Stochastic Traffic Regulator for End-to-End Network Delay Guarantees

Massieh Kordi Boroujeny, Brian L. Mark, Yariv Ephraim
Dept. of Electrical and Computer Engineering
George Mason University, Fairfax, VA

***Abstract*—Providing end-to-end network delay guarantees in packet-switched networks such as the Internet is highly desirable for mission-critical and delay-sensitive data transmission, yet it remains a challenging open problem. Due to the looseness of the deterministic bounds, various frameworks for stochastic network calculus have been proposed to provide tighter, probabilistic bounds on network delay, at least in theory. However, little attention has been devoted to the problem of regulating traffic according to stochastic burstiness bounds, which is necessary in order to guarantee the delay bounds in practice. We propose and analyze a stochastic traffic regulator that can be used in conjunction with results from stochastic network calculus to provide probabilistic guarantees on end-to-end network delay. Numerical results are provided to demonstrate the performance of the proposed traffic regulator.**[1]

***Index Terms*—Stochastic network calculus, traffic shaper, end-to-end delay, traffic burstiness bounds.**

I. Introduction

Currently, the Internet does not provide end-to-end delay guarantees for traffic flows. Even if the path taken by a given traffic flow is fixed, e.g., via mechanisms such as software-defined networking (SDN) or multi-protocol label switching (MPLS), network congestion arising from other flows can result in highly variable delays. The variability and random nature of traffic flows in a packet-switched network make it very challenging to provide any type of performance guarantees. The standard approach to providing network performance guarantees consists of two basic elements:

1) *Admission control:* A new flow should only be admitted to the network if sufficient resources are available for the new flow, as well as existing flows, to maintain their performance guarantees.
2) *Traffic regulation:* The traffic flow must be regulated to ensure that it does not use more resource than what was negotiated by the admission control scheme.

Admission control relies on a means of characterizing the traffic. On the other hand, the random and bursty nature of traffic flows in packet-switched networks make them difficult to characterize. Even if the flows can be modeled as random arrival processes, the problem of developing a resource allocation scheme to guarantee end-to-end performance in a network is generally intractable.

In his seminal work, Cruz [1], [2] proposed the so-called (σ, ρ) characterization of traffic, which imposes a deterministic bound on the burstiness of a traffic flow. By bounding traffic flows according to (σ, ρ) parameters, Cruz developed a network calculus which determined how these parameters propagate through network elements and from which end-to-end delay bounds could be derived. An important feature of the (σ, ρ) characterization is that it could be enforced by a traffic regulator. The (σ, ρ) framework of Cruz provides a machinery for end-to-end delay guarantees. In practice, however, the deterministic (σ, ρ) characterization leads to delay bounds that are very loose, which would lead to low utilization of the network resources. Nevertheless, the (σ, ρ) characterization was the basis for further research into stochastic bounds on traffic burstiness and stochastic network calculus to provide tighter, probabilistic end-to-end delay guarantees.

The development of stochastic network calculus and associated performance bounds remains an active topic of research. However, little attention has been devoted to the problem of traffic regulation to ensure that the input traffic of a network conforms to a stochastic traffic bound. In the deterministic network calculus of Cruz, the (σ, ρ) traffic regulator is tightly coupled to the (σ, ρ) traffic characterization. In effect, the (σ, ρ) traffic characterization is defined operationally in terms of a (σ, ρ) traffic regulator. To our knowledge, a traffic regulator to enforce a stochastic traffic bound has not been addressed previously in the literature.

In this paper, we develop a traffic regulator to enforce the so-called *generalized Stochastically Bounded Burstiness* (gSBB) traffic bound in [3], [4]. We refer to our proposed regulator as a stochastic (σ^*, ρ) regulator, since the burst size parameter can take on one of finite set of values. We describe the design and basic properties of the stochastic (σ^*, ρ) regulator and describe a practical implementation. Our analytical results establish that it enforces the gSBB bound.[2] We demonstrate the operation of the (σ^*, ρ) regulator via a numerical example involving a particular gSBB bound.

The remainder of the paper is organized as follows. In Section II, we review the deterministic (σ, ρ) regulator of Cruz. In Section III, we develop our proposed stochastic (σ^*, ρ) regulator. The numerical example is presented in Section IV. Concluding remarks are given in Section V.

[1]This work was supported in part by the U.S. National Science Foundation under Grant No. 1717033.

[2]Due to space constraints proofs are omitted.

II. Deterministic (σ, ρ) Regulator

The (σ, ρ) traffic regulator was introduced by Cruz in [1], [2] as a key element in a network calculus to bound the end-to-end delay of a traffic flow in a network. Let $R = \{R(t) : t \geq 0\}$ denote a traffic rate process such that the amount of traffic arriving in an interval $[s, t]$ is given by

$$A(s,t;R) := \int_s^t R(\tau)\, \mathrm{d}\tau, \tag{1}$$

in units of bits. In this paper, we assume here that traffic processes are in continuous-time, although our development could also be framed in discrete-time. A traffic stream R is said to be (σ, ρ)-bounded, denoted as $R \sim (\sigma, \rho)$, if

$$A(s,t;R) \leq \rho(t-s) + \sigma, \quad \forall s \in [0,t]. \tag{2}$$

For an idealized fluid model of input traffic, a (σ, ρ) traffic regulator ($\sigma, \rho \geq 0$) ensures that the *output* traffic stream $R_{\mathrm{o}} \sim (\sigma, \rho)$ and traffic departs the regulator in the same order as it arrives to the regulator, i.e., the service discipline is first-come first-served (FCFS). In the practical case in which the traffic consists of discrete packets of maximum length L and the input/output links to the regulator have finite capacity C, the output traffic stream satisfies $R_{\mathrm{o}} \sim (\sigma + \delta, \rho)$, where $\delta := (1 - \rho/C)L$ (see Fig. 1). Traffic regulation can be accomplished by dropping, low-priority marking, or delaying packets. In the first two cases, the traffic regulator is sometimes referred to as a *traffic policer* whereas in the third case it is referred to as a *traffic shaper*. The traffic regulators discussed in this paper will be of the traffic shaper variety.

Suppose a traffic stream R is offered to an infinite-buffer FCFS system with constant service rate ρ. The (virtual) workload of the system at time t is defined by

$$W_\rho(t;R) := \max_{0 \leq s \leq t} [A(s,t;R) - \rho(t-s)]. \tag{3}$$

Clearly, $W_\rho(t;R)$ is a decreasing function of ρ. It can easily be shown that $R \sim (\sigma, \rho)$ if and only if

$$W_\rho(t;R) \leq \sigma, \quad \forall t \geq 0. \tag{4}$$

Equation (4) provides a useful alternative characterization of a (σ, ρ)-bounded traffic stream.

Now suppose that the input and output traffic links to and from a (σ, ρ) regulator have a finite capacity $C > \rho$. Consider an input traffic stream R_{i} to the regulator. Let s_j denote the arrival time of the jth packet, t_j its departure time, and L_j its length in bits. The jth packet begins arriving at time s_j and is received completely at the regulator at time $a_j := s_j + L_j/C$. We assume that a packet does not arrive when the previous one is being received. i.e., $a_j < s_{j+1}$.

The operation of the regulator can be described in terms of the workload $W_\rho(s_j; R_{\mathrm{i}})$. At time s_j, if $W_\rho(s_j; R_{\mathrm{i}}) > \sigma$, the regulator delays the packet such that at its departure time t_j, the condition $W_\rho(t_j; R_{\mathrm{o}}) \leq \sigma$ holds. Hence, the delay incurred on the jth packet in the regulator is [1]

$$d_j = t_j - s_j = [W_\rho(s_j; R_{\mathrm{i}}) - \sigma]^+/\rho, \tag{5}$$

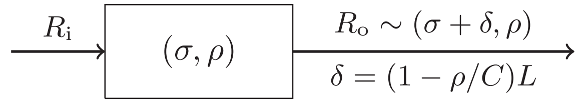


Fig. 1. Deterministic (σ, ρ) regulator with input/output links of capacity C.

where $[x]^+ := \max\{x, 0\}$. The jth packet completely departs the regulator at time

$$b_j := t_j + L_j/C. \tag{6}$$

At times other than the departure time, the workload may not necessarily be bounded by σ, but always satisfies [1]

$$W_\rho(t; R_{\mathrm{o}}) \leq \sigma + (1 - \rho/C)L, \quad \forall t \geq 0, \tag{7}$$

Thus, $R_{\mathrm{o}} \sim (\sigma + \delta, \rho)$, where δ can be viewed as the maximum error margin in regulating packetized traffic when the input/output links have capacity C (see Fig. 1).

As shown Fig. 2, when a packet is being received by the regulator, e.g., during $[s_j, a_j]$, the workload $W_\rho(t; R_{\mathrm{i}})$ increases linearly with slope $C - \rho$. Conversely, during the time between the complete arrival of a packet and the initial arrival of the next packet to the system, e.g., during $[a_j, s_{j+1}]$, the workload $W_\rho(t; R_{\mathrm{i}})$ decreases linearly with slope $-\rho$. Similarly, when a packet departs the regulator, e.g., during $[t_j, b_j]$, the workload $W_\rho(t; R_{\mathrm{o}})$ increases linearly with slope $C - \rho$. When packets are not departing the system, e.g., during $[b_j, t_{j+1}]$, $W_\rho(t; R_{\mathrm{o}})$ decreases linearly with slope $-\rho$. Assume that the buffer of the regulator is empty at $t = s_1$. Let $\delta_j := (1 - \rho/C)L_j$ denote the error margin due to regulating the jth packet. The governing equations for a (σ, ρ) regulator in terms of the workloads $W_\rho(t; R_{\mathrm{i}})$ and $W_\rho(t; R_{\mathrm{o}})$ are as follows [1]:

$$W_\rho(s_j; R_{\mathrm{i}}) = [W_\rho(a_{j-1}; R_{\mathrm{i}}) - \rho(s_j - a_{j-1})]^+, \tag{8}$$

$$W_\rho(a_j; R_{\mathrm{i}}) = W_\rho(s_j; R_{\mathrm{i}}) + \delta_j, \tag{9}$$

$$W_\rho(t_j; R_{\mathrm{o}}) = \begin{cases} \sigma, & \text{if } W_\rho(s_j; R_{\mathrm{i}}) > \sigma, \\ W_\rho(s_j; R_{\mathrm{i}}), & \text{if } W_\rho(s_j; R_{\mathrm{i}}) \leq \sigma, \end{cases} \tag{10}$$

$$W_\rho(b_j; R_{\mathrm{o}}) = W_\rho(t_j; R_{\mathrm{o}}) + \delta_j, \tag{11}$$

for $j = 1, 2, \ldots$.

III. Stochastic (σ^*, ρ) Regulator

The (σ, ρ) regulator may enforce a loose deterministic bound on the traffic or incur unnecessarily large delays on the traffic. To address these issues, we propose a *stochastic* traffic regulator, which is designed to enforce a probabilistic bound on the traffic:

$$\mathsf{P}\{W_\rho(t;R) \geq \gamma\} \leq f(\gamma), \quad \forall \gamma \in [0, T], \tag{12}$$

where $f(\gamma)$ is a non-increasing positive bounding function and T is a limit on the tail distribution of the workload (see [5]). We refer to a regulator that enforces (12) as a stochastic (σ^*, ρ) regulator, where the burstiness bound σ^* is variable. When $T = \infty$, (12) is equivalent to the definition of *generalized* Stochastically Bounded Burstiness (gSBB) proposed in [4],

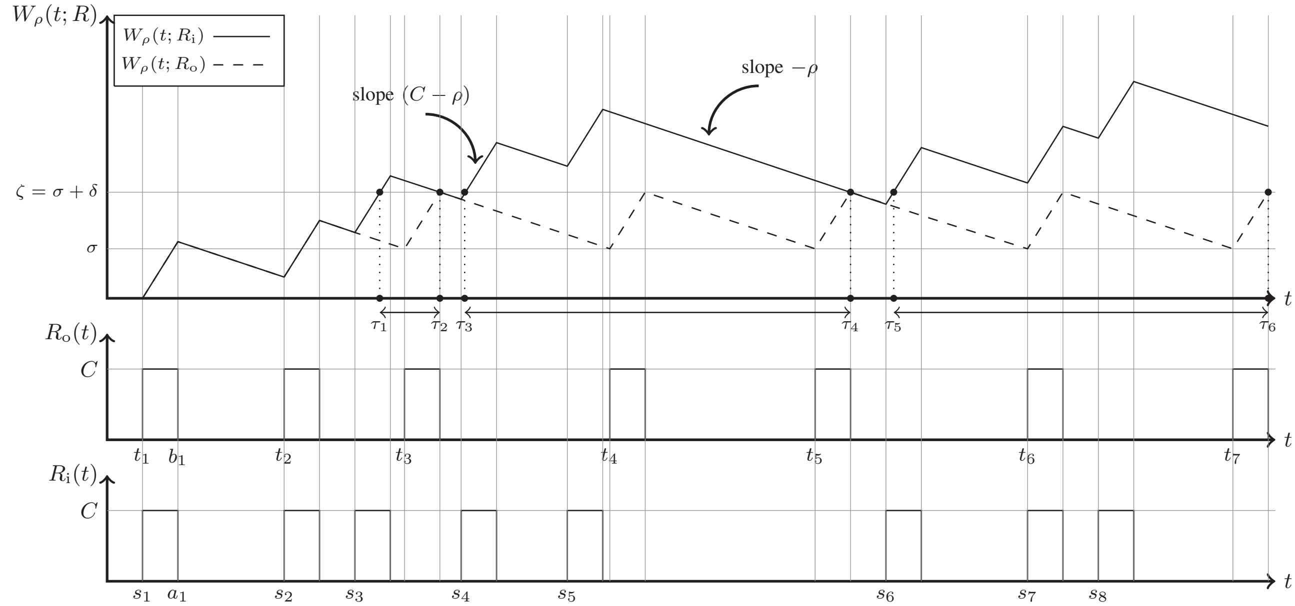


Fig. 2. Example of the operation of a (σ, ρ) traffic regulator.

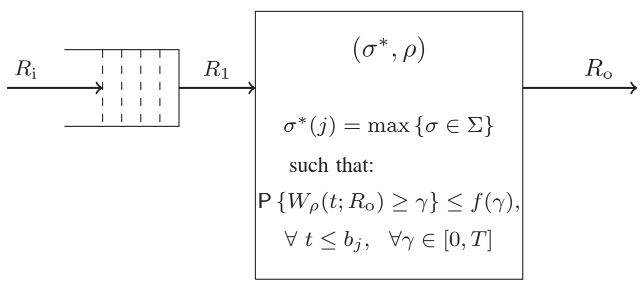


Fig. 3. Stochastic (σ^*, ρ) traffic regulator.

which is an extension of Stochastically Bounded Burstiness (SBB) introduced by [6]. The SBB concept is a generalization of the Exponentially Bounded Burstiness (EBB) developed earlier in [7], [8]. The EBB concept is closely related to the exponential bounds derived in [9] (cf. [10]).

Users specify their traffic flows with a parameter $(\rho, f(\gamma), T)$ in terms of a bound of the form (12). By applying results from the gSBB network calculus, the admissibility of a given set of traffic flows with respect to a certain probabilistic end-to-end delay constraint can be determined. However, such an end-to-end delay guarantee can only be provided if the user traffic streams conform to their stated traffic parameters. The stochastic traffic regulator developed in this paper can be applied at the network edge to ensure that a user's traffic stream does not violate the traffic parameter provided to the admission control unit.

A. Basic Design and Properties

A schematic of the stochastic (σ^*, ρ) regulator is shown in Fig. 3. The input and output links of the regulator are assumed to have capacity C. A buffer at the front-end of the regulator delays incoming packets until all previous packets have departed, thus ensuring a FCFS service discipline. Let R_i and R_o denote, respectively, the input traffic to and output traffic from the regulator. We denote internal traffic stream departing from the front-end buffer as R_1. Let s_j and $\tilde{s}_j$ denote, respectively, the arrival and departure time of the jth packet at the buffer. The regulator incurs a delay on the jth packet such that it begins departing the buffer at time t_j and completely leaves the regulator at time b_j. The departure times t_j are chosen with the objective of incurring minimal delay while satisfying

$$\mathsf{P}\{W_\rho(t;R_\mathrm{o}) \geq \gamma\} \leq f(\gamma), \quad \forall\, t \leq b_j, \ \forall \gamma \in [0,T]. \tag{13}$$

Since the front-end buffer delays each packet until the complete departure time of the previous packet from the regulator, we have

$$\tilde{s}_j = \max\{s_j, b_{j-1}\}. \tag{14}$$

We now discuss how the departure time t_j is determined. As in a deterministic (σ, ρ) traffic regulator, the rate parameter ρ must be greater than or equal to the long-term average input traffic rate, i.e.,

$$\rho \geq \lim_{t\to\infty} \frac{1}{t-s} \int_s^t R_\mathrm{i}(\tau)\, \mathrm{d}\tau, \quad \forall s \geq 0, \tag{15}$$

to avoid incurring an unbounded packet delay in the long-term. Unlike the burstiness parameter σ in the deterministic regulator, the proposed stochastic regulator involves a *variable* parameter chosen for the jth packet as $\sigma^*(j) \in \Sigma$, where $\Sigma := \{\sigma_1, \sigma_2, \ldots, \sigma_M\}$ is a finite set of values ordered such that $\sigma_1 > \sigma_2 > \ldots > \sigma_M$. The values in Σ arise from approximating the bounding function $f(\gamma)$ in the interval

$[0, T]$ by a stepwise function $\tilde{f}(\gamma)$, consisting of M steps (see Fig. 4) such that

$$\tilde{f}(\gamma) = \begin{cases} 1, & 0 \leq \gamma \leq T_M, \\ f(T_{M-1}), & T_M \leq \gamma \leq T_{M-1}, \\ \vdots & \vdots \\ f(T_1), & T_2 \leq \gamma \leq T_1, \\ 0, & T_1 \leq \gamma, \end{cases} \tag{16}$$

where T_i σ_i, and M, are defined as follows:

$$T_1 = T; \; T_i - T_{i+1} \geq \delta; \; \sigma_i = T_i - \delta \geq 0, \tag{17}$$

for $i = 1, 2, \ldots, M$. If T and δ are fixed, the maximum possible value of M is $\lfloor T/\delta \rfloor - 1$. We set $\tilde{f}(\gamma) = 1$ for $\gamma \in [0, T_M]$ because $\mathsf{P}\{W_\rho(t; R_\mathrm{o}) \geq 0\} = 1$. Since $\tilde{f}(\gamma) \geq f(\gamma)$ in this range, the bound may not be enforced for these values of γ. However, the violation probability is bounded by T_M/T, which can be made as small as $\delta/T \ll 1$. We also remark that setting $\tilde{f}(\gamma) = 0$ for $\gamma \geq T_1$ limits the burst size of the output traffic to T. Strictly speaking, this is not necessary for enforcing (12) and can be avoided by setting $\tilde{f}(\gamma)$ to a positive value in this range.

Similar to the deterministic regulator (cf. (5)), the delay incurred on the jth packet by the (σ^*, ρ) regulator can be computed as follows:

$$d_j = t_j - s_j = [W_\rho(s_j; R_\mathrm{i}) - \sigma^*(j)]^+/\rho, \tag{18}$$

where the burstiness parameter for the jth packet is determined according to

$$\sigma^*(j) = \max\Big\{\sigma \in \Sigma : \mathsf{P}\{W_\rho(t; R_\mathrm{o}) \geq \gamma\} \leq \tilde{f}(\gamma), \\ \forall\; t \in [b_{j-1}, b_j], \forall \gamma \in [0, T]\Big\}. \tag{19}$$

The set enclosed in braces in (19) is non-empty since σ_M satisfies the specified conditions. Thus, $\sigma^*(j)$ is well-defined. The packet j begins departing the regulator at time t_j, given by (18), and completely leaves at time b_j, given by (6). Essentially, the stochastic regulator determines the minimum delay that can be incurred on an incoming packet such that the upper bound in (19) is met during the time between the complete departures of packets $j-1$ and j. To achieve this, $\sigma^*(j)$ is chosen as the *maximum* value of $\sigma \in \Sigma$ such that the bound is satisfied.

Upon arrival of the jth packet at time s_j, the value of $\sigma^*(j)$ determined using according to (19) as follows. First, initialize $\sigma^*(j) = \sigma_1$. Then t_j is computed using (18) and the departure time b_j is computed from (6). Then based on the departure time, t_j, $W_\rho(t; R_o)$ is calculated and if the condition in (19) is met the procedure stops. Otherwise, the procedure is repeated with the next value in the ordered set Σ, i.e., $\sigma^*(j) = \sigma_2$. The procedure stops when $\sigma^*(j)$ is set equal to the largest value in Σ such that the condition in (19) is satisfied. Two difficulties with this procedure are 1) evaluation of the probability in (19) and 2) testing that the condition holds for all $\gamma \in [0, T]$ and for all $t \in [b_{j-1}, b_j]$. In Section III-B, we develop a practical implementation of the stochastic (σ^*, ρ) regulator, which addresses these issues.

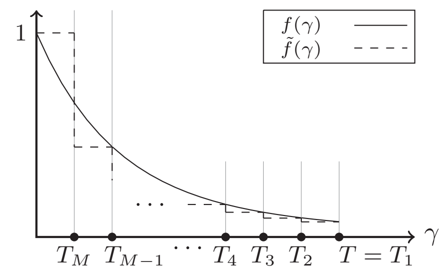


Fig. 4. Stepwise approximating function for $f(\gamma)$.

B. Practical Implementation

To analyze the stochastic (σ^*, ρ) regulator, it is convenient to introduce the internal traffic stream R_1 (see Fig. 7), Analogous to equations (8)–(11) for the deterministic (σ, ρ) regulator, the following equations involving R_1, with $\sigma^*(j) = \sigma$, can be derived:

$$W_\rho(\tilde{s}_j; R_1) = [W_\rho(b_{j-1}; R_\mathrm{o}) - \rho(\tilde{s}_j - b_{j-1})]^+, \tag{20}$$

$$W_\rho(t_j; R_\mathrm{o}) = \begin{cases} \sigma, & \text{if } W_\rho(\tilde{s}_j; R_1) > \sigma, \\ W_\rho(\tilde{s}_j; R_1), & \text{if } W_\rho(\tilde{s}_j; R_1) \leq \sigma, \end{cases} \tag{21}$$

$$W_\rho(b_j; R_\mathrm{o}) = W_\rho(t_j; R_\mathrm{o}) + \delta_j, \tag{22}$$

$$t_j = [W(\tilde{s}_j; R_1) - \sigma]^+/\rho + \tilde{s}_j. \tag{23}$$

Equation (20) follows from the equality $W_\rho(b_{j-1}; R_o) = W_\rho(b_{j-1}; R_1)$, since at most one packet is in the server part of the regulator at any given time.

The computations involved in the stochastic (σ^*, ρ) regulator can be simplified by introducing the concept of overshoot duration and by assuming stationarity and ergodicity of the input traffic stream R_i.

Definition 1. Given a threshold value $\zeta > 0$ and a traffic stream R, an *overshoot interval* with respect to R and ζ is a maximal interval of time η such that $W_\rho(\tau; R) \geq \zeta$ for all $\tau \in \eta$. Let $|\eta|$ denote the length of interval η. Let $\mathcal{O}(t)$ denote the set of overshoot intervals contained in $[0, t]$. Then the *overshoot duration* up to time t is defined as

$$\tilde{T}_\zeta(t; R) = \sum_{\eta \in \mathcal{O}(t)} |\eta|. \tag{24}$$

In Fig. 2, the overshoot set with respect to threshold value ζ until the end of time domain depicted in the figure consists of three intervals $[\tau_1, \tau_2]$, $[\tau_3, \tau_4]$ and $[\tau_5, \tau_6]$. Note that, as the maximum value for σ^* is σ_1, it can be easily verified using (17) and (7) that

$$\tilde{T}_{T_1}(t; R_\mathrm{o}) = 0, \quad \forall t > 0. \tag{25}$$

Given a time interval $[a, b]$, define $W_1 := W_\rho(a; R_\mathrm{o})$ and $W_2 := W_\rho(b; R_\mathrm{o})$. We define the increment in overshoot duration when the workload of the output process is *increasing* due to a packet departure from the regulator as follows:

$$\alpha(a,b,\zeta)=\begin{cases} b-a, & \zeta\le W_1,\\ \frac{W_2-\zeta}{C-\rho}, & W_1\le\zeta\le W_2\\ 0, & W_2<\zeta.\end{cases} \tag{26}$$

We define the increment in overshoot duration when the workload is *decreasing* due to the packet inter-departure time as follows:

$$\beta(a,b,\zeta)=\begin{cases} b-a, & \zeta\le W_2,\\ \frac{W_1-\zeta}{\rho}, & W_2\le\zeta\le W_1,\\ 0, & W_1<\zeta.\end{cases} \tag{27}$$

The following proposition shows how to compute $\tilde{T}_\zeta(t;R_o)$ at times $t=t_j$ and $t=b_j$ of packet j. See Figs. 7 and 5.

Proposition 1.

$$\tilde{T}_\zeta(t_j;R_\mathrm{o})=\tilde{T}_\zeta(b_{j-1};R_\mathrm{o})+\beta(b_{j-1},t_j,\zeta), \tag{28}$$

$$\tilde{T}_\zeta(b_j;R_\mathrm{o})=\tilde{T}_\zeta(t_j;R_\mathrm{o})+\alpha(t_j,b_j,\zeta), \tag{29}$$

If the input traffic R_i is stationary and ergodic, the output traffic R_o satisfies

$$\frac{\tilde{T}_\zeta(t;R_\mathrm{o})}{t}\sim \mathsf{P}\left\{W_\rho(t;R_\mathrm{o})\ge\zeta\right\}. \tag{30}$$

Therefore, for sufficiently large t, $\mathsf{P}\left\{W_\rho(t;R_\mathrm{o})\ge\gamma\right\}$ in (19) can be approximated by the left-hand side of (30). The following theorem simplifies the procedure for finding $\sigma^*(j)$ to satisfy the constraint of (19) and provides the basis for a practical implementation of the stochastic (σ^*,ρ) regulator.

Theorem 1. If the input traffic, R_i, to a stochastic (σ^*,ρ) regulator is stationary and ergodic, $\sigma^*(j)$ can be obtained as follows. Let $\mathcal{B}_j=\{\sigma\in\Sigma:\sigma>W_\rho(\tilde{s}_j;R_\mathrm{o})\}$ and

$$\Sigma_j=\begin{cases}\Sigma, & \text{if } \mathcal{B}_j=\emptyset,\\ \Sigma\backslash\mathcal{B}_j\cup\{\min\mathcal{B}_j\}, & \text{otherwise.}\end{cases} \tag{31}$$

Next, let

$$\mathcal{A}_j:=\left\{\sigma\in\Sigma_j\backslash\{\sigma_M\}:\frac{\tilde{T}_\zeta(b_j;R_\mathrm{o})}{b_j}\le\tilde{f}(\sigma+\delta)\right\} \tag{32}$$

where b_j and ζ are functions of σ. Here, $b_j(\sigma)$ is given by (6) and (23), and $\zeta(\sigma)=\max\{\gamma\in\Sigma:\gamma<\sigma\}+\delta$. Then we set

$$\sigma^*(j)=\begin{cases}\max\mathcal{A}_j, & \mathcal{A}_j\ne\emptyset,\\ \sigma_M, & \text{otherwise.}\end{cases} \tag{33}$$

According to Theorem 1, instead of verifying the upper bound in (19) for all $\gamma\in[0,T]$ and all $t\in[b_{j-1},b_j]$, we need only compute the set $\mathcal{A}_j$ in (32). For each $\sigma\in\Sigma_j\backslash\{\sigma_M\}$, we compute a temporary value for t_j from (23) and the corresponding values of b_j and ζ. We then update the values of $\tilde{T}_\zeta(t_j;R_\mathrm{o})$ and $\tilde{T}_\zeta(b_j;R_\mathrm{o})$ using Proposition 1 and test the condition in (32). After $\mathcal{A}_j$ has been determined, $\sigma^*(j)$ is computed using (33). The value of t_j used to delay packet j in the regulator is then computed using (23) with $\sigma=\sigma^*(j)$.

For the stochastic (σ^*,ρ) regulator, computation of the departure time, t_j, of the jth packet requires updates to at most $2(M-1)$ values of $\tilde{T}_{T_i}(t_j;R_\mathrm{o})$ and $\tilde{T}_{T_i}(b_j;R_\mathrm{o})$ for threshold values $i=2,\dots,M$. Once the departure time t_j is determined, $2(M-1)$ values of $\tilde{T}_{T_i}(t_j;R_\mathrm{o})$ and $\tilde{T}_{T_i}(b_j;R_\mathrm{o})$, for $i=2,\dots,M$, need to be updated. Thus, the overall computational complexity of the procedure outlined in Theorem 1 is $O(M)$ per packet.

IV. Numerical Results

We consider a system in which the packets sizes L_j are drawn randomly according to

$$L_j\sim \text{UniformI}[L_{\min},L_{\max}], \tag{34}$$

where UniformI[a,b] denotes the uniform distribution over the set of integers contained in the interval $[a,b]$. The inter-arrival times of the packets, $s_{j+1}-s_j$, are determined as follows:

$$s_{j+1}-s_j\sim U_j+\frac{L_j}{C}, \tag{35}$$

where $U_j\sim\text{Exp}(\lambda)$, i.e., $\{U_j\}$ is an i.i.d. sequence of exponentially distributed random variables with parameter λ. By adopting (35) to model the inter-arrival times, we ensure that packets are received after the previous ones have been fully received, i.e., the packets will not overlap with each other. In a system described by (34)–(35), $\rho^{-1}W_\rho(s_j)$ is equal to the waiting time experienced by the jth customer in a $G/G/1$ system in which the service time of the jth customer is given by $S_j=(\rho^{-1}-C^{-1})L_j$ and the inter-arrival time between the jth and $(j+1)$th customer is U_j [1], [11].

In this example, we set $L_{\min}=5$, $L_{\max}=10$, and $\lambda=0.25$, and $\rho=0.65$. We use the following bounding function:

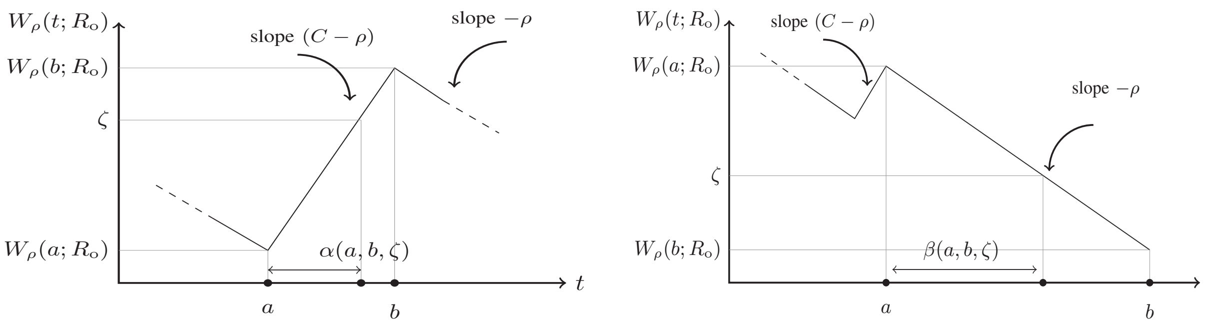
$$f(\sigma):=\begin{cases}-2.5\times10^{-3}\sigma+1, & 0\le\sigma\le40,\\ -5\times10^{-3}\sigma+1.1, & 40\le\sigma\le200.\end{cases} \tag{36}$$

In Fig. 6, $\tilde{f}(\sigma)$ is defined by approximating $f(\sigma)$ by a stepwise function with $M=10$ and $T_{i+1}-T_i=20$ for $i=1,2,\dots,M-1$. Observe that the output traffic is shaped to satisfy the desired bound.

Using the same model for inter-arrival and packet lengths, we have investigated the impact of the parameter M on traffic shaping of the input traffic. From Fig. 7, we see that as M is increased, a tighter fit of the output traffic to the desired bound can be achieved. In our example, the maximum possible value of M is $M=56$, for which a very tight fit to the bound is achieved. Table I presents the average delay and standard variation of the delay for the packets. Note that as M increases the average delay decreases and the standard deviation of the packet delay also decreases. These results are expected, since an increase in M implies that the delay incurred on a packet can increase in smaller increments, resulting smaller overall variance. In addition, a larger value of M results in a smaller average delay since there are more smaller choices of delay for a packet in order to maintain the burstiness bound.

V. Conclusion

The stochastic traffic regulator developed in this paper addresses an open problem in the application of stochastic network calculus to provide probabilistic end-to-end delay



(a) $\alpha(a, b, \zeta)$ (b) $\beta(a, b, \zeta)$

Fig. 5. Calculating the increment in the overshoot duration.

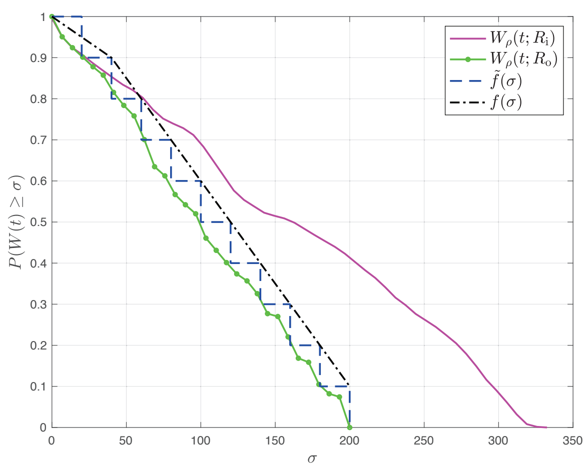


Fig. 6. Performance of the stochastic (σ^*, ρ) traffic regulator.

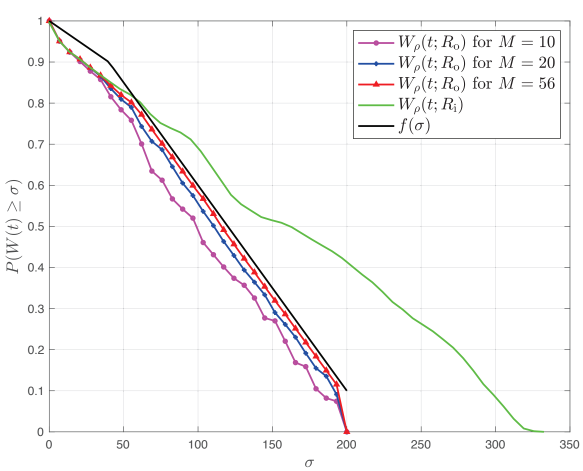


Fig. 7. Traffic regulator performance with different M values.

TABLE I
TRAFFIC SHAPING DELAY WITH DIFFERENT M VALUES.

M	Average Delay	Standard Deviation of Delay
10	93	113
20	81	107
56	75	99

guarantees. The key property of the proposed stochastic (σ^*, ρ) regulator is the enforcement of the gSBB traffic burstiness bound defined in [4] for an arbitrary bounding function. The performance of the stochastic regulator was demonstrated in a numerical example with a particular gSBB bounding function. The (σ^*, ρ) regulator could be applied in conjunction with the phase-type bounds proposed in [5].

REFERENCES

[1] R. L. Cruz, "A calculus for network delay. I. Network elements in isolation," *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 114–131, Jan. 1991.

[2] ——, "A calculus for network delay. II. Network analysis," *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 132–141, Jan. 1991.

[3] Q. Yin, Y. Jiang, S. Jiang, and P. Y. Kong, "Analysis on generalized stochastically bounded bursty traffic for communication networks," in *Proc. IEEE Local Comput. Netw. (LCN)*, Nov. 2002, pp. 141–149.

[4] Y. Jiang, Q. Yin, Y. Liu, and S. Jiang, "Fundamental calculus on generalized stochastically bounded bursty traffic for communication networks," *Comput. Netw.*, vol. 53, no. 12, pp. 2011 – 2021, Aug. 2009.

[5] M. Kordi Boroujeny, B. L. Mark, and Y. Ephraim, "Tail-limited phase-type burstiness bounds for network traffic," in *Proc. Inf. Sci. Sys. (CISS)*, Mar. 2019, pp. 1–6.

[6] D. Starobinski and M. Sidi, "Stochastically bounded burstiness for communication networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 1, pp. 206–212, Jan. 2000.

[7] O. Yaron and M. Sidi, "Performance and stability of communication networks via robust exponential bounds," *IEEE/ACM Trans. Netw.*, vol. 1, no. 3, pp. 372–385, Jun. 1993.

[8] ——, "Generalized processor sharing networks with exponentially bounded burstiness arrivals," in *IEEE INFOCOM Proc.*, vol. 2, Jun. 1994, pp. 628–634.

[9] C. S. Chang, "Stability, queue length, and delay of deterministic and stochastic queueing networks," *IEEE Trans. Autom. Control*, vol. 39, no. 5, pp. 913–931, May 1994.

[10] C.-S. Chang, *Performance Guarantees in Communication Networks*, ser. Telecommunication Networks and Computer Systems. London: Springer-Verlag, 2000.

[11] L. Kleinrock and H. H. Goldstine, *Queueing Systems. Volume II: Computer Applications*. New York: John Wiley, 1976.