45,44]. Other approaches, such as system identification...
techniques and neural networks, can also be used to identify the parameters of the fluid flow equation.

We illustrate the derivation of the state equation for an M/M/1 queue following [42]. We assume that the link has a First-In-First-Out (FIFO) service discipline and a common (shared) buffer. The following standard assumptions are made: the packets arrive according to a Poisson process; packet transmission time is proportional to the packet length; and that the packets are exponentially distributed with mean length 1. Then, from the M/M/1 queueing formulas, for a constant arrival rate to the queue the average number in the system at steady state is \( \lambda/(\lambda - \lambda) \). Requiring that \( x(t) = \lambda/(\lambda - \lambda) \) when \( x = 0 \), the state model becomes

\[
\dot{x}(t) = -\frac{x(t)}{1 + x(t)} C(t) + \lambda(t), \quad x(0) = x_0
\]

(3)

The validity of this model for packet based and circuit switched networks has been verified by a number of researchers, including [44,45].

Here we present an example for modelling ABR traffic competing with guaranteed traffic for the finite sever capacity in ATM based networks. Note that using similar arguments for packet based networks leads to the same fluid flow model [42]. We consider a series of \( M \), ATM output buffered switching nodes that route cells from a set of incoming links to a set of outgoing links. The ABR traffic is modelled as a one-way connection between an Origin-Destination (OD) pair spanning M switching nodes (see Figures 1). The nodes are connected by links \((1,\ldots,M)\), associated with deterministic propagation delays \( \tau_i \) \((i = 1,\ldots,M)\). Guaranteed traffic \((\lambda_1^g,\ldots,\lambda_M^g)\), appearing across every ATM switch, requires some access to the shared resources of each link. The queue at each link provides for statistical multiplexing of the incoming traffic streams.
Using fluid flow arguments, we can represent ABR traffic as a series of interconnected M/M/1 queues interfered by guaranteed traffic competing for the common resources. The ABR model, an extension of the M/M/1 queue model (1), is:

$$
\dot{x}_i(t) = -C_i(t) \left( \frac{x_i(t)}{1 + x_i(t)} \right) + \lambda_{ABR}^i(t) + \lambda_{eq}^i(t)
$$

$$
\dot{x}_i(t) = -C_i(t) \left( \frac{x_i(t)}{1 + x_i(t)} \right) + \gamma_{ABR}^i(t) + \lambda_{eq}^i(t) \quad i = 2, ..., M.
$$

where

$C_i(t)$ - bandwidth (capacity, cell service rate) allocated to the ABR traffic at node $i$,

$x_i(t)$ - state of the queue (i.e. ensemble average of number of cells in the queue) at node $i$,

$\lambda_{ABR}^i(t)$ - arrival rate due to ABR traffic

$\lambda_{eq}^i(t)$ - arrival rate at cell-queue $i$ due to Guaranteed traffic,

$\gamma_{ABR}^i(t)$ - ABR traffic entering node $i$, arriving from the previous node $i-1$, delayed by a deterministic amount $\tau_{i-1}$ due to the transmission propagation.

For $i = 2, ..., M - 1$

$$
\gamma_{ABR}^i(t) = \left( C_{i-1}(t - \tau_{i-1}) \frac{x_{ABR}^{i-1}(t - \tau_{i-1})}{1 + x_{ABR}^{i-1}(t - \tau_{i-1})} \right).
$$

where $x_{ABR}^i(t)$ is the ensemble average of the number of cell places in the buffer occupied by ABR traffic.

This model (4) can be used to represent all possible ABR traffic paths for any origin destination pair. The validity of using an M/M/1 queue to approximately describe the queueing delays with fixed packet length is discussed by Gerla et al [48]. Furthermore, in the following section we provide a simulative comparison.
3. **Simulative comparison of fluid flow model with discrete event based simulation**

In this section we provide a simulative comparison of the fluid flow model derived above with a discrete event cell based simulation (using OPNET) of an ATM switch fed by a bursty source (Fig. 2).

![Network Topology of simulated model](image)

**Figure 2 Network Topology of simulated model**

We compare the time evolution of the state of the queue system, as given by the solution of the fluid flow model presented above, with the state of the queue observed from the discrete event cell based simulation of an ATM switch. Both model and simulation consider the same input representing a bursty on-off source, shown in Figure 3 (general) and Figure 4 (actual Opnet output). The active and idle periods of the connection are shown in part (a), the packet activity in part (b), and the cell activity in part (c). The idle period has a geometric distribution with the mean value chosen to adjust the network load. During each active period a number of packets are generated with a geometric distribution and mean number of packets $N$. The packet size also has a geometric distribution with mean size of 8 kbytes, while the pause period is exponential with mean value equal to 0.5 msec. For ATM based networks, each packet is segmented into cells and transmitted at a constant cell rate of 342170 cells/sec.
The time evolution of the queue state from both the model and OPNET simulation are presented in Figure 5. From Figure 5 we can observe that there is a reasonable agreement between the proposed model and the observed one, which demonstrates confidence to the model.
Figure 5. Time evolution of network system queue state obtained using OPNET simulation (broken line) and solution of the fluid flow model (solid line)

Note that similar fluid flow models in both a discrete and continuous form have been used by a number of researchers for designing, or analysing the behaviour of network systems under control [49,11,17,13,50]. For example, Rohrs [11] using similar fluid flow arguments derived the following discrete fluid flow model of the state of the buffer at the output port of an ATM switch.

\[
x(nT) = \max\{x(n-1)T + (\lambda(n-1)T - C(n))T, 0\} + n_q
\]

(5)

where \( T \) is the sampling period, and \( n_q \) is a noise term representing the difference between the model and the actual queue system, and uses this model to evaluate the performance of a binary Backward Explicit Congestion Notification (BECN) control algorithm. He demonstrates the undesired cyclic behaviour of the controlled system. This (undesired) cyclic behaviour is also presented in [13] for TCP/IP, using dynamic models of the behaviour of the different phases of the TCP/IP congestion algorithms (slow start and congestion avoidance
phase) for high bandwidth-delay products and random loss. Their results are demonstrated using simulations. In [50] for ATM congestion control they make use of a similar model, as given by (5), and using intuition they design an ABR flow control strategy (referred to as queue control function) to keep the queue well controlled, according to some known function (step, linear, hyperbolic and inverse hyperbolic). They use analysis and simulation to evaluate the proposed strategy. It is worth noting that many other types of models have been proposed, either using queueing theory arguments, or others, but in most cases the derived models are too complex for deriving simple to understand and implement controllers. As a result, motivated by the desire to derive simple controllers, often the dynamic aspects of the network system are ignored. For example, in [51] the analysis of the performance of simple (binary) reactive congestion control algorithms is carried out using a queueing theory approach model, which is limited to steady state analysis only due to the inability to handle the resultant computational complexity for the dynamic case.

4. Conclusions

Based on the results derived from our simulations tests we argue that the fluid flow model appears reasonable to pursue further work on using it for designing non-linear congestion controllers. Combined with the fact that the model we use is based on simple fluidic arguments motivates us to extend our work further on investigating universal dynamic models for both TCP/IP and ATM. Future work includes the detailed evaluation of universal dynamic models derived from simple fluid flow arguments that capture the fundamental characteristics of network behavior. Our objective is to derive models that are simple, but accurate enough, for synthesizing simple to implement congestion controllers with proven analytic properties. The presented fluid-flow model and the simulative comparison presented in this paper suggest that this model is a good starting candidate.

REFERENCES:


