

On Network-Wide Packet Pacing

Tilman Wolf, Yong Huang, Weibo Gong and Yan Cai
Technical Report: TR-08-CSE-07
Department of Electrical and Computer Engineering
University of Massachusetts, Amherst, MA USA
{wolf,yhuang,gong,ycai}@ecs.umass.edu

Abstract—A major challenge in wide deployment of optical packet switches with small buffers in next-generation networks is the network-wide impact of occasional traffic bursts. We describe the design of an architecture that implements packet pacing such that suitable statistical properties of network traffic can be ensured and small-buffer networks can be operated efficiently. We introduce a mathematical model for tandem fluid queues with cross-traffic that allows us to analyze the burst impact on an entire network rather than a single buffer. We demonstrate – theoretically and through simulation – that a small amount of packet pacing deployed throughout the network can avoid traffic bursts and significantly improve network operation.

I. INTRODUCTION

As data communication has become increasingly important to our society, demands on the networking infrastructure have reached a point where it is desirable to implement an optical packet switching core. The prohibitive cost of large optical buffers has led to many next-generation network architecture designs based on optical routers with small buffers. A major challenge in this context is to envision efficient operation scenarios that justify these small buffer networks. For core routers with millions of flows it can be argued that the assumption of flow independence and the law of large numbers is sufficient [6], [7]. However, for access network routers with fewer flows and more correlated traffic efficient operation cannot solely rely on the smoothing effect of aggregation.

We present the design and analysis of an architecture that implements packet pacing on network links such that suitable statistical properties of network traffic can be ensured and small-buffer networks can be operated efficiently. Our analysis shows in this paper that this pacing feature is a crucial component of future networks as bursty sources that traverse multiple buffers quickly fill router buffers and cause poor network performance. We also demonstrate through analysis and simulation that the enforcement of a small amount of pacing at access points can avoid traffic bursts and significantly improve network operation.

We believe that addressing issues related to the theory, implementation, and deployment of a network-wide pacing architecture is particularly important at this point in time as we see four current developments in networks:

- **Move towards an optical networks:** Many efforts in the current Internet as well as in the next-generation Internet design aim at the use of optical networks. It is unlikely that large optical buffers will become practical and thus

network operations need to be designed around small buffer networks.

- **TCP dominates the transport layer:** The window-based operation of TCP causes bursty transmissions from most traffic sources. With developments of new flavors of TCP with larger window sizes, this burstiness increases.
- **Bandwidth dominates as traffic metric:** Throughout academia and industry, it is common to measure the impact of traffic on the network by quantifying the bandwidth of a transmission. As we show in our paper, packet pacing also has a significant impact and thus needs to be considered.
- **Design of next-generation Internet architecture:** Efforts in the networking community are aimed at designing the next-generation Internet architecture. Considering packet pacing will be a crucial aspect of this design and thus it is highly beneficial to study this topic in the context of network-wide deployment at this point in time.

These trends not only illustrate that pacing per se is an important aspect of network design and operation. They also show that it is important to consider packet pacing from a network-wide view point rather than from a single-buffer point of view as much of the prior work has done.

Apart from the general contribution of the design and detailed analysis of a pacing architecture, we derive the following key observations:

- **Traffic bursts cause global burstiness.** Traffic that shares a common queue can be impacted by bursts from a single source (e.g., TCP bursts). For example, evenly spaced traffic from one source can become bursty when a burst of packets from another flow fills a shared queue and decreases the spacing between packets. This effect can ripple downstream for all flows in the buffer and thus expand across the entire network.
- **Pacing helps globally.** Burstiness of traffic not only impacts the first buffer along a path, but can cause increasing queue lengths downstream as our analysis of tandem queuing systems shows. Cross-traffic can be affected by this burstiness and thus cause poor queuing behavior even on nodes that are off-path. Pacing eliminates these effects and thus is of global benefit for the network. We also observe that even small amounts of pacing are globally beneficial.
- **Indiscriminate pacing is effective.** It is not necessary

to pace traffic on a per-flow basis. Indiscriminate pacing is effective in achieving desirable network-wide traffic properties. This observation is important since traffic that paces itself voluntarily is at a disadvantage to unpaced traffic. Using indiscriminate pacing throughout the network can ensure that fairness is provided.

The paper is structured as follows. Section II discusses related work in the area of traffic pacing. Section III describes the overall pacing architecture. The main focus of the paper lies on Section IV, where the analysis of tandem queuing systems with and without cross traffic is derived and the impact of bursty traffic on a network of queues is explored. Section V presents simulation results that validate the analysis and show the benefits of packet pacing. Our contributions and results are summarized in Section VI.

II. RELATED WORK

An optical network infrastructure promises high bandwidth connectivity with data rates that vastly exceed the capacity of networks with electronic transmission links. However, the high cost of large optical buffers requires optical networks to be built on the premise of having only small (or even no) buffers. Recent studies suggest that for core routers with millions of flows small buffers suffice based on the assumption of flow independence and the law of large numbers [6], [7]. However, it is still an open question if the assumptions underlying the argument small buffer are realistic [5].

Rather than relying on desirable statistical properties of traffic to ensure efficient operation of the optical network, it is possible to pace traffic before it reaches the optical domain. By spacing packets evenly, burstiness is avoided and small buffers are more likely to absorb short-timescale variations in traffic. Sivaraman et al. [16], [17] have proposed such traffic conditioning at the optical edge based on the arrival curve of traffic. This approach is promising, but exhibits limitations in terms of the complexity of the pacing process and the performance requirements for pacing at the optical edge. They also propose a placement of the pacing component at the entrance of optical packet switching networks based on the analysis of the source arrival curve and the deadline curve. Our approach aims at pacing throughout the network, in particular in access networks where data rates are not as high and pacing can be implemented more efficiently.

In another major difference to the approach by Sivaraman et al. that bases pacing on algorithms using complex arrival curves [16], we employ a single parameter pacing algorithm that is easy to implement and easy to understand. This simplicity is important when considering deployment issues and when exploring complex game-theoretical interactions that can appear when different administrative entities attempt to optimize their benefits. These issues have arisen in the context of existing protocols (e.g., TCP unfriendly practices [11]).

Packet pacing has also been studied in the context of video streaming; see for example [3], [4], [14], [15]. More recently, pacing has been explored as a tool to reduce the burst impact caused by TCP to optical packet switching routers [1], [6],

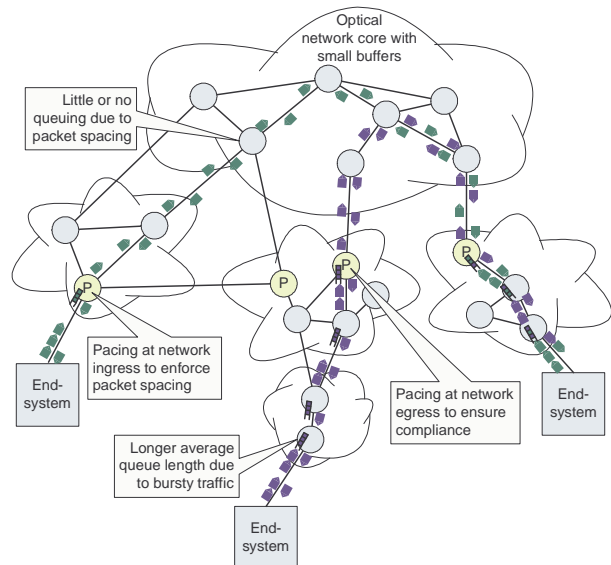


Fig. 1. Pacing Architecture.

[16]–[18] and optical burst switching (OBS) networks [12]. These studies have focused on a subset of all traffic and the case of a single bottleneck in the end-to-end path. While these studies point out the importance of considering spacing as a traffic smoothing technique, they do not make a convincing case for the significance of pacing to the operation of the entire Internet. We attempt to address this network-wide issue in this paper, where we not only consider multi-buffer networks, but also the impact of cross-traffic.

There is a large amount of research work on network traffic statistics and their impact on queues. Our study is focused on the impact of small time scale burstiness on tandem and networked queues. Our observation that such burstiness could impact all queues in the flow path in a linear way seems to be new and important.

III. PACING ARCHITECTURE

We begin our discussion of pacing from a systems point of view. In this section, we discuss the design of a pacing architecture for networks. This architecture provides a background and context for the queuing analysis in Section IV and the simulation results in Section V.

A. Pacing Nodes

It is neither necessary nor feasible to implement packet pacing on every node in the network. As our later results show, pacing can benefit nodes that are multiple hops downstream and thus does not need to be performed at every node. Therefore, we envision that pacing nodes are placed only at a few locations throughout the network. This design is illustrated in Figure 1, where a ‘P’ indicates a pacing node.

Each pacing node has two basic functionalities to identify traffic that could cause bursts and avoid propagating these bursts downstream:

- **Monitoring of Packet Spacing:** In order to determine if traffic does or does not meet pacing requirements, it is necessary for the pacing node to monitor incoming traffic. It can then determine if the outgoing aggregate does or does not adhere to necessary packet spacing requirements. One metric that can be monitored for this purpose is the auto-correlation time constant that we introduce in Section IV-A2 for on-off traffic models.
- **Active Pacing:** When traffic does not meet the required packet spacing, packets are buffered and transmitted with increased inter-packet times to generate a less bursty pattern. In general (when considering the operation of the network in its entirety), it is sufficient to perform this pacing indiscriminately on all traffic. However, in order to ensure fairness among flows, this process can be extended to perform active pacing on a selected subset of traffic. Flows that show particularly bursty patterns could have their packets delayed before flows that adhere to pacing requirements.

B. Control of Pacing and Incentives

With the capabilities to identify when traffic is bursty and to pace this traffic, a pacing node can be used for two purposes:

- **Compliance:** A pacing node can be used to check that certain traffic meets the requirements for packet spacing and space packets more evenly if necessary. This capability is important on egress links of a network to ensure that traffic complies to the desired level of packet spacing.
- **Enforcement:** A pacing node can also be used to enforce packet spacing on ingress links to a network. In this case, traffic that does not comply is buffered and spaced.

While the two functions of compliance and enforcement are very similar on the data plane (i.e., check burstiness and buffer if necessary), the resulting control plane interactions are very different. Compliance to pacing requirements by all nodes in the network is the ideal scenario, but as in any distributed system that is not managed by a single authority, it is important to consider the case where nodes do not comply.

While the design of a detailed incentive system is beyond the scope of this paper, it can be envisioned that such an approach requires a game-theoretical consideration of different interest groups that are involved in generating and forwarding traffic. The principle idea could follow the process that is used in today's power grid. While most consumers are familiar with being charged for the amount of energy used (i.e., bandwidth used in the network context), power companies also consider so-called "reactive power" (due to the phase shift between voltage and current) when transmitting electrical power in power grid. This reactive power is comparable to traffic bursts in the network. In power transmissions, users (e.g., a factory) often need to install reactive power compensators to reduce reactive power that would cause current to flow back and forth over the power grid and waste considerable amounts of "bandwidth" on the power grid [9]. Such a compensator is comparable to a pacing node that smoothes out any potential traffic bursts that could be transmitted. As indicated earlier, it

is not sufficient that pacing be limited to the end-systems only as a single bursty source can cause ripple effects throughout the network. Thus, pacing nodes need to be placed *inside* the network.

If pacing nodes with the above capabilities are deployed throughout the network, traffic bursts can be avoided. Why this provides a significant benefit for the operation of the network is analyzed in detail in the following section.

IV. QUEUING BEHAVIOR IN TANDEM QUEUES

In this section, we present the mathematical foundations on why pacing is indeed an effective mechanism that reduces burstiness network-wide and thus makes it feasible to operate small-buffer at a high level of efficiency.

We first show the impact of bursty sources on a sequence of buffers, called "tandem queues." The analysis requires the use of Poisson Counter Drive Stochastic Differential Equations to develop a fluid tandem queuing model. We then expand the concept of tandem queues to also consider cross-traffic. We first represent cross-traffic with an on-off model and then generalize it to any type of traffic. Finally, we illustrate our result with a more visual representation of the impact of bursty traffic on networks.

A. Impact of Source Bursts on Network

The impact of bursty traffic on network buffers has been studied in the context of a single queue, but not in the context of the entire network. This is partly due to the difficulty in dealing with analytic models of networks of queues with correlated arrivals. In this section, we show an approach to dealing with this problem and obtaining analytic models for tandem queues with on-off sources.

1) *Poisson Counter Driven Stochastic Differential Equations:* First, we introduce a mathematical method for the analysis of the packet spacing impact in networks. Our method is based on the calculus of Poisson Counter Drive Stochastic Differential Equations (PCSDE). This calculus combines the independent increment property of the Poisson process (or Poisson Counter) with ideas from differential equations [2]. We show that it is a powerful tool in dealing with correlated traffic passing through a sequence of queues.

The basic concepts of PCSDE are as follows. Consider a stochastic integral equation

$$x(t) = x(0) + \int_0^t f(x(\tau), \tau) d\tau + \int_0^t g(x(\tau), \tau) dN_\tau, \quad (1)$$

where N_τ is a Poisson counter and $g()$ is a differentiable function. The solution of Equation 1 is defined as follows:

Definition: $x(\cdot)$ is a solution of Equation 1 in the Itô sense if, on an interval where N is constant, x satisfies $\dot{x} = f(x, t)$ and if, when N jumps at t_1 , x changes according to

$$\lim_{t \rightarrow t_1^+} x(t) = g \left(\lim_{t \rightarrow t_1^-} x(t), t_1 \right) + \lim_{t \rightarrow t_1^-} x(t) \quad (2)$$

and $x(\cdot)$ is taken to be continuous from the left. Equation 1 is often written as

$$dx(t) = f(x, t)dt + g(x)dN \quad (3)$$

and is called the Poisson Counter Driven Stochastic Differential Equation (PCSD E).

Some direct consequences of the definition above are the following. Consider a stochastic differential equation driven by n independent Poisson counters N_1, \dots, N_n :

$$dx = f(x)dt + \sum_{i=1}^n g_i(x)dN_i, \quad x \in R^n. \quad (4)$$

A significant point is that, due to the fact that $x(t)$ is continuous from the left and the Poisson counter is taken to be continuous from the right, we have

$$\frac{d}{dt}E[x(t)] = E[f(x(t))] + \sum_{i=1}^n E[g_i(x(t))]\lambda_i, \quad (5)$$

where λ_i is the rate for counter N_i .

Also, we have the following ‘‘Itô rule:’’ If $\psi : R^n \rightarrow R$ is a differentiable function, then

$$d\psi(t) = \left\langle \frac{\partial \psi}{\partial x}, f(x) \right\rangle dt + \sum_{i=1}^n [\psi(x(t) + g_i(x(t))) - \psi(x(t))] dN_i. \quad (6)$$

2) *A Fluid Tandem Queuing Model:* We now turn to the application of PCSD E to the study of the impact of bursts on networks. We start with considering a tandem fluid model that consists of a series of intermediate servers in row. We show that the PCSD E is capable of establishing an explicit expression relating the average queue length to the burstiness of input traffic. For clarity we first consider a single on-off source and then present results for multiple on-off sources at the end of this subsection. Our solutions reveal important relations between the source correlation and the backlog in *all* the queues in routers. Figure 2 illustrates a typical tandem queue network with one Markov on-off source.

The sample path description for the single source tandem queue is

$$\begin{cases} dx(t) &= (1 - x(t))dN_1 - x(t)dN_2 \\ dv_1(t) &= -c_1\mathbf{I}(v_1)dt + c_0x(t)dt \\ dv_2(t) &= -c_2\mathbf{I}(v_2)dt + c_1\mathbf{I}(v_1)dt \\ dv_3(t) &= -c_3\mathbf{I}(v_3)dt + c_2\mathbf{I}(v_2)dt \\ &\vdots \\ dv_n(t) &= -c_n\mathbf{I}(v_n)dt + c_{n-1}\mathbf{I}(v_{n-1})dt, \end{cases} \quad (7)$$

where $\mathbf{I}(v_i)$ denotes the corresponding indicator function, taking on the value ‘1’ for $v_i \geq 0$. To avoid trivialities we assume $c_0 > c_1 > c_2 > \dots > c_n > 0$. Note that if a node has larger capacity (for the concerned flow) than the upstream ones it can be neglected in the analysis without loss of generality. Stability demands that $c_n \geq c_0E[x(t)]$.

We omit the solution procedures due to space constraints, but state that by combining the modeling power of Poisson

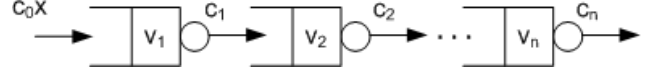


Fig. 2. A Typical Tandem Queue Network.

counters and differential equations we are able to obtain an explicit solution for the tandem fluid queue:

$$\left(\frac{\lambda_1}{\lambda_1 + \lambda_2} c_0 - c_n \right) E[v_n] + (c_{n-1} - c_n) \cdot \left(E[v_{n-1}] + \dots + E[v_1] + \frac{E[c_0x]}{\lambda_1 + \lambda_2} \right) = 0. \quad (8)$$

To reveal the structure of this formula we introduce some notation. We call $r = E[c_0x]$ the arrival rate, $\rho_k = (c_{k-1} - r)/(c_k - r)$ the k^{th} excess capacity ratio, $\tau = 1/(\lambda_1 + \lambda_2)$ the source auto-correlation time constant. Now we can write

$$\begin{aligned} E[v_n] &= (\rho_n - 1)(E[v_{n-1}] + E[v_{n-2}] + \dots + E[v_1] + \tau r) \\ &= \tau r (\rho_n - 1) \rho_{n-1} \rho_{n-2} \dots \rho_1. \end{aligned} \quad (9)$$

This leads to the summation formula

$$E[v_n + v_{n-1} + \dots + v_1] = \tau r \left(\frac{c_0 - r}{c_n - r} + 1 \right). \quad (10)$$

This formula is significant since it tells us that if the arrival rate $E[c_0x]$ is kept constant, say both λ_1 and λ_2 are simultaneously reduced to half, then the time constant for the auto-correlation function for the arrival process $x(t)$, $1/(\lambda_1 + \lambda_2)$, will double, and consequently *linearly* affect the backlog at *every* queue! In other words, if one breaks the arrival burst into smaller pieces (e.g., through pacing), it would significantly reduce the congestion of the network.

The above results can be generalized to tandem queues with multiple, inhomogeneous Markov on-off sources. For example, if there are M Markov on-off sources that feed into an n -stage tandem queue, we have

$$\left(c_n - \sum_{k=1}^n \frac{c_{0k} \lambda_{i1}}{\lambda_{i1} + \lambda_{i2}} \right) E[v_n] = (c_{n-1} - c_n) \left(E[v_n] + E[v_{n-1}] + \dots + E[v_1] + \sum_{i=1}^M \frac{1}{\lambda_{i1} + \lambda_{i2}} E[c_{0i}x_i] \right). \quad (11)$$

It can be seen that if every source reduces the auto-correlation time constant by half and keeps the rates fixed, the average backlog at every queue will also be reduced by half. Results for more complex systems are omitted here but can be obtained from [2], [8].

The explicit expressions of a tandem queue length indicates that even with the same input traffic rate, the average queue length increases as the auto-correlation time constant goes up. This observation is perfectly in accordance with the well-acknowledged observation that the burstiness of network traffic is able to impact the performance of intermediate routers. Note that although we only show the impact on the average queue length, it can be seen from later discussion that the impact also applies to the instantaneous queue length, and thus affects packet loss on routers with limited buffers.

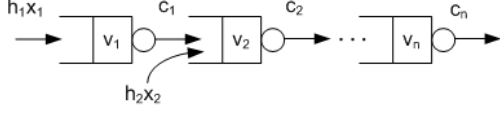


Fig. 3. A Tandem Queue Network with Markov On-Off Cross-Traffic.

B. Tandem Queue Networks with Cross Traffic

With an understanding of the impact of source bursts, we further analyze the impact of cross-traffic on a tandem queue system. We are interested in how burst of cross traffic affects the average queue length. Using the stochastic differential equations, we obtain an explicit expression of the average queue length with generalized cross-traffic at steady state.

1) *Markov On-Off Cross Traffic*: To introduce our notation, consider Figure 3, which shows a tandem queue network with a Markov on-off source and Markov on-off cross traffic. Source 1 flows into the first queue and the cross traffic flows into the second queue. Variables x_1 and x_2 present the state of each source, where $x_i = 1$ indicates that source i is on and $x_i = 0$ means that it is off. The two sources have peak rate of h_1 and h_2 , respectively. The Poisson rate of on and off states are λ_{i1} and λ_{i2} , where $i = 1, 2$. The system dynamics are described by following stochastic differential equations:

$$\begin{cases} dx_1(t) &= (1 - x_1(t))dN_{11} - x_1(t)dN_{12} \\ dx_2(t) &= (1 - x_2(t))dN_{21} - x_2(t)dN_{22} \\ dv_1(t) &= -c_1\mathbf{I}(v_1)dt + h_1x_1(t)dt \\ dv_2(t) &= -c_2\mathbf{I}(v_2)dt + c_1\mathbf{I}(v_1)dt + h_2x_2(t)dt \\ dv_3(t) &= -c_3\mathbf{I}(v_3)dt + c_2\mathbf{I}(v_2)dt \\ &\vdots \\ dv_n(t) &= -c_n\mathbf{I}(v_n)dt + c_{n-1}\mathbf{I}(v_{n-1})dt \end{cases} \quad (12)$$

We make two assumptions similar to those in previous section. The first assumption is that the capacities of queues are decreasing and are always less than the peak rate of sources: $h_1 > c_1 > c_2 > \dots > c_n > 0$ and $h_2 > c_2 > \dots > c_n > 0$ (capacities that do not follow this decreasing order can be ignored). The second assumption is that the average throughput of all flows into each queue should be less than the capacity: $E[h_1x_1] + E[h_2x_2] < c_n$. We can then obtain steady state solutions:

$$\begin{aligned} E[v_1] &= \left(c_1 - E[h_1x_1]\right)^{-1} \frac{h_1 - c_1}{\lambda_{11} + \lambda_{12}} E[h_1x_1] \quad \text{and} \\ E[v_2] &= \left(c_2 - E[h_1x_1] - E[h_2x_2]\right)^{-1} \\ &\quad \left[\left(c_1 + E[h_2x_2] - c_2\right) E[v_1] + \frac{c_1 + E[h_2x_2] - c_2}{\lambda_{11} + \lambda_{12}} E[h_1x_1] \right. \\ &\quad \left. + E[h_2x_2] \left(\frac{h_2 + E[h_1x_1] - c_2}{\lambda_{21} + \lambda_{22}} \right) \right]. \end{aligned} \quad (13)$$

More generally, the average queue length of queue n ($n > 2$)

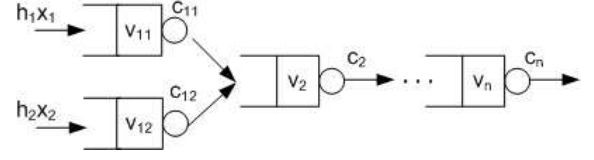


Fig. 4. A Tandem Queue Network with More General Cross-Traffic.

at steady state is:

$$\begin{aligned} E[v_n] &= \left(c_n - E[h_1x_1] - E[h_2x_2]\right)^{-1} (c_{n-1} - c_n) \\ &\quad \left(\frac{E[h_1x_1]}{\lambda_{11} + \lambda_{12}} + \frac{E[h_2x_2]}{\lambda_{21} + \lambda_{22}} + E[v_1] + E[v_2] + \dots + E[v_{n-1}] \right). \end{aligned} \quad (14)$$

If we halved λ_{11} , λ_{12} and λ_{21} and λ_{22} , the average throughput of Source 1, $E[h_1x_1]$ and source 2, $E[h_2x_2]$ would be same. However, $E[v_1]$ would be doubled. Also, according to Equations (13), and (14), $E[v_2]$ and $E[v_n]$ would all be doubled as well.

2) *More General Cross Traffic*: We now extend our analysis to a system with more general cross traffic shown in Figure 4. The cross traffic goes through a queue with length $v_{12}(t)$ before injecting into the tandem queue network.

This system can be formulated as:

$$\begin{cases} dx_1(t) &= (1 - x_1(t))dN_{11} - x_1(t)dN_{12} \\ dx_2(t) &= (1 - x_2(t))dN_{21} - x_2(t)dN_{22} \\ dv_{11}(t) &= -c_{11}\mathbf{I}(v_{11})dt + h_1x_1(t)dt \\ dv_{12}(t) &= -c_{12}\mathbf{I}(v_{12})dt + h_2x_2(t)dt \\ dv_2(t) &= -c_2\mathbf{I}(v_2)dt + c_{11}\mathbf{I}(v_{11})dt + c_{12}\mathbf{I}(v_{12})dt \\ dv_3(t) &= -c_3\mathbf{I}(v_3)dt + c_2\mathbf{I}(v_2)dt \\ &\vdots \\ dv_n(t) &= -c_n\mathbf{I}(v_n)dt + c_{n-1}\mathbf{I}(v_{n-1})dt. \end{cases} \quad (15)$$

We can solve steady state $E[v_{11}]$, $E[v_{12}]$, $E[v_2]$, \dots , $E[v_n]$ with similar assumptions as in previous sections: the decreasing capacity assumption, $h_1 > c_{11} > c_2 > \dots > c_n > 0$ and $h_2 > c_{22} > c_2 > \dots > c_n > 0$, and the assumption that the average flow rate into a network should be less than the capacity of the bottleneck link, $E[h_1x_1] + E[h_2x_2] < c_n$.

The steady state solutions are:

$$\begin{aligned} E[v_{11}] &= \frac{h_1 - c_{11}}{\lambda_{11} + \lambda_{12}} E[h_1x_1] \left(c_{11} - E[h_1x_1]\right)^{-1}, \\ E[v_{12}] &= \frac{h_2 - c_{12}}{\lambda_{21} + \lambda_{22}} E[h_2x_2] \left(c_{12} - E[h_2x_2]\right)^{-1}, \text{ and} \\ E[v_2] &= \left(c_2 - E[h_1x_1] - E[h_2x_2]\right)^{-1} \\ &\quad \left[\left(c_{11} + E[h_2x_2] - c_2\right) \left(\frac{E[h_1x_1]}{\lambda_{11} + \lambda_{12}} + E[v_{11}] \right) + \right. \\ &\quad \left. \left(c_{12} + E[h_1x_1] - c_2\right) \left(\frac{E[h_2x_2]}{\lambda_{21} + \lambda_{22}} + E[v_{12}] \right) \right]. \end{aligned} \quad (16)$$

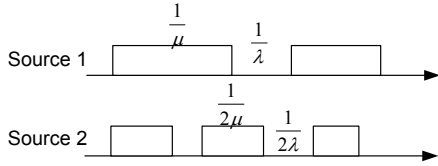


Fig. 5. Two Sources with Different On-Off Rates.

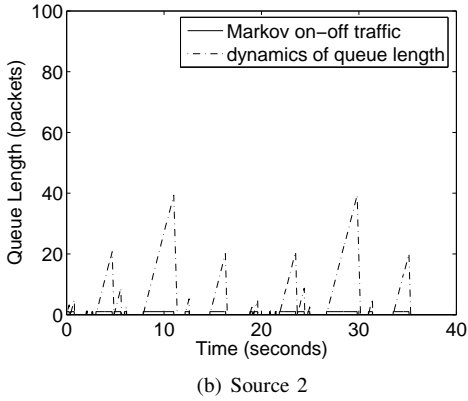
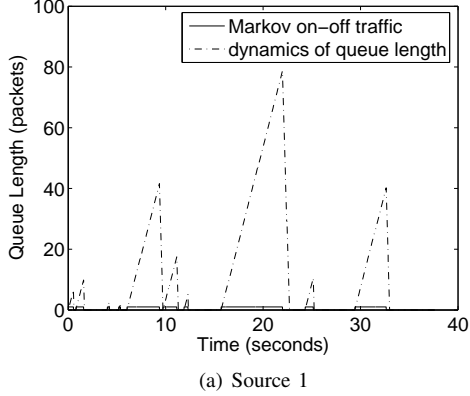


Fig. 6. Comparison of Queue Length for Different Sources. The queue length caused by Source 1 is compressed along the x-axis and along the y-axis for Source 2 indicating lower buffer requirements for spaced packet transmissions.

For $n > 2$, we have:

$$E[v_n] = (c_n - E[h_1x_1] - E[h_2x_2])^{-1} (c_{n-1} - c_n) \left(\frac{E[h_1x_1]}{\lambda_{11} + \lambda_{12}} + \frac{E[h_2x_2]}{\lambda_{21} + \lambda_{22}} + E[v_{n-1}] + E[v_{n-2}] + \dots + E[v_2] + E[v_{11}] + E[v_{12}] \right) \quad (17)$$

The analytic results show that given the same average throughput of $E[h_1x_1]$ and $E[h_2x_2]$, if the auto-correlation time constants are doubled, the average queue length at every stage will be doubled, too.

C. A Visualization of Pacing Impact for an Entire Network

The above analysis yields an important relationship between burstiness in traffic and queue length on routers. The results are not completely general due to some assumptions on the type of sources and the tandem topology. To provide an intuition

that the conclusions are also valid for a more general case, we introduce a graphical explanation of the above results. This helps us to understand the implications for any type of network.

Consider the first buffer in a series of tandem buffers. We compare two sources with equal average bandwidth but different levels of burstiness as illustrated in Figure 5. Source 1 is burstier due to longer on and longer off periods. Source 2 can be considered “paced.” Specifically, we assume $\lambda_1 = 1/2s^{-1}$, $\lambda_2 = 1/2s^{-1}$, $c_0 = 1.0\text{Mbps}$ and $c_1 = 0.9\text{Mbps}$ with UDP packets of 1000 bytes to obtain concrete simulation results.

Figure 6 shows the instantaneous queue length for both sources. Figure 6(a) shows higher queue lengths due to longer on periods in Source 1. The paced source in Figure 6(b) yields much lower queue lengths due to shorter bursts. To quantify the comparison, one can observe that Figures 6(a) and 6(b) are geometrically similar to each other. The three dominating peaks in Figure 6(a) (at times 9s, 22s, and 33s) appear also in Figure 6(b) (at times 4.5s, 11s, and 16.5s). Apart from this $2\times$ compression along the x-axis, a $2\times$ compression along the y-axis is also observable. This compression is due to shorter bursts in Source 2, which yields in turn shorter queue lengths. This observation of compression actually holds for general arrival processes and general network topologies, and thus is a fundamental argument for assessing the impact of burstiness and pacing.

The compression of 1:4 between Figures 6(a) and 6(b) should come as no surprise since we assume infinite buffer size. This makes the queueing system a “linear” system and thus to obtain the queue length curve in Figure 6(b) from the one in Figure 6(a) we only need to double the unit for both axis.

Several important comments are in order here. First, the above observation of proportionality does not depend on the fact that the traffic is on-off, it does not depend on the Markov assumption either. Second, this proportionality does not depend on the stationarity of the traffic. Third, the proportionality observation would hold for many queues. Forth, the queue length curves in the figures are the instantaneous queue length and our analytical results about the average queue length proportionality are only a minor consequence of the current observation.

In this visualization of the burstiness impact, the only crucial assumption that does not match with reality is the infinite buffer size. However when buffer size is limited and losses occur, the behavior changes. For UDP traffic, the impact is to remove the peaks of the queue length curve when they exceed the buffer size. For TCP traffic, the behavior is more complicated due to the retransmission. One can argue that the result for TCP traffic similar to that in Figure 6(a) would do even worse than traffic similar to that in Figure 6(b) due to more retransmission and the possibility of causing synchronization among many TCP sources sharing a common bottleneck [13]. Packet pacing would reduce the overall delay which is beneficial to TCP flows.

This visualization of the impact of traffic bursts on the Inter-

net operation gives us a principle concept of how significantly packet pacing would help. It also illustrates that any sort of packet pacing at any location would help. It is not required that pacing happens at the entry point to the optical network as suggested in [16]. It could be at the end system, the access network, or anywhere else. For example, being able to perform pacing close to the source where data rates are lower and transmissions are still performed in the electronic domain is much more practical to implement than pacing traffic in the optical domain.

V. SIMULATION RESULTS

In this section, we present ns-2 simulation [10] results that validate our analysis and compare the performance of networks with pacing nodes with that of networks that do not employ pacing.

A. Validation

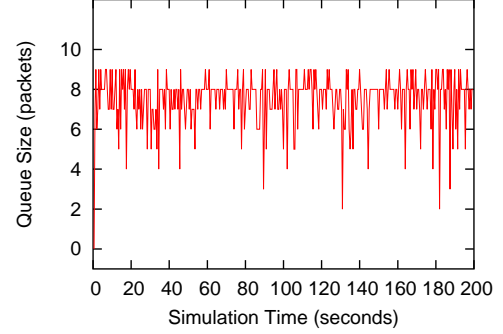
To validate the analytical results presented in previous section, we simulate a network with 3 tandem queues and cross traffic, illustrated in Figure 4. The link capacities are set to $c_{11} = c_{12} = 850\text{kbps}$, $c_2 = 800\text{kbps}$, and $c_3 = 700\text{kbps}$. Two Markov on-off sources generate traffic with peak rates of $h_1 = h_2 = 900\text{ kbps}$ and an average utilization of $E[x_1] = E[x_2] = 0.3333$. The auto-correlation time constant for both sources varies from 0.0333 second to 0.0833 seconds. Figure 7 compares the analytic result with simulations for queues over all stages. These results validate that our analytic results match simulation results. They also highlight one of our main observation: the average queue length increases linearly with the auto-correlation time constant and with every stage of the tandem queue network with the presence of general cross traffic.

B. Simulation of Network-Wide Pacing

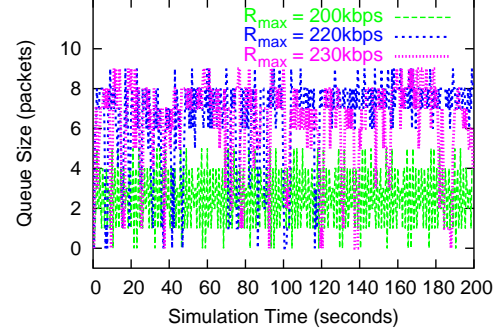
The functionality of the pacing node that we use in this simulation study is such that the pacer for each source enforces a maximum transmission rate R_{max} . If a node transmits faster than this maximum rate, the pacer holds packets that exceed the permissible rate in an infinite internal buffer and transmit them with a delay. This approach is conceptually similar to a leaky bucket that enforces a maximum rate, but packets that do not meet the requirements are not dropped but delayed. If the transmission rate is lower than R_{max} , the pacer does not change the traffic pattern. In a deployed network, the value of R_{max} would be determined via the control plane (e.g., through a compliance-enforcement relationship between peering nodes). The development of a more detailed pacing algorithm is the scope of future work.

We are interested in how this simple pacing technology affect the performance of the whole network and the individual sources in terms of the following metrics:

- **Average Queue Length in Every Stage:** As illustrated by the analytical results, busty traffic affects the average queue length of every stages. We show how pacing can reduce the average queue length.



(a) Instant Queue Length Without Pacing.



(b) Instant Queue Length With Pacing at Different Rates.

Fig. 9. Comparison of Instant Queue Length at Bottleneck Link.

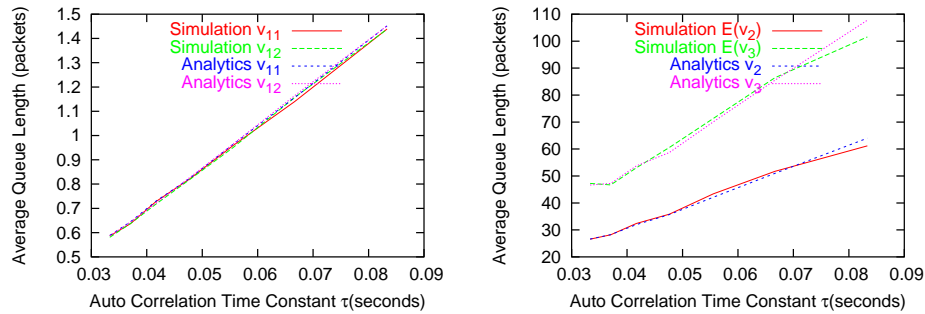
- **Cumulative Dropped Packets:** In practical networks, every queue has a finite size and packet drops occur when the buffer overflows. Especially for networks with small buffers, the number of dropped packets is an important metric, because it illustrates the inefficiency of the network operation (i.e., wasted effort by upstream nodes to transmit packets that eventually get dropped).
- **Throughput of Individual Sources:** Pacing smoothes the transmission from sources and caps their peak rate. It is important to consider the performance that is obtained by paced sources to explore efficiency and fairness issues.

We simulate a tandem queue network with the presence of non-cooperative cross-traffic shown in Figure 8. All links have a bandwidth of 10 Mbps and three CBR flows with data rate 7 Mbps, 8 Mbps, and 9 Mbps reduce the available bandwidth on each link. Five TCP sources are considered.

Figure 9 compares the instant queue length between unpaced and paced traffic. Figure 9(a) shows that the burstiness of unpaced traffic causes a mostly full queue with occasional overflows and underflows. Figure 9(b) shows that pacing reduces the oscillation of instant queue length.

Figure 10 shows a direct comparison of queue length moving average. Figure 11 compares the cumulative dropped packet. The finite buffer size causes a large number of dropped packets for unpaced traffic. As expected, pacing dramatically reduces the number of dropped packets.

Figure 12 compares the cumulative throughput of the five TCP sources. Figure 12(b) and 12(c) show the cumulative



(a) Average Queue Length at First Queue for Both Types of Traffic. (b) Average Queue Length at Second and Third Queue.

Fig. 7. Comparison of Simulation Results to Analytic Results for Tandem Queues with Cross-Traffic.

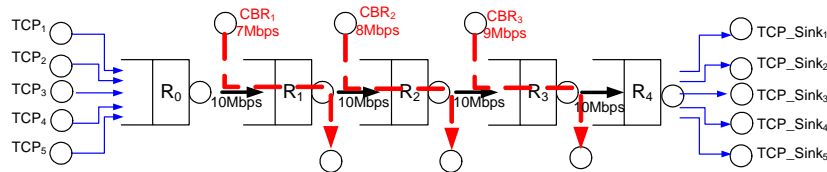


Fig. 8. Simulated Tandem Queue System with Cross-Traffic.

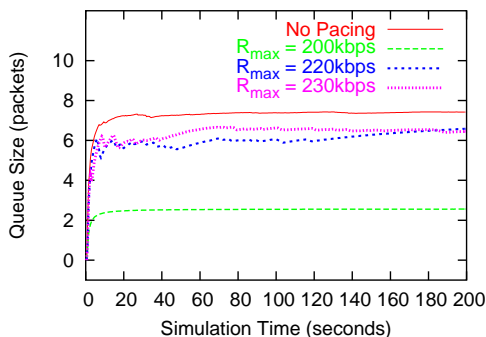


Fig. 10. Comparison of Queue Length Moving Average at Bottleneck Link.

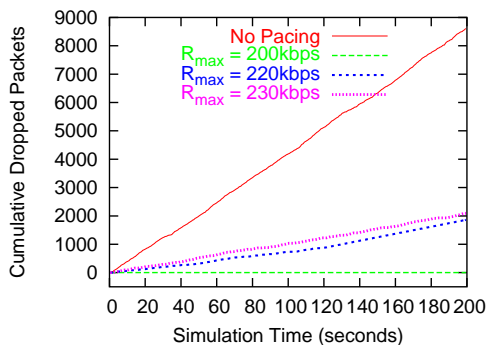


Fig. 11. Cumulative Dropped Packets at Bottleneck Link (With Cross-Traffic).

throughput when R_{max} is 220kbp and 200kbps, respectively. In both cases, the throughput of individual TCP source converges to a steady point quickly due to the competition with cross-traffic. However, average TCP throughput is less than that of no pacing since the sources are less competitive due to pacing.

To explore this issue of competitiveness further, Table I lists a summary of metrics at a finer resolution than shown in the previous figures. The average TCP throughput is calculated as the average of steady throughput among the five TCP sources. The metrics of average queue length and packing drop rate clearly demonstrate the benefit of pacing. The average queue length with pacing is either smaller or at the same level as that of unpaced sources. At the same time, pacing dramatically reduces the packet drop rate. As R_{max} increases, the average TCP throughput also increases, but is always smaller than that of unpaced traffic. This result demonstrates the inherent tension between optimizing network operation versus optimizing the performance of a source. On one hand, pacing helps the operation of networks as it reduces queue lengths and avoids transmissions of packets that later get dropped. On the other hand, pacing puts a cap on the maximum rate at which a node can transmit and thus weakens its competitive power when traversing a node with limited buffer space. In our simulation, the five TCP sources are competing with CBR3 for bandwidth on the bottleneck link. As R_{max} decreases, the throughput of CBR3 increases. This illustrates that it is important to implement pacing throughout the network to ensure that an aggressively transmitting node does not take advantage of other traffic that chooses to comply with pacing (or has already been forced through a pacing node upstream). This pacing can be done indiscriminately or on a per-flow basis. As the queue length and drop rate shows, either approach is beneficial to

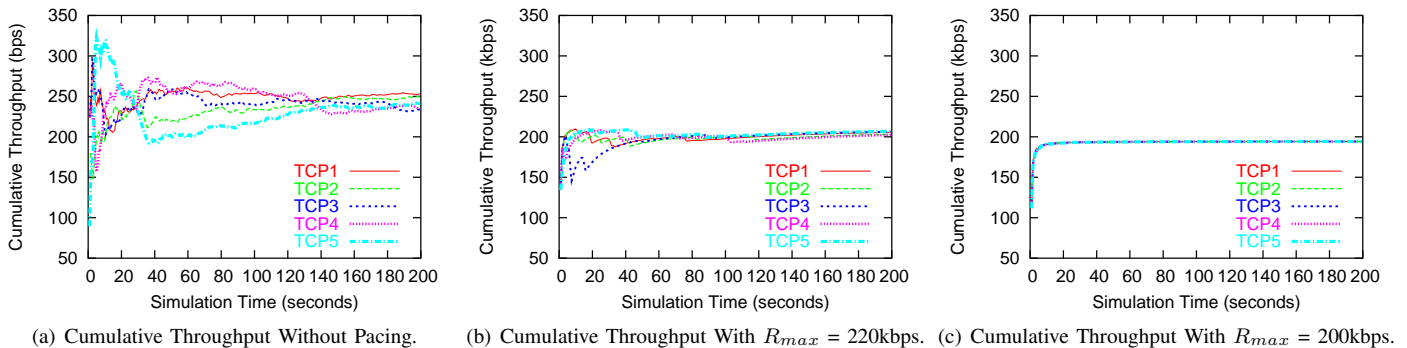


Fig. 12. Cumulative Throughput of Five TCP Sources with Pacing and Different R_{max} .

TABLE I
COMPARISON OF AVERAGE QUEUE LENGTH, PACKET DROP RATE, AVERAGE TCP THROUGHPUT, AND CROSS-TRAFFIC THROUGHPUT (WITH CROSS-TRAFFIC).

Pacing R_{max}	Avg. Queue Length (pkt)			Packet Drop Rate (pkt/sec)			Avg. TCP (kbps)	CBR3 (kbps)
	Queue 1	Queue 2	Queue 3	Queue 1	Queue 2	Queue 3		
200kbps	0.28	0.84	2.58	0.0	0.0	0.0	199.8	8955.0
220kbps	0.28	0.51	6.63	0.0	0.0	9.14	204.1	8915.0
230kbps	0.23	0.45	6.86	0.0	0.0	10.56	209.7	8891.2
Unpaced	0.21	0.83	7.48	0.01	0.07	43.80	252.3	8713.5

networks operation.

VI. SUMMARY AND CONCLUSIONS

In summary, this paper presents the design and analysis of a network-wide approach to pacing traffic. The motivation for this work lies in the design of optical networks, which can only buffer a few packets. Instead of relying on traffic statistics that may change as network use changes, we propose to actively monitor and pace traffic to ensure suitable queuing behavior.

Our analysis of tandem queue networks expands the study of pacing effects on router queues beyond the consideration of a single buffer. We also show how cross-traffic can be tied into our model to consider network-wide effects. The simulation results that we present validate our analytical approach and give us further insight into the effects of pacing. We show that pacing improves the operation of the network by shortening queues in routers and reducing packet drops. Even if pacing is not performed at the exact fair share of the bottleneck link rate, these benefits can be observed. Since we also show that there is an incentive not to pace one's own traffic, we conclude that a network-wide deployment of pacing nodes is necessary to avoid that certain sources can game the system.

We believe that our analysis and simulation results provide insights into the effects of pacing on a network-wide level. It is possible to deploy such an architecture in the current Internet and in future network architectures. Specifically, network-wide pacing can help in making the operation of small-buffer optical networks feasible and efficient.

REFERENCES

- [1] A. Aggarwal, S. Savage, and T. Anderson. Understanding the performance of TCP pacing. In *Proc. of IEEE INFOCOM 2000*, pages 1157–1165, Tel Aviv, Israel, Mar. 2000.
- [2] R. W. Brockett, W. Gong, and Y. Guo. Stochastic analysis for fluid queueing systems. In *Proceedings of the 38th IEEE Conference on Decision and Control (CDC)*, pages 3077–3082, Phoenix, AZ, Dec. 1999.
- [3] G. Cao, W.-c. Feng, and M. Singhal. Online variable-bit-rate video traffic smoothing. *Computer Communications*, 26(7):639–651, May 2003.
- [4] R.-I. Chang, M.-C. Chen, J.-M. Ho, and M.-T. Ko. An effective and efficient traffic smoothing scheme for delivery of online VBR media streams. In *Proceedings of the Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '99)*, pages 447–454, New York, NY, Mar. 1999.
- [5] A. Dhamdhere and C. Dovrolis. Open issues in router buffer sizing. *SIGCOMM Computer Communication Review*, 36(1):87–92, Jan. 2006.
- [6] M. Enachescu, Y. Ganjali, A. Goel, N. McKeown, and T. Roughgarden. Part III: Routers with very small buffers. *SIGCOMM Computer Communication Review*, 35(3):83–90, July 2005.
- [7] M. Enachescu, Y. Ganjali, A. Goel, N. McKeown, and T. Roughgarden. Routers with very small buffers. In *Proceedings of the Twentyfifth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2006)*, Barcelona, Spain, Apr. 2006.
- [8] W. Gong, D. Towsley, Y. Wu, and D. Ye. A stochastic differential equation model for queues with traffic smoothing. *Proceedings of the 42nd Annual Allerton Conference on Communication, Control and Computing*, Oct. 2002.
- [9] S. Hao and A. Papalexopoulos. Reactive power pricing and management. *IEEE Transactions on Power Systems*, 12(1):95–104, Feb. 1997.
- [10] LBNL, Xerox PARC, UCB, and USC/ISI, <http://www.isi.edu/nsnam/ns/>. *The Network Simulator - ns-2*.
- [11] Y. Liu, W. Gong, and P. Shenoy. On the impact of concurrent downloads. In *Proceedings of the Winter Simulation Conference*, pages 1300–1305, Arlington, VA, Dec. 2001.
- [12] C. Qiao and M. Yoo. Optical burst switching (obs) – a new paradigm for an optical internet. *Journal of High Speed Networks*, 8(1):69–84, Mar. 1999.
- [13] G. Raina, D. Towsley, and D. Wischik. Part II: control theory for buffer sizing. *SIGCOMM Computer Communication Review*, 35(3):79–82, July 2005.
- [14] J. D. Salehi, Z.-L. Zhang, J. F. Kurose, and D. F. Towsley. Supporting stored video: Reducing rate variability and end-to-end resource requirements through optimal smoothing. *IEEE/ACM Transactions on Networking*, 6(4):397–410, Aug. 1998.
- [15] S. Sen, J. Rexford, J. K. Dey, and J. F. Kurose. Online smoothing of

- variable-bit-rate streaming video. *IEEE Transactions on Multimedia*, 2(1):37–48, Mar. 2000.
- [16] V. Sivaraman, H. Elgindy, D. Moreland, and D. Ostry. Packet pacing in short buffer optical packet switched networks. In *Proceedings of the Twentyfifth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2006)*, Barcelona, Spain, Apr. 2006.
- [17] V. Sivaraman, D. Moreland, and D. Ostry. A novel delay-bounded traffic conditioner for optical edge switches. In *Proceedings of the IEEE Workshop on High-Performance Switching and Routing (HPSR 2005)*, Hong Kong, May 2005.
- [18] D. Wischik and G. Raina. Buffer sizes for large multiplexers: TCP queueing theory and instability analysis. In *Proceedings of 1st Conference on Next Generation Internet Networks (EuroNGI 2005)*, Rome, Italy, Apr. 2005.