A Deterministic Definition of Burstiness For Network Traffic Characterization*

Sami Ayyorgun   Wu-chun Feng

Research And Development in Advanced Network Technology (RADIANT)
Los Alamos National Laboratory
P.O. Box 1663, MS D451
Los Alamos, New Mexico 87545

{sami,feng}@lanl.gov

January 4, 2004

Abstract—The burstiness of network traffic has a profound impact on the performance of many network protocols. However, a widely accepted definition of burstiness, be it either deterministic or probabilistic, does not exist in the networking community. Deterministic definitions of burstiness provide an initial insight into both traffic characterization and performance analysis.

We propose a deterministic definition of burstiness for network traffic characterization, based on service curves. The proposed definition facilitates; (1) performance analyses for both average and scalar worst-case performance guarantees at the same time, (2) a simple systematic approach to performance guarantees, analytically (we show that the queue size, output traffic, virtual-delay, aggregate traffic, etc. at various points in a network can easily be characterized within the framework of the proposed definition), (3) a systematic analytical framework for measurement based analysis of probabilistic performance guarantees that would be inferred via sample-path analyses.

We also discuss the notion of burstiness. We point out that one might want to perceive the burstiness of a flow from the perspective of a network element; specifically, based on the queue size behavior that it induces on a network element of interest. We indicate that it is the decay rate of the tail of the queue size distribution that we care about in deciding the degree of burstiness of a flow with respect to another one, after some appropriate normalizations of the flows. The faster the decay rate is the less bursty the traffic is, and vice versa.

1 Introduction

With the rapid spread of data networks within the last decade, it has become apparent that network traffic exhibits bursty behavior. However, a widely accepted definition of burstiness, be it either deterministic or probabilistic, does not exist in the networking community. Yet, it is known that the burstiness of network traffic has a profound impact on the performance of many network protocols in areas such as congestion control (e.g. TCP), multiple-access (e.g. CSMA), routing (e.g. BGP), and switching and multiplexing in general.

Deterministic definitions of burstiness provide an initial insight into both traffic characterization and performance analysis, though they are not appealing for statistical gains. A handful of deterministic characterization of traffic with respect to burstiness is introduced in the early 90's [1, 2, 3]—these are almost all the deterministic characterizations which might be considered as currently relevant. One of these characterizations [1], and its companion service model (*service curve* model [4, 5]), has received a considerable attention from the networking community, recently resulting in two books on the subject [6, 7], primarily because they have facilitated a systematic analytical approach to performance guarantees in communication networks, much like what we have in linear system theory.

None of these characterizations mentioned above, however, considers average performance guarantees. Their main focus is on some scalar worst-case performance metrics; such as delay, backlog, or jitter being less than or equal to a certain scalar quantity, at every point in time at a network element. Average performance guarantees, on the other hand, are in demand by many applications. Motivated partly

*Los Alamos National Laboratory, Technical Report LA-UR-03-4477, June 2003.

by such demands, a few works on the implications of average case performance guarantees due to the characterization in [1] have appeared in the literature in recent years, [8, 9].

In this study, we propose a deterministic definition of burstiness which could address both average and scalar worst-case performance guarantees at the same time. Our motivation for this study is threefold: (1) We would like to try to clarify the notion of burstiness, and how we might want to perceive it. (2) We would like to come up with a deterministic definition of burstiness, and hence a traffic characterization, for both average and scalar worst-case performance guarantees, and still retain a systematic analytical approach as presented by the characterization in [1] and its companion service model that we have mentioned earlier. (3) We would like the deterministic characterization that we come up with to be directly applicable to measurement based analysis of probabilistic performance guarantees that would be inferred via sample-path analyses.

The notion of burstiness, and how we might want to perceive it, will be discussed in section 3. As for our second motivation, we would like a traffic characterization to have the following properties: Let $\mathcal{C}$ denote a traffic characterization, then

1. if two traffic flows are characterized according to $\mathcal{C}$, then the aggregate of the flows should also be easily characterized according to $\mathcal{C}$,

2. if a traffic flow characterized according to $\mathcal{C}$ is fed into a network element commonly used in practice, then

   (a) the output flow should also be easily characterized according to $\mathcal{C}$,

   (b) and, both the queue size and the virtual delay should also be easily characterized in the same framework as $\mathcal{C}$,

3. the characterization $\mathcal{C}$ should be "stationary" in the sense that the characterization of a time-shifted traffic does not change with respect to that of the unshifted traffic.

We show that the traffic characterization provided by the burstiness definition that we propose here satisfies all of the above properties. These properties facilitate a systematic analytical treatment of performance guarantees in communication networks.

For our third motivation, we would like the burstiness definition that we come up with be such that both the definition itself and its implications on performance guarantees could be viewed from the stand point of relative frequency interpretation of probability. The burstiness definition that we propose here has inherently this aspect.

The rest of the paper is organized as follows: Section 2 provides a background. Section 3 discusses the notion of burstiness, and how we might want to perceive it. Section 4 introduces the new burstiness definition. Sections 4.1 and 4.2 examine the implications of the new definition, by itself and for a single network element, respectively. Section 4.2.1 gives the average performance guarantees for a single network element. Section 4.3 examines the performance guarantees according to the new burstiness definition, over a tandem of network elements. Section 5 provides some discussions about the new characterization. Finally, section 6 gives conclusions.

2 Background and Convention

We adopt a discrete-time formulation for the simplicity of exposition. Time is slotted into fixed-length intervals, and marked by the integers. The unit of transmission for communication is referred to as a *packet*, in this study. A *flow* is a non-decreasing function defined from the integers to the non-negative integers. The value $R(n)$ of a flow R at time n denotes the total number of packets that has arrived by time n (inclusive) for a *connection*. The *rate-function* r of a flow R is defined as

$$r(n) = R(n) - R(n-1) \quad \text{for all } n,$$

which denotes the total number of packets that has arrived at time n. A *network element* is an input-output device that accepts packets at its input, processes them, and delivers them at its output. A network element is said to be *passive* if it does not generate any packet internally. Network elements are assumed to be passive in this study, for the simplicity of exposition. Packets are assumed to be able to instantaneously arrive and depart at a network element, i.e. a whole packet could arrive instantaneously at time k, and depart later at time n where $n \geqslant k$. Note that a packet could depart in the same interval in which it has arrived; this is sometimes referred to as *cut-through* operation. The capacity $c(n)$ of a network element at time n is the total number of packets that it could deliver (serve) at time n. The function c is called the *instantaneous capacity rate*, or just the *rate*, of the network element.

We denote the set of all the integers by $\mathbb{Z}$, and the set of all the positive integers by $\mathbb{Z}^+$. Given a statement A which could be true or false, the notation $[A]$ stands for 1 if A is true, and 0 otherwise.[1] The logical negation of a statement A is denoted by $\bar{A}$, i.e. if A is true, $\bar{A}$ is false, and if A is false, $\bar{A}$ is true. Finally, all functions are assumed to be defined from the integers to the integers, unless otherwise noted from here on.

For convenience, the end of proofs in the text are marked by '$\blacksquare$', and the end of examples are marked by '$\lrcorner$', where both marks are flushed to the right margin.

We utilize the following definitions in this study, which have been previously introduced in the literature.

Definition 1 *Let f and g be any two functions. The* min-+ convolution *of f and g, denoted*[2] *as $f \triangledown g$, is defined as*

$$(f \triangledown g)(n) = \min_{k \leqslant n}\{f(k) + g(n-k)\} \qquad \textit{for all } n.$$

The convolution $f \triangledown g$ is read as "f min-convolved with g", or as "the min-plus convolution of f with g".

Consider the problem of finding a function X whose min-+ convolution with g is f, i.e. find an X such that $X \triangledown g = f$. One can proceed to find X as follows;

$$\begin{aligned}
(X \triangledown g)(n) &= f(n) \\
\min_{k \leqslant n}\{X(k) + g(n-k)\} &= f(n) \\
X(k) + g(n-k) &\geqslant f(n) && \text{for all } k \leqslant n \\
X(k) &\geqslant f(n) - g(n-k) && \text{for all } k \leqslant n \\
X(k) &\geqslant \max_{n \geqslant k}\{f(n) - g(n-k)\} \\
&= \max_{n \geqslant 0}\{f(n+k) - g(n)\}\,.
\end{aligned}$$

Hence, it is convenient to define the min-+ *deconvolution* of two functions f and g as follows.

Definition 2 *Let f and g be two functions. The* min-+ deconvolution *of f and g, denoted as $f \,\underline{\triangledown}\, g$, is defined as*

$$(f \,\underline{\triangledown}\, g)(n) = \max_{k \geqslant 0}\{f(n+k) - g(k)\} \qquad \textit{for all } n.$$

The deconvolution $f \,\underline{\triangledown}\, g$ is read as "f min-deconvolved with g", or as "the min-plus deconvolution of f with g"

A companion operator to the min-+ convolution is called the max-+ *convolution*, and defined as follows.

[1] We have adopted this notation from [13].

[2] One reason to choose this notation over some others, for example '$*$', is that there are a companion and related other operators to this operator, which are employed later in the work. We believe that this choice of notation provides a better choice of notations for these other operators in a fitting manner.

Definition 3 *Let f and g be two functions. The* max-+ convolution *of f and g, denoted as $f \vartriangle g$, is defined as*

$$(f \vartriangle g)(n) = \max_{k \leqslant n}\{f(k) + g(n-k)\} \qquad \textit{for all } n.$$

The convolution $f \vartriangle g$ is read as "f max-convolved with g", or as "the max-plus convolution of f with g".

Again, consider the problem of finding a function X whose max-+ convolution with g is f; that is, find an X such that $X \vartriangle g = f$. One can proceed to find X as follows;

$$\begin{aligned}
(X \vartriangle g)(n) &= f(n) \\
\max_{k \leqslant n}\{X(k) + g(n-k)\} &= f(n) \\
X(k) + g(n-k) &\leqslant f(n) && \text{for all } k \leqslant n \\
X(k) &\leqslant f(n) - g(n-k) && \text{for all } k \leqslant n \\
X(k) &\leqslant \min_{n \geqslant k}\{f(n) - g(n-k)\} \\
&= \min_{n \geqslant 0}\{f(n+k) - g(n)\}\,.
\end{aligned}$$

Hence, it is also convenient to define the max-+ *deconvolution* of two functions f and g as follows.

Definition 4 *Let f and g be two functions. The* max-+ deconvolution *of f and g, denoted as $f \,\bar{\vartriangle}\, g$, is defined as*

$$(f \,\bar{\vartriangle}\, g)(n) = \min_{k \geqslant 0}\{f(n+k) - g(k)\} \qquad \textit{for all } n.$$

The deconvolution $f \,\bar{\vartriangle}\, g$ is read as "f max-deconvolved with g", or as "the max-plus deconvolution of f with g".

Definition 5 *An S-*server *with* service curve S *is a network element that when fed with an input flow R, the corresponding output flow G satisfies*

$$G(n) \geqslant (R \triangledown S)(n) \quad \textit{for all } n$$

for any R. A service curve S is a non-decreasing function defined from the integers to the non-negative integers, and takes on the value zero for non-positive values (i.e. $S(n) = 0$ for all $n \leqslant 0$).

An *S-server with equality* is an S-server such that the inequality in definition 5 becomes an equality, i.e. the output of an S-server with equality is given by $G(n) = (R \triangledown S)(n)$ for all n.

Note that a *work-conserving*[3] server with the capacity of serving packets with a constant integer rate ρ is an S-server with equality with service curve $S(n) = \max\{0, \rho \cdot n\}$, since the output G of a work-conserving server with an input flow R is given[4] by

$$G(n) = \min_{k \leqslant n}\{R(k) + \rho\,(n-k)\} \quad \text{for all } n$$

which could also be represented equivalently as

$$G(n) = (R \triangledown S)(n) \quad \text{for all } n.$$

[3] A network element is said to be *work-conserving* if it serves packets whenever it has packets to serve, unconditionally of any other criteria.

[4] This is often referred to as Reich's result [10]. See [11], for example, for a simple derivation.

3 The Notion of Burstiness

The burstiness of an arrival process, roughly, has to do with the "proximity" of arrival instances to each other, and also with the "variation" of arrival amounts from one arrival instance to another. A burstiness definition in coming up with a traffic characterization, tries to restrict these two aspects of variations in arrivals, in a combined fashion so that some provable bounds on a specified set of performance metrics of interest could be given with ease. This is in general the case for both deterministic and probabilistic definitions of burstiness, which could be followed for most of the burstiness definitions and traffic characterizations surveyed in [11].

A catch in the above paragraph is the phrase 'so that some provable bounds on a specified set of performance metrics of interest could be given with ease'. A key view in coming up with a burstiness definition with *utility* is to have a focus on some performance metrics of interest. This often implies that one would base his/her perception of burstiness on the behavior of an arrival process through a network element (which is typically a variant of a work-conserving server). In other words, one would tend to perceive the burstiness of a traffic from the perspective of a server (a network element).

A good example to this where this view is squarely placed at the heart of the definition is the (σ, ρ) model, which also leads to the concept of *arrival curves* in the general case, [1, 5]. A flow R is said to be (σ, ρ) constrained (or, (σ, ρ) smooth) if it satisfies the following condition

$$R(n+k) - R(k) \leqslant \sigma + \rho \cdot n \quad \text{for all } k \text{ and } n \geqslant 0.$$

Now, if a (σ, ρ) constrained flow R is fed into a work-conserving server with constant rate ρ, it is not difficult to show that the backlog $Q(n)$ at any time n is upper bounded by σ; let G be the corresponding output flow, then

$$\begin{aligned} Q(n) &= R(n) - G(n) \\ &= R(n) - \min_{k \leqslant n}\{R(k) + \rho \cdot (n-k)\} && \text{by the work-conserving server} \\ &= \max_{k \leqslant n}\{R(n) - R(k) - \rho \cdot (n-k)\} \\ &\leqslant \max_{k \leqslant n}\{\sigma + \rho \cdot (n-k) - \rho \cdot (n-k)\} && \text{by the characterization of } R \\ &= \sigma\,. \end{aligned}$$

A similar statement could also be given for a flow conforming to an arrival curve. A flow R is said to be conformal to an arrival curve A (or, A smooth) if it satisfies the following condition

$$R(n+k) - R(k) \leqslant A(n) \quad \text{for all } k \text{ and } n \geqslant 0,$$

where A is a non-decreasing function defined from the integers to the non-negative integers.

At first, this view of having some performance metrics of interest in focus, in coming up with a burstiness definition, might not become apparent. Rather, one might be inclined towards giving a definition by somehow comparing a traffic by itself. An example to this is the following traffic characterization that we might suggest; a flow R is said to be proportionally constrained by $\alpha(n)$ if it satisfies the following inequality

$$\frac{R(n+k) - R(k)}{\max\{R(n+l) - R(l), 1\}} \leqslant \alpha(n)\,,$$

for all k, l, and n, where $(k, n+k]$ and $(l, n+l]$ are non-overlapping intervals and their boundaries lie in between some start and end time of flow R that we might easily define.

Although such definitions may seem to be intuitive and natural at first sight, they do not necessarily lend themselves very well to get bounds on some simple performance metrics. In other words, the burstiness perceived by human eye, without taking into account any consideration regarding the behavior of

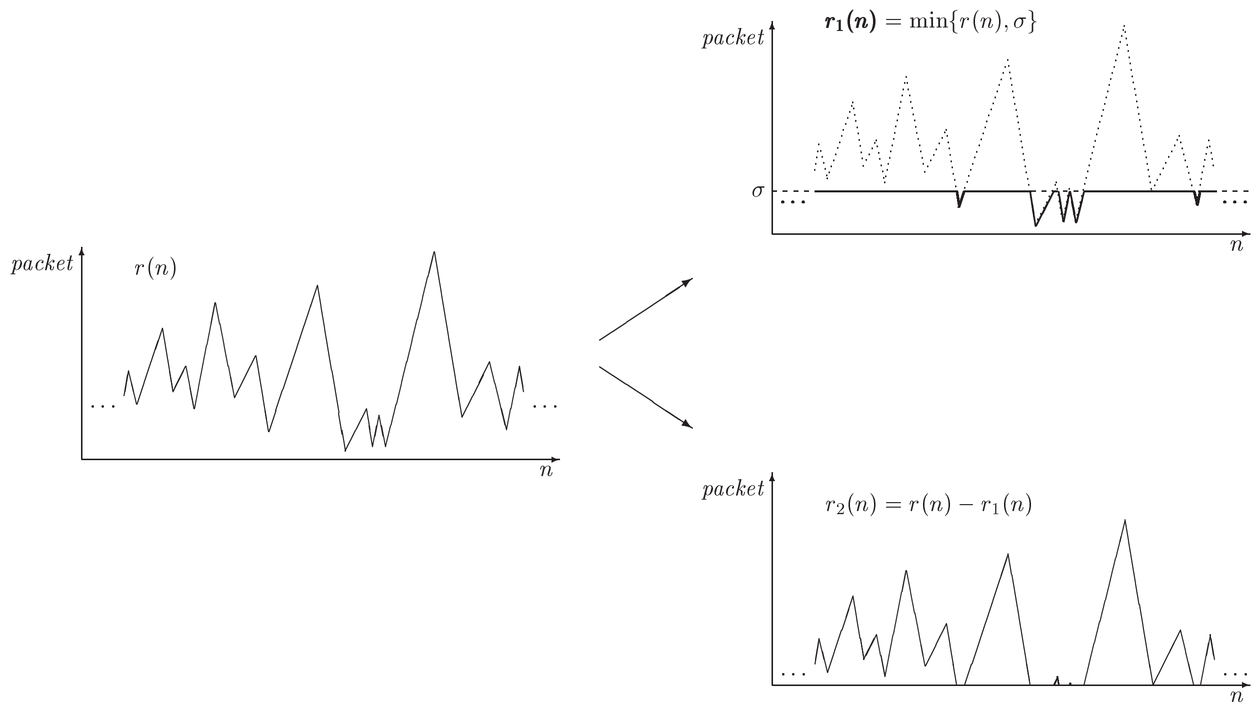

Figure 1: Given a rate-function $r(n)$ and a trim-off level σ, the trimmed rate-function $r_1(n)$ and the trimmed-off rate-function $r_2(n)$.

a traffic through a server (a network element in general), does not necessarily suit well for performance analysis, and is not necessarily the same as burstiness that one might perceive from the perspective of a server. Said differently, bare human perception of burstiness does not necessarily incorporate simple performance metrics (that we are interested in, in general, in performance analysis) in our cognition.

Thus, we choose to base our perception of burstiness of a flow on its behavior through a network element, specifically on the queue size behavior.

3.1 Burstiness via Level-crossing

Given a flow R, consider its rate-function r. Let us trim r at a certain level σ, as given below where the trimmed rate-function is denoted by r_1,

$$r_1(n) = \min\{r(n), \sigma\} \quad \text{for all } n.$$

Let us also consider the part of the rate-function which has been trimmed off, i.e.

$$r_2(n) = r(n) - r_1(n) \quad \text{for all } n.$$

Vaguely, we would tend to perceive that $r_2(n)$ is at least as bursty as $r(n)$, and $r_1(n)$ is at most as bursty as $r(n)$, for some values of σ. This could be viewed in Figure 1 where time-slots are drawn very close to each other for convenience.

To put it more precisely, if we were to multiply both r_1 and r_2 by some constants a_1 and a_2, respectively, such that the rate-functions $\lfloor a_1 \cdot r_1(n) \rfloor$ and $\lceil a_2 \cdot r_2(n) \rceil$ would have the same average rates as $r(n)$, and feed $r(n)$, $\lfloor a_1 \cdot r_1(n) \rfloor$, and $\lceil a_2 \cdot r_2(n) \rceil$ into identical network elements with labels 0, 1, and 2, respectively, where the average rate that each network element serves packets is grater than that of its input, we would observe for some values of σ that the queue size behavior in network element 1 is less "erratic" than that of 0, and the queue size behavior in network element 2 is more "erratic" than that of 0, considering the entire duration of the observations.

In other words, if we select r from a random sample space, and feed all three processes r, $\lfloor a_1 \cdot r_1(n) \rfloor$, and $\lceil a_2 \cdot r_2(n) \rceil$ for each selection of r into network elements 0, 1, and 2, respectively, and observe the queue size distributions; for some values of σ, the decay rate of the tail of the queue size distribution (i.e. the complementary cumulative distribution function of the queue size) in network element 1 is faster that that of network element 0 after a certain queue size value, and it is slower in network element 2 than that of network element 0 again after a certain queue size value.

So, it is the decay rate of the tail of the queue size distribution that we care about in deciding the degree of burstiness of a flow with respect to another one, after some appropriate normalizations of the flows as we have indicated above. The faster the decay rate is the less bursty the traffic is, and vice versa.

More precisely, we could decide the burstiness of a traffic source A with respect to another one B, from the perspective of a network element of interest, as follows: Normalize the sources such that the average rate of traffic coming out of source A is equal to that of source B. Feed the traffic coming out of each source A and B into identical network elements α and β, respectively, where the average rate that each network element serves packets is grater than that of its input. Observe the queue size distribution in each network element α and β. If there exists a queue size level σ_0 beyond which the decay rate of the tail of the queue size distribution in network element α is slower than that of network element β, then source A is more bursty than source B.

Returning to the trimming of rate-function r, and in light of the above discussions, we might want to lower bound the rate of the overshoot of r above each trim-off level σ, in an attempt to restrict its burstiness. We might also consider to upper bound it as well.

However, recalling our perception of burstiness of a flow via the queue size behavior that it induces on a network element, we might actually want to lower and/or upper bound the rate of the overshoot of not r, but instead for example, $\max\{0, r(n) - \rho\}$, if the network element at hand is a work-conserving server with constant rate ρ.

The lower-bound suggested above, in the case where the network element at hand is a work-conserving server with constant rate ρ, could be expressed as follows;

$$L(\sigma, n) \leqslant \frac{1}{n} \sum_{k < i \leqslant n+k} \Big[\max\{0, r(i) - \rho\} > \sigma \Big] \,.$$

Notice that above we would like to have the subscript '$k < i \leqslant n + k$' of the sum as such, since we would like to have a "stationary" characterization as we have indicated in the introduction.

It is not difficult to show that (which could be inferred from the proof of theorem 2 in getting the line tagged $(*)$ from the previous line there) the above lower-bound implies the following lower-bound on the tail of the queue size distribution;

$$L(\sigma, n) \leqslant \frac{1}{n} \sum_{k < i \leqslant n+k} [Q(i) > \sigma] \,.$$

which also implies a lower-bound on the decay rate of the tail of the queue size distribution, and hence on the burstiness of r.

The upper-bound that we have mentioned above could also be expressed as

$$\frac{1}{n} \sum_{k < i \leqslant n+k} \Big[\max\{0, r(i) - \rho\} > \sigma \Big] \leqslant U(\sigma, n) \,.$$

Notice that this bound would be of no use to get neither an upper-bound nor a lower-bound on the tail of the queue size distribution as we have indicated above, in the case where the network element to which the flow is fed is an S-server. Again, this could be inferred from the proof of theorem 2 in getting the line tagged $(*)$ from the previous line there.

However, again, recalling our perception of burstiness of a flow via the *queue size* behavior that it induces on a network element, we might actually want to lower and/or upper bound the rate of overshoot of the queue size *itself* directly, instead of a variant of the rate-function. With this view, we propose the following definition of burstiness, and hence a traffic characterization, in the following section where we utilize an upper-bound as we have indicated above. As for a study regarding a lower-bound, we leave it to a future work.

4 A Deterministic Definition of Burstiness

With the discussions and motivations given in section 3, we propose the following definition of burstiness, and hence a traffic characterization.

Definition 6 *A traffic flow R is said to be* bursty *with* service curve S *and level-crossing function* $U(\sigma, n)$*, and denoted as* $R \sim (S, U)$*, if*

$$\frac{1}{n} \sum_{k<i\leqslant n+k} \Big[R(i) - R(j) > S(i-j) + \sigma, \quad \textit{for some } j < i \Big] \leqslant U(\sigma, n) \tag{1}$$

for all k, $n > 0$, and σ, where $U(\sigma, n)$ is defined from $\mathbb{Z} \times \mathbb{Z}^+$ to the non-negative real numbers.

We assume that the following properties hold for any level-crossing function $U(\sigma, n)$, without loss of generality:

1. $U(\sigma, n)$ is non-increasing in σ, since the quantity corresponding to a σ on the left-hand-side of inequality (1) is non-increasing with σ.

2. Clearly, $U(\sigma, n) \leqslant 1$ for any σ and $n > 0$. Furthermore, we assume for mathematical convenience that $U(\sigma, n) = 1$ for all $\sigma < 0$ and for any $n > 0$, unless otherwise noted from here on.

3. $\lim_{\sigma \to \infty} U(\sigma, n) = 0$ for any $n > 0$ (this is certainly the case if flow R is bounded, which we could always assume without loss of generality for almost all practical purposes).

In the rest of this section, we examine some of the properties/implications of definition 6. Specifically, we will show that it satisfies all the properties of a traffic characterization that we have sought to have as stated earlier in the introduction.

4.1 Implications on Aggregate Flows and Average Rate

We first show that the traffic characterization provided by definition 6 satisfies the property 1 of a traffic characterization that we have sought to have as stated in the introduction. That is, we show that the aggregate of flows where each flow is characterized according to this characterization could also be easily characterized by the same characterization. This is stated more precisely in the following theorem, and proved thereafter.

Theorem 1 *Let R_1 and R_2 be two flows that $R_1 \sim (S_1, U_1)$ and $R_2 \sim (S_2, U_2)$. The aggregate flow $R_1 + R_2$ is bursty with service curve $S_1 + S_2$ and level-crossing function $U_1 \triangledown U_2$ where the convolution is carried out over the first arguments (i.e. σ) of U_i's. In other words, $R_1 + R_2 \sim (S_1 + S_2, U_1 \triangledown U_2)$.*

Let us adopt a convention from here on that whenever we refer to a min-+ convolution of any two bivariate functions, we mean the min-+ convolution carried out over their first arguments.

Proof: The proof follows by considering the following statements for any k, $n > 0$, i that $k < i \leqslant n + k$, σ, and $u \leqslant \sigma$;

$$\begin{aligned}
A &: (R_1 + R_2)(i) - (R_1 + R_2)(j) > (S_1 + S_2)(i - j) + \sigma, \quad \text{for some } j < i \\
A_1 &: R_1(i) - R_1(j) > S_1(i - j) + u, \quad \text{for some } j < i \\
A_2 &: R_2(i) - R_2(j) > S_2(i - j) + \sigma - u, \quad \text{for some } j < i.
\end{aligned}$$

Notice that $A \Rightarrow (A_1 \text{ OR } A_2)$, since clearly $(\bar{A_1} \text{ AND } \bar{A_2}) \Rightarrow \bar{A_1}$. Thus, we have

$$\begin{aligned}[A] &\leqslant [A_1 \text{ OR } A_2] \\ &\leqslant [A_1] + [A_2] .\end{aligned}$$

Since the above inequality holds for any k, $n > 0$, i that $k < i \leqslant n + k$; and since $R_1 \sim (S_1, U_1)$ and $R_2 \sim (S_2, U_2)$, we also have

$$\frac{1}{n} \sum_{k<i\leqslant n+k} [A] \leqslant U_1(u, n) + U_2(\sigma - u, n) .$$

Furthermore, since the last inequality holds for any $u \leqslant \sigma$, we get

$$\begin{aligned}\frac{1}{n} \sum_{k<i\leqslant n+k} [A] &\leqslant \min_{u\leqslant\sigma}\{U_1(u, n) + U_2(\sigma - u, n)\} \\ &= (U_1 \triangledown U_2)(\sigma, n) .\end{aligned}$$

One might think that we would have had to use infimum 'inf' above instead of minimum 'min', since U_i's are real-valued. However, it turns out that this is not the case since the above minimum is effectively taken over a set of finite number of elements due to the fact that $U_i(s, n) = 1$ for all $s < 0$ and for any $n > 0$, for i equals to both 1 and 2. Moreover, since $\lim_{\sigma\to\infty} U(\sigma, n) = 0$ for any U, note that we have $(U_1 \triangledown U_2)(\sigma, n) = 1$ for all $\sigma < 0$ and $n > 0$.[5] This completes the proof. ∎

The burstiness characterization provided by definition 6 has also an implication on the long-term average rate of a flow. This is stated in the following theorem, and proved thereafter. Before we give that, however, we would like to provide the following lemma.

Lemma 1 *There holds for any non-negative real number a that*

$$a = \int_0^\infty [a > x] \; dx .$$

Proof: The proof follows by the following simple manipulation;

$$\begin{aligned}\int_0^\infty [a > x] \; dx &= \int_0^{a^-} [a > x] \; dx + \int_a^\infty [a > x] \; dx \\ &= \int_0^{a^-} 1 \; dx + \int_a^\infty 0 \; dx \\ &= a .\end{aligned}$$

∎

The discrete-time version of the above lemma is given below

$$a = \sum_{n=0}^{\infty} [a > n]$$

where a is any non-negative integer, whose proof also follows similarly.[6]

[5]This is why we have assumed for mathematical convenience that a level-crossing function in burstiness definition 6 satisfies $U(s, n) = 1$ for all $s < 0$ and $n > 0$.

[6]In general, we have

$$a = \int_{-\infty}^0 [a < x] \, dx + \int_0^\infty [a > x] \, dx ,$$

where a is any real number.

Theorem 2 *Given a flow $R \sim (S, U)$, the long-term average rate μ of flow R satisfies*

$$\mu = \limsup_{(n-k)\to\infty} \frac{R(n) - R(k)}{n-k} \leqslant \limsup_{n\to\infty} \frac{S(n)}{n} + \limsup_{n\to\infty} \sum_{\sigma \geqslant 0} U(\sigma, n)\,.$$

Proof: Let $\rho \equiv \limsup_{n\to\infty} \frac{S(n)}{n}$.

The proof follows by considering the total number of packets $R(n) - R(k)$ that would arrive in any interval $(k, n]$, and upper bounding it as;

$$\begin{aligned}
R(n) - R(k) &= \sum_{k<i\leqslant n} r(i) \\
&= \sum_{k<i\leqslant n} \int_0^\infty [r(i) > \sigma]\; d\sigma \\
&= \int_0^\infty \sum_{k<i\leqslant n} [r(i) > \sigma]\; d\sigma \\
&= \int_0^\rho \sum_{k<i\leqslant n} [r(i) > \sigma]\; d\sigma + \int_\rho^\infty \sum_{k<i\leqslant n} [r(i) > \sigma]\; d\sigma \\
&\leqslant \int_0^\rho \sum_{k<i\leqslant n} 1\; d\sigma + \int_\rho^\infty \sum_{k<i\leqslant n} [r(i) > \sigma]\; d\sigma \\
&= \int_0^\rho (n-k)\; d\sigma + \int_0^\infty \sum_{k<i\leqslant n} [r(i) > \rho + \sigma]\; d\sigma \\
&= \rho\cdot(n-k) + \int_0^\infty \sum_{k<i\leqslant n} [r(i) > \rho + \sigma]\; d\sigma \\
&\leqslant \rho\cdot(n-k) + \int_0^\infty \sum_{k<i\leqslant n} [R(i) - R(j) > \rho\cdot(i-j) + \sigma, \quad \text{for some } j < i]\; d\sigma \qquad (*) \\
&\leqslant \rho\cdot(n-k) + \int_0^\infty \sum_{k<i\leqslant n} [R(i) - R(j) > S(i-j) + \sigma, \quad \text{for some } j < i]\; d\sigma \\
&\leqslant \rho\cdot(n-k) + \sum_{\sigma=0}^{\infty} \sum_{k<i\leqslant n} [R(i) - R(j) > S(i-j) + \sigma, \quad \text{for some } j < i] \\
&\leqslant \rho\cdot(n-k) + \sum_{\sigma=0}^{\infty} U(\sigma, n-k)\cdot(n-k)\,.
\end{aligned}$$

Hence, we get

$$\begin{aligned}
\mu &= \limsup_{(n-k)\to\infty} \frac{R(n)-R(k)}{n-k} \\
&\leqslant \limsup_{(n-k)\to\infty} \frac{\rho\cdot(n-k)+\sum_{\sigma=0}^{\infty}U(\sigma,n-k)\cdot(n-k)}{n-k} \\
&= \limsup_{(n-k)\to\infty} \left(\rho+\sum_{\sigma=0}^{\infty}U(\sigma,n-k)\right) \\
&= \rho + \limsup_{(n-k)\to\infty} \sum_{\sigma=0}^{\infty}U(\sigma,n-k) \\
&= \rho + \limsup_{n\to\infty} \sum_{\sigma=0}^{\infty}U(\sigma,n) \\
&= \limsup_{n\to\infty} \frac{S(n)}{n} + \limsup_{n\to\infty} \sum_{\sigma=0}^{\infty}U(\sigma,n) \qquad \blacksquare
\end{aligned}$$

Example 1 For service curves of the form $S(n) = \max\{0, \rho(n-D)\}$, the long-term average rate μ of $R \sim (S,U)$ is upper bounded by

$$\rho + \limsup_{n\to\infty} \sum_{\sigma=0}^{\infty} U(\sigma,n)\,.$$

Whenever we refer to a service curve of the form $S(n) = \max\{0, \rho(n-D)\}$, we assume ρ to be a positive integer, and D to be a non-negative finite integer, just for the sake of simplicity. We could have also adopted that $S(n) = \max\{0, \lfloor\rho(n-D)\rfloor\}$, however that would have cluttered the arguments in examples.

In the following section, we study the implications of definition 6 for an S-server, whereby we show some of the other properties of this characterization pursuant to the properties of a traffic characterization that we have sought to have as stated earlier in the introduction.

4.2 Implications for an S-server

If a flow $R \sim (S,U)$ is fed into an S-server, the queue size Q at the server is also similarly upper bounded by U. This is stated more precisely in the following theorem, and proved thereafter. The queue size $Q(n)$ is the total number of packets which resides in the server at time n; that is, if R and G denote the aggregates of the flows at the input and at the output of the server, respectively, then $Q(n) = R(n) - G(n)$.

Theorem 3 *If an input flow $R \sim (S,U)$ is fed into an S-server with service curve S, then the queue size Q at the server satisfies*

$$\frac{1}{n}\sum_{k<i\leqslant n+k} [Q(i) > \sigma] \leqslant U(\sigma,n)$$

for all k, $n > 0$, and σ.

Proof: Let the corresponding output flow be denoted by G. The proof follows by considering the following statements and their evaluations for any k, $n > 0$, i that $k < i \leqslant n + k$, and σ;

$$\begin{aligned}
[Q(i) > \sigma] &= [R(i) - G(i) > \sigma] \\
&\leqslant \left[R(i) - \min_{j \leqslant i}\{R(j) + S(i-j)\} > \sigma\right] \qquad (*) \\
&= \left[\max_{j \leqslant i}\{R(i) - R(j) - S(i-j)\} > \sigma\right] \\
&= \left[R(i) - R(j) - S(i-j) > \sigma, \quad \text{for some } j < i\right] \\
&= \left[R(i) - R(j) > S(i-j) + \sigma, \quad \text{for some } j < i\right]
\end{aligned}$$

hence, we get

$$\begin{aligned}
\frac{1}{n} \sum_{k < i \leqslant n+k} [Q(i) > \sigma] &\leqslant \frac{1}{n} \sum_{k < i \leqslant n+k} \left[R(i) - R(j) > S(i-j) + \sigma, \quad \text{for some } j < i\right] \qquad (*) \\
&\leqslant U(\sigma, n)\,. \qquad \blacksquare
\end{aligned}$$

Notice that this result could in fact be used to measure a level-crossing function U of an input traffic R for any given service curve S. To do this, we would need to have an S-server with equality (note the definition given right after the definition of an S-server). Then, upon feeding a flow R into a such S-server, we would observe the queue size Q at the server. The supremum of the empirical values of the quantity on the left-hand-side of the inequality in theorem 3 would give the tightest (i.e. the smallest) possible level-crossing function U in characterizing the flow R as $R \sim (S, U)$. This is facilitated by the fact that when we have an S-server with equality, the relations in lines tagged $(*)$ above in the proof of the theorem 3 become an equality.

Given a flow R with unknown characterization according to definition 6, we can find a characterization of it by utilizing theorem 2 and theorem 3, if the long-term average rate μ of R is know—it is often not difficult to get a fairly good approximation of long-term average rate of a flow by a variant of the Law of Large Numbers. By theorem 2, we could pick a service curve S that $\limsup_{n\to\infty} \frac{S(n)}{n} \geqslant \mu$. Thus, service curve $S(n) = \max\{0, \lceil \mu \rceil \cdot n\}$ is a good candidate to find a characterization of R according to definition 6, as well as service curves of the from $S(n) = \max\{0, \lceil \mu \rceil (n - D)\}$. Then, by theorem 3 and the discussions given in the previous paragraph, we could find a tight level-crossing function U to complete the characterization.

The output flow of an S-server fed by a flow $R \sim (S, U)$ could also be easily characterized according to definition 6. This is stated in the following theorem, and proved thereafter.

Theorem 4 *The output flow G of an S-server with service curve S, fed by a flow $R \sim (S, U)$, is bursty with service curve $S \,\bar{\triangledown}\, S$ and the level-crossing function U. In other words, $G \sim (S \,\bar{\triangledown}\, S, U)$.*

Proof: The proof follows by considering the following statements and their evaluations for any k, $n > 0$, i that $k < i \leqslant n + k$, $j < i$, and σ;

$$
\begin{aligned}
\Big[G(i) - G(j) > (S \mathbin{\underline{\triangledown}} S)(i-j) + \sigma\Big] &\leqslant \Big[G(i) - (R \triangledown S)(j) > (S \mathbin{\underline{\triangledown}} S)(i-j) + \sigma\Big] \\
&\leqslant \Big[R(i) - (R \triangledown S)(j) > (S \mathbin{\underline{\triangledown}} S)(i-j) + \sigma\Big] \\
&= \Big[R(i) - \min_{h \leqslant j}\{R(h) + S(j-h)\} > (S \mathbin{\underline{\triangledown}} S)(i-j) + \sigma\Big] \\
&= \Big[\max_{h \leqslant j}\{R(i) - R(h) - S(j-h)\} > (S \mathbin{\underline{\triangledown}} S)(i-j) + \sigma\Big] \\
&= \Big[R(i) - R(h) - S(j-h) > (S \mathbin{\underline{\triangledown}} S)(i-j) + \sigma, \quad \text{for some } h \leqslant j\Big] \\
&= \Big[R(i) - R(h) > (S \mathbin{\underline{\triangledown}} S)(i-j) + S(j-h) + \sigma, \quad \text{for some } h \leqslant j\Big] \\
&\leqslant \Big[R(i) - R(h) > S(i-h) - S(j-h) + S(j-h) + \sigma, \quad \text{for some } h \leqslant j\Big] \\
&= \Big[R(i) - R(h) > S(i-h) + \sigma, \quad \text{for some } h \leqslant j\Big] \\
&= \Big[R(i) - R(h) > S(i-h) + \sigma, \quad \text{for some } h \leqslant j < i\Big] \\
&\leqslant \Big[R(i) - R(h) > S(i-h) + \sigma, \quad \text{for some } h < i\Big].
\end{aligned}
$$

Taking the logical ' OR ' of the statements appearing on either side in the last line above, over all $j < i$ (note that the logical ' OR ' of the statement on the right-hand-side is just equal to itself), we get

$$
\begin{aligned}
&\Big[G(i) - G(j) > (S \mathbin{\underline{\triangledown}} S)(i-j) + \sigma, \quad \text{for some } j < i\Big] \leqslant \\
&\qquad\qquad \Big[R(i) - R(h) > S(i-h) + \sigma, \quad \text{for some } h < i\Big].
\end{aligned}
$$

Hence, we obtain the desired result as shown below, by appropriate summations and divisions;

$$
\begin{aligned}
&\frac{1}{n} \sum_{k < i \leqslant n+k} \Big[G(i) - G(j) > (S \mathbin{\underline{\triangledown}} S)(i-j) + \sigma, \quad \text{for some } j < i\Big] \leqslant \\
&\qquad\qquad \frac{1}{n} \sum_{k < i \leqslant n+k} \Big[R(i) - R(h) > S(i-h) + \sigma, \quad \text{for some } h < i\Big] \\
&\qquad\qquad \leqslant U(\sigma, n)
\end{aligned}
$$

■

We might actually need to slightly rectify the above result by replacing the service curve $S \mathbin{\underline{\triangledown}} S$ in characterizing the output flow, by S_o given below

$$
S_o(n) = \begin{cases} 0 & \text{if } n \leqslant 0 \\ (S \mathbin{\underline{\triangledown}} S)(n) & \text{else.} \end{cases}
$$

We would like to have this rectification for two reasons: (1) A service curve is defined to take on the value zero for non-negative values of its argument. (2) We would only need to have $S \mathbin{\underline{\triangledown}} S$ for positive values of its argument (as this could be noted from the first line of the proof of the above result). We have not done this rectification in the body of the theorem in order not to clutter the result.

Example 2 For service curves of the form $S(n) = \max\{0, \rho(n - D)\}$, the min-+ deconvolution of S with itself is equal to $\max\{0, \rho \cdot n\}$; whereby also note that for service curves of the form $S(n) = \max\{0, \rho \cdot n\}$ (i.e. when $D = 0$), the deconvolution of S with itself is equal to itself. $\lrcorner$

The burstiness characterization provided by definition 6 has also an implication on the *virtual-delay* at an S-server. *Virtual-delay*, which has been previously introduced in the literature, is defined below.

Definition 7 *The* virtual-delay $D(n)$ *at any time* n *for an input flow* R *at a network element is defined as*

$$D(n) = \min\{\delta : \delta \geqslant 0,\, G(n+\delta) \geqslant R(n)\}$$

where G *is the corresponding output flow.*

The virtual-delay $D(n)$ is basically the delay experienced by the packets arriving at time n, through the network element, if the packets are to be served in the order in which they have arrived.

Theorem 5 *If a flow* $R \sim (S, U)$ *is fed into an* S*-server with service curve* S*, then the virtual-delay* $D(n)$ *at the server satisfies*

$$\frac{1}{n} \sum_{k < i \leqslant n+k} [D(i) > \sigma] \leqslant U\big((S \mathbin{\bar{\vartriangle}} S)(\sigma), n\big)$$

for all k*,* $n > 0$*, and* σ*.*

Proof: The proof follows by considering the following statements and their evaluations for any k, $n > 0$, i that $k < i \leqslant n+k$, and σ;

$$\begin{aligned}
[D(i) > \sigma] &= [G(i+\sigma) < R(i)] \\
&\leqslant \big[(R \triangledown S)(i+\sigma) < R(i)\big] \\
&= \big[R(i) - (R \triangledown S)(i+\sigma) > 0\big] \\
&= \Big[R(i) - \min_{j \leqslant i+\sigma} \{R(j) + S(i+\sigma-j)\} > 0\Big] \\
&= \Big[\max_{j \leqslant i+\sigma} \{R(i) - R(j) - S(i+\sigma-j)\} > 0\Big] \\
&= \Big[R(i) - R(j) - S(i+\sigma-j) > 0, \quad \text{for some } j < i\Big]
\end{aligned}$$

(note that a j above can not be greater than or equal to i, since in that case the left-hand-side of the inequality could not become positive)

$$\begin{aligned}
&= \Big[R(i) - R(j) > S(i+\sigma-j), \quad \text{for some } j < i\Big] \\
&= \Big[R(i) - R(j) > S(i-j) + S(i-j+\sigma) - S(i-j), \quad \text{for some } j < i\Big] \qquad (*) \\
&\leqslant \Big[R(i) - R(j) > S(i-j) + (S \mathbin{\bar{\vartriangle}} S)(\sigma), \quad \text{for some } j < i\Big]
\end{aligned}$$

hence, by appropriate summations and divisions, and applying the burstiness characterization of R, we obtain the desired result as shown below;

$$\begin{aligned}
\frac{1}{n} \sum_{k < i \leqslant n+k} [D(i) > \sigma] &\leqslant \frac{1}{n} \sum_{k < i \leqslant n+k} \Big[R(i) - R(j) > S(i-j) + (S \mathbin{\bar{\vartriangle}} S)(\sigma), \quad \text{for some } j < i\Big] \\
&\leqslant U\big((S \mathbin{\bar{\vartriangle}} S)(\sigma), n\big)\,. \qquad \blacksquare
\end{aligned}$$

We could actually slightly improve the above result. This could be done if we would replace the subscript '$k \geqslant 0$' in taking the max-+ deconvolution of S with itself by '$k > 0$', as it could be noted by the line tagged $(*)$ in the derivation of the above result, since in that line we have '$j < i$'.

This result could in fact be further improved if we set a time origin for flows. That is; if we assume for almost all practical purposes that flow R satisfies $R(-\infty) = 0$, i.e. there is a certain point in time before which no packet has arrived in flow R, and call that point as the origin (i.e. $n = 0$), then we could replace '$(S \bar{\vartriangle} S)(\sigma)$' in the above result by

$$\min_{0 < u \leqslant (n+k)} \{S(u+\sigma) - S(u)\}$$

which is greater than or equal to $(S \bar{\vartriangle} S)(\sigma)$. This again could be noted by the line tagged $(*)$ in the derivation. However, if we do that, we would obtain a more time-dependent result (i.e. the first argument of the level-crossing function on the right-hand-side of the inequality in theorem 5 also depends on time n), whereas the result we have here is less time-dependent.

Example 3 For service curves of the form $S(n) = \max\{0, \rho(n-D)\}$, the max-+ deconvolution of S with itself is given by

$$(S \bar{\vartriangle} S)(n) = \begin{cases} -\rho \cdot n & \text{if } n < 0 \\ 0 & \text{if } 0 \leqslant n < D \\ \rho \cdot (n - D) & \text{else} \end{cases}$$

which is also equal to $\min\{\rho \cdot n, S(n)\}$ if we were to express it more compactly. Hence for service curves of this form, the bound on the distribution of the virtual-delay $D(n)$ is given by[7]

$$U\big((S \bar{\vartriangle} S)(\sigma), n\big) = \begin{cases} 1 & \text{if } \sigma < 0 \\ U(0, n) & \text{if } 0 \leqslant \sigma < D \\ U(\rho\,(\sigma - D), n) & \text{else.} \end{cases}$$

These results with a little bit of more work, facilitate a systematic treatment of performance guarantees in tandem networks analytically. This is the subject of a later section.

4.2.1 Average Performance Guarantees at an S-server

It immediately follows by the results in section 4.2 that we could also give average performance guarantees at an S-server with the burstiness characterization provided by definition 6. Specifically, by theorems 3 and 5, we could see that time-averaged virtual-delay and backlog are also bounded. These are pointed out by the following corollaries.

Corollary 1 *If an input flow $R \sim (S, U)$ is fed into an S-server with service curve S, the average queue size $Q(n)$ at the server is upper bounded as*

$$\frac{1}{n-k} \sum_{k < i \leqslant n} Q(i) \leqslant \sum_{\sigma=0}^{\infty} U(\sigma, n-k) \quad \textit{for all } k < n,$$

and hence, the long-term average queue size is upper bounded as

$$\limsup_{(n-k)\to\infty} \frac{1}{n-k} \sum_{k < i \leqslant n} Q(i) \leqslant \limsup_{n \to \infty} \sum_{\sigma=0}^{\infty} U(\sigma, n)\,.$$

[7] With the first improvement that we have mentioned before, we could actually replace 'D' by '$D-1$', and obtain a tighter bound on the distribution of the virtual-delay.

Proof: The proof follows immediately by theorem 3, as shown below; it holds for all $k < n$ that

$$\begin{aligned}\sum_{k<i\leqslant n} Q(i) &= \sum_{k<i\leqslant n}\sum_{\sigma\geqslant 0}[Q(i)>\sigma] \\ &= \sum_{\sigma\geqslant 0}\sum_{k<i\leqslant n}[Q(i)>\sigma] \\ &\leqslant \sum_{\sigma\geqslant 0}(n-k)U(\sigma,n-k)\end{aligned}$$

dividing both sides by $n-k$, we get

$$\frac{1}{n-k}\sum_{k<i\leqslant n} Q(i) \leqslant \sum_{\sigma\geqslant 0} U(\sigma,n-k)\,.$$

Consequently, the long-term average queue size is upper bounded as

$$\limsup_{(n-k)\to\infty}\frac{1}{n-k}\sum_{k<i\leqslant n} Q(i) \leqslant \limsup_{n\to\infty}\sum_{\sigma=0}^{\infty} U(\sigma,n)\,. \qquad \blacksquare$$

Note that, given a flow $R \sim (S,U)$ where U is the tightest possible level-crossing function, if the area under U is not finite for some n (i.e. the sum $\sum_{\sigma=0}^{\infty} U(\sigma,n) = \infty$ for some n), then when flow R is fed into an S-server with equality (note the definition given right after the definition of an S-server), the the long-term average queue size becomes unbounded.

Similarly, the time-averaged virtual-delay is also upper bounded, which is given below by the following corollary.

Corollary 2 *If an input flow $R \sim (S,U)$ is fed into an S-server with service curve S, the average virtual-delay $D(n)$ at the server is upper bounded as*

$$\frac{1}{n-k}\sum_{k<i\leqslant n} D(i) \leqslant \sum_{\sigma=0}^{\infty} U\big((S \bar{\vartriangle} S)(\sigma), n-k\big) \quad \textit{for all } k<n,$$

and hence, the long-term average virtual-delay is upper bounded as

$$\limsup_{(n-k)\to\infty}\frac{1}{n-k}\sum_{k<i\leqslant n} D(i) \leqslant \limsup_{n\to\infty}\sum_{\sigma=0}^{\infty} U\big((S \bar{\vartriangle} S)(\sigma), n\big)\,.$$

Proof: The proof follows immediately by theorem 5, as shown below; it holds for all $k < n$ that

$$\begin{aligned}\sum_{k<i\leqslant n} D(i) &= \sum_{k<i\leqslant n}\sum_{\sigma\geqslant 0}[D(i)>\sigma] \\ &= \sum_{\sigma\geqslant 0}\sum_{k<i\leqslant n}[D(i)>\sigma] \\ &\leqslant \sum_{\sigma\geqslant 0}(n-k)\,U\big((S \bar{\vartriangle} S)(\sigma), n-k\big)\end{aligned}$$

dividing both sides by $n-k$, we get

$$\frac{1}{n-k}\sum_{k<i\leqslant n} D(i) \leqslant \sum_{\sigma=0}^{\infty} U\big((S \bar{\vartriangle} S)(\sigma), n-k\big)\,.$$

Consequently, the long-term average queue size is upper bounded as

$$\limsup_{(n-k)\to\infty} \frac{1}{n-k} \sum_{k<i\leqslant n} D(i) \leqslant \limsup_{n\to\infty} \sum_{\sigma=0}^{\infty} U\big((S \bar{\wedge} S)(\sigma), n-k\big) . \qquad \blacksquare$$

Example 4 For service curves of the form $S(n) = \max\{0, \rho(n-D)\}$, we have pointed out earlier what $U\big((S \bar{\wedge} S)(\sigma), n\big)$ would be in example 3. So, for service curves of this form, the average virtual-delay $D(n)$ at an S-server is upper bounded by the above corollary as

$$\begin{aligned}
\limsup_{(n-k)\to\infty} \frac{1}{n-k} \sum_{k<i\leqslant n} D(i) &\leqslant \limsup_{n\to\infty} \sum_{\sigma=0}^{\infty} U\big((S \bar{\wedge} S)(\sigma), n\big) \\
&= D \cdot \limsup_{n\to\infty} U(0,n) + \limsup_{n\to\infty} \sum_{\sigma=D}^{\infty} U(\rho(\sigma - D), n) \\
&= D \cdot \limsup_{n\to\infty} U(0,n) + \limsup_{n\to\infty} \sum_{\sigma=0}^{\infty} U(\rho \cdot \sigma, n) .
\end{aligned}$$

Since, $U(\sigma, n) \leqslant 1$ for all σ and $n > 0$, we also have

$$\limsup_{(n-k)\to\infty} \frac{1}{n-k} \sum_{k<i\leqslant n} D(i) \leqslant D + \limsup_{n\to\infty} \sum_{\sigma=0}^{\infty} U(\rho \cdot \sigma, n) .$$

4.3 Performance Guarantees Over A Tandem of Network Elements

Performance guarantees over a tandem of network elements, in the framework of definition 6, follow almost directly from the results in section 4.2 where implications of the burstiness characterization provided by definition 6 is studied for an S-server. In order to discover such performance guarantees, we would need to generalize the results in section 4.2, slightly. These generalizations are realized when we consider feeding a flow characterized according to definition 6 with service curve S^*, into an S-server with service curve S where S is not necessarily equal to S^*. In the following subsection, we provide these generalizations. The results in this section could further be improved slightly if we pay a close attention to some of the variables, as we have just pointed out earlier after the proof of theorem 5. However, doing that would clutter the exposure of the results, hence we have chosen not to do this in this paper. Such improvements could be followed similarly as presented in the corresponding section in [12].

4.3.1 Generalized Implications for an S-server

Derivations to obtain the results in this section are similar to those of section 4.2, and can be followed in parallel. Some of the explanations that might be needed further for the results in this section could be found in section 4.2 where corresponding results are provided.

Theorem 6 *If an input flow $R \sim (S^*, U)$ is fed into an S-server with service curve S, the queue size Q at the server satisfies*

$$\frac{1}{n} \sum_{k<i\leqslant n+k} [Q(i) > \sigma] \leqslant U\big((S \bar{\wedge} S^*)(0) + \sigma, n\big)$$

for all k, $n > 0$, and σ.

Proof: Let the corresponding output flow be denoted by G. The proof follows by considering the following statements and their evaluations for any k, $n > 0$, i that $k < i \leqslant n + k$, and σ;

$$
\begin{aligned}
[Q(i) > \sigma] &= [R(i) - G(i) > \sigma] \\
&\leqslant \left[R(i) - \min_{j \leqslant i}\{R(j) + S(i-j)\} > \sigma\right] \\
&= \left[\max_{j \leqslant i}\{R(i) - R(j) - S(i-j)\} > \sigma\right] \\
&= \left[R(i) - R(j) - S(i-j) > \sigma, \quad \text{for some } j < i\right] \\
&= \left[R(i) - R(j) > S(i-j) + \sigma, \quad \text{for some } j < i\right] \\
&= \left[R(i) - R(j) > S^*(i-j) + S(i-j) - S^*(i-j) + \sigma, \quad \text{for some } j < i\right] \\
&\leqslant \left[R(i) - R(j) > S^*(i-j) + (S \barwedge S^*)(0) + \sigma, \quad \text{for some } j < i\right]
\end{aligned}
$$

hence, we get

$$
\begin{aligned}
\frac{1}{n} \sum_{k < i \leqslant n+k} [Q(i) > \sigma] &\leqslant \frac{1}{n} \sum_{k < i \leqslant n+k} \left[R(i) - R(j) > S^*(i-j) + (S \barwedge S^*)(0) + \sigma, \quad \text{for some } j < i\right] \\
&\leqslant U\big((S \barwedge S^*)(0) + \sigma, n\big) \qquad \blacksquare
\end{aligned}
$$

Note that if service curve S of the S-server is greater than or equal to S^* of the flow at every point in time, then the bound on the queue size distribution becomes $U(\sigma, n)$.

Example 5 For service curves of the form

$$
\begin{aligned}
S^*(n) &= \max\{0, \rho^*(n - D^*)\} \\
S(n) &= \max\{0, \rho(n - D)\}
\end{aligned}
$$

note the following calculations:

1. If $\rho^* > \rho$, then we have $(S \barwedge S^*)(0) = -\infty$. Hence, the bound of the queue size as given in theorem 6 becomes 1, and no use. Thus, one assume for practical purposes that $\rho^* \leqslant \rho$.

2. For $\rho^* \leqslant \rho$, if $D^* \geqslant D$, then $(S \barwedge S^*)(0) = 0$. In this case, the bound of the queue size as given in theorem 6 becomes equal to $U(\sigma, n)$.

3. For $\rho^* \leqslant \rho$ and $D^* < D$, we have $(S \barwedge S^*)(0) = -\rho^*(D - D^*)$. Hence, the bound of the queue size as given in theorem 6 becomes equal to $U(\sigma - \rho^*(D - D^*), n)$.

The generalized result for the output flow of an S-server fed by a flow $R \sim (S^*, U)$ is given below. We would again leave the rectification of the service curve in the characterization of the output flow after the theorem, as we have done before for the corresponding result in theorem 4.

Theorem 7 *The output flow G of an S-server with service curve S, fed by a flow $R \sim (S^*, U)$, is bursty with service curve $S^* \veebar S$ and the level-crossing function U. In other words, $G \sim (S^* \veebar S, U)$.*

Proof: The proof follows by considering the following statements and their evaluations for any k, $n > 0$, i that $k < i \leqslant n + k$, $j < i$, and σ;

$$
\begin{aligned}
\Big[G(i) - G(j) > (S^* \underline{\triangledown} S)(i-j) + \sigma\Big] \leqslant \Big[G(i) - (R \triangledown S)(j) > (S^* \underline{\triangledown} S)(i-j) + \sigma\Big]& \\
\leqslant \Big[R(i) - (R \triangledown S)(j) > (S^* \underline{\triangledown} S)(i-j) + \sigma\Big]& \\
= \Big[R(i) - \min_{h \leqslant j}\{R(h) + S(j-h)\} > (S^* \underline{\triangledown} S)(i-j) + \sigma\Big]& \\
= \Big[\max_{h \leqslant j}\{R(i) - R(h) - S(j-h)\} > (S^* \underline{\triangledown} S)(i-j) + \sigma\Big]& \\
= \Big[R(i) - R(h) - S(j-h) > (S^* \underline{\triangledown} S)(i-j) + \sigma, \quad \text{for some } h \leqslant j\Big]& \\
= \Big[R(i) - R(h) > (S^* \underline{\triangledown} S)(i-j) + S(j-h) + \sigma, \quad \text{for some } h \leqslant j\Big]& \\
\leqslant \Big[R(i) - R(h) > S^*(i-h) - S(j-h) + S(j-h) + \sigma, \quad \text{for some } h \leqslant j\Big]& \\
= \Big[R(i) - R(h) > S^*(i-h) + \sigma, \quad \text{for some } h \leqslant j\Big]& \\
= \Big[R(i) - R(h) > S^*(i-h) + \sigma, \quad \text{for some } h \leqslant j < i\Big]& \\
\leqslant \Big[R(i) - R(h) > S^*(i-h) + \sigma, \quad \text{for some } h < i\Big]&.
\end{aligned}
$$

Taking the logical ‘ OR ’ of the statements appearing on either side in the last line above, over all $j < i$ (note that the logical ‘ OR ’ of the statement on the right-hand-side is just equal to itself), we get

$$
\begin{aligned}
\Big[G(i) - G(j) > (S^* \underline{\triangledown} S)(i-j) + \sigma, \quad \text{for some } j < i\Big] \leqslant& \\
\Big[R(i) - R(h) > S^*(i-h) + \sigma, \quad \text{for some } h < i\Big]&.
\end{aligned}
$$

Hence, we obtain the desired result as shown below, by appropriate summations and divisions;

$$
\begin{aligned}
\frac{1}{n} \sum_{k < i \leqslant n+k} \Big[G(i) - G(j) > (S^* \underline{\triangledown} S)(i-j) + \sigma, \quad \text{for some } j < i\Big] \leqslant& \\
\frac{1}{n} \sum_{k < i \leqslant n+k} \Big[R(i) - R(h) > S^*(i-h) + \sigma, \quad \text{for some } h < i\Big]& \\
\leqslant U(\sigma, n)& \qquad \blacksquare
\end{aligned}
$$

We might again need to slightly rectify the above result by replacing the service curve $S^* \underline{\triangledown} S$ in characterizing the output flow, by S_o given below

$$
S_o(n) = \begin{cases} 0 & \text{if } n \leqslant 0 \\ (S^* \underline{\triangledown} S)(n) & \text{else.} \end{cases}
$$

We would like to have this rectification by the same reasons given for the corresponding rectification done after theorem 4.

Example 6 For service curves of the form

$$
\begin{aligned}
S^*(n) &= \max\{0, \rho^*(n - D^*)\} \\
S(n) &= \max\{0, \rho(n - D)\}
\end{aligned}
$$

note the following calculations:

1. If $\rho^* > \rho$, regardless of the values of D^* and D, $(S^* \veebar S)(n)$ becomes infinite for any positive n. In this case, the result we have here does not provide any useful information for burstiness characterization of the output flow, since in that case its bounding function could be set as $U(\sigma, n) = 0$ for all $\sigma \geqslant 0$ and for all $n > 0$, without loss of generality. Thus, we can assume for practical purposes that $\rho^* \leqslant \rho$.
2. With this assumption, regardless of the values of D^* and D, we have

$$(S^* \veebar S)(n) = \min\{0, \rho^*(n - (D^* - D))\} \quad \text{for all } n.$$

Note however that the shape of S_o will be different for $D^* \geqslant D$ and $D^* < D$.

The generalized result for virtual-delays experienced at an S-server fed by a flow $R \sim (S^*, U)$ is given below. The definition of virtual delay is given by definition 7.

Theorem 8 *If a flow $R \sim (S^*, U)$ is fed into an S-server with service curve S, then the virtual-delay $D(n)$ at the server satisfies*

$$\frac{1}{n} \sum_{k<i\leqslant n+k} [D(i) > \sigma] \leqslant U\big((S \,\underline{\vartriangle}\, S^*)(\sigma), n\big)$$

for all k, $n > 0$, and σ.

Proof: The proof follows by considering the following statements and their evaluations for any k, $n > 0$, i that $k < i \leqslant n + k$, and σ;

$$\begin{aligned}
[D(i) > \sigma] &= [G(i+\sigma) < R(i)] \\
&\leqslant \big[(R \triangledown S)(i+\sigma) < R(i)\big] \\
&= \big[R(i) - (R \triangledown S)(i+\sigma) > 0\big] \\
&= \Big[R(i) - \min_{j \leqslant i+\sigma} \{R(j) + S(i+\sigma-j)\} > 0\Big] \\
&= \Big[\max_{j \leqslant i+\sigma} \{R(i) - R(j) - S(i+\sigma-j)\} > 0\Big] \\
&= \Big[R(i) - R(j) - S(i+\sigma-j) > 0, \quad \text{for some } j < i\Big]
\end{aligned}$$

(note that a j above can not be greater than or equal to i, since in that case the left-hand-side of the inequality could not become positive)

$$\begin{aligned}
&= \Big[R(i) - R(j) > S(i+\sigma-j), \quad \text{for some } j < i\Big] \\
&= \Big[R(i) - R(j) > S^*(i-j) + S(i-j+\sigma) - S^*(i-j), \quad \text{for some } j < i\Big] \\
&\leqslant \Big[R(i) - R(j) > S(i-j) + (S \,\underline{\vartriangle}\, S^*)(\sigma), \quad \text{for some } j < i\Big]
\end{aligned}$$

hence, by appropriate summations and divisions, and applying the burstiness characterization of R, we obtain the desired result as shown below;

$$\begin{aligned}
\frac{1}{n} \sum_{k<i\leqslant n+k} [D(i) > \sigma] &\leqslant \frac{1}{n} \sum_{k<i\leqslant n+k} \Big[R(i) - R(j) > S(i-j) + (S \,\underline{\vartriangle}\, S^*)(\sigma), \quad \text{for some } j < i\Big] \\
&\leqslant U\big((S \,\underline{\vartriangle}\, S^*)(\sigma), n\big) .
\end{aligned}$$

$\blacksquare$

Example 7 For service curves of the form

$$\begin{aligned} S^*(n) &= \max\{0, \rho^*(n - D^*)\} \\ S(n) &= \max\{0, \rho(n - D)\} \end{aligned}$$

note the following calculations:

1. If $\rho^* > \rho$, then we have $(S \mathbin{\overline{\vartriangle}} S^*)(\sigma) = -\infty$ for any $\sigma \geqslant 0$. Hence, the bound of the virtual-delay as given in theorem 8 becomes 1, and no use. Thus, one assume for practical purposes that $\rho^* \leqslant \rho$.

2. For $\rho^* \leqslant \rho$, if $D^* \geqslant D$, then $(S \mathbin{\overline{\vartriangle}} S^*)(\sigma) = S(\sigma)$. In this case, the bound on the virtual-delay as given in theorem 8 becomes $U(S(\sigma), n)$.

3. For $\rho^* \leqslant \rho$ and $D^* < D$, we have for $\sigma \geqslant 0$

$$\begin{aligned} &\text{if } \sigma \leqslant D - D^* && (S \mathbin{\overline{\vartriangle}} S^*)(\sigma) = -\rho^*(D - D^* - \sigma) \\ &\text{if } \sigma > D - D^* && (S \mathbin{\overline{\vartriangle}} S^*)(\sigma) = S(\sigma)\,. \end{aligned}$$

   Hence, the bound on the virtual-delay as given in theorem 8 becomes

$$\text{for } \sigma \leqslant D - D^* \quad U\big((S \mathbin{\overline{\vartriangle}} S^*)(\sigma), n\big) = \begin{cases} 1 & \text{if } \sigma < D - D^* \\ U(0, n) & \text{if } \sigma = D - D^* \end{cases}$$

$$\begin{aligned} \text{for } \sigma > D - D^* \quad U\big((S \mathbin{\overline{\vartriangle}} S^*)(\sigma), n\big) &= U\big(S(\sigma), n\big) \\ &= \begin{cases} U(0, n) & \text{if } \sigma < D \\ U\big(\rho(\sigma - D), n\big) & \text{else.} \end{cases} \end{aligned}$$

All of the theorems in this section essentially facilitate a systematic analytical treatment of performance guarantees over a tandem of network elements. To further clarify this, we provide the results in the following subsection.

4.3.2 Considering a Tandem of Network Elements

In order to clarify the claimed systematic treatment of performance guarantees over a tandem of network elements, we would like to give the following two lemmas and three corollaries.

Lemma 2 *Let f, g, and h be any three functions. The following inequality holds*

$$((f \mathbin{\underline{\triangledown}} g) \mathbin{\underline{\triangledown}} h)(n) \leqslant (f \mathbin{\underline{\triangledown}} (g \triangledown h))(n) \quad \textit{for all } n.$$

Proof: The proof follows directly from the definitions of min-+ deconvolution and min-+ convolution, and is given below. It holds for all n that

$$\begin{aligned} ((f \mathbin{\underline{\triangledown}} g) \mathbin{\underline{\triangledown}} h)(n) &= \max_{k \geqslant 0} \{(f \mathbin{\underline{\triangledown}} g)(n + k) - h(k)\} \\ &= \max_{k \geqslant 0} \left\{ \max_{l \geqslant 0} \{f(n + k + l) - g(l)\} - h(k) \right\} \\ &= \max_{k \geqslant 0} \left\{ \max_{l \geqslant 0} \{f(n + k + l) - g(l) - h(k)\} \right\} \\ &= \max_{\substack{k \geqslant 0 \\ l \geqslant 0}} \{f(n + k + l) - (g(l) + h(k))\} \end{aligned}$$

notice that for all k and l such that $k+l$ would remain the same, $f(n+k+l)$ does not change its value for any given n, hence the maximum above for that fixed value of $k+l$ will occur for the minimum value of $g(l)+h(k)$ over all such k's and l's; that is, we have

$$
\begin{aligned}
&= \max_{k+l\geqslant 0}\Big\{f(n+k+l) - \min_{0\leqslant u\leqslant k+l}\big\{g(u)+h(k+l-u)\big\}\Big\} \\
&\leqslant \max_{k+l\geqslant 0}\Big\{f(n+k+l) - \min_{u\leqslant k+l}\big\{g(u)+h(k+l-u)\big\}\Big\} \\
&= \max_{k+l\geqslant 0}\big\{f(n+k+l) - (g\,\nabla h)(k+l)\big\} \\
&= \max_{u\geqslant 0}\big\{f(n+u) - (g\,\nabla h)(u)\big\} \\
&= (f\,\overline{\nabla}\,(g\,\nabla h))(n)\,. \qquad \blacksquare
\end{aligned}
$$

Lemma 3 *Let f, g, and h be any three functions. The following inequality holds*

$$(f\,\overline{\Delta}\,(g\,\overline{\nabla}\,h))(n) \geqslant ((f\,\nabla h)\,\overline{\Delta}\,g)(n) \quad \textit{for all } n.$$

Proof: The proof follows directly from the definitions of max-+ and min-+ deconvolutions, and is given below. It holds for all n that

$$
\begin{aligned}
(f\,\overline{\Delta}\,(g\,\overline{\nabla}\,h))(n) &= \min_{k\geqslant 0}\{f(n+k) - (g\,\overline{\nabla}\,h)(k)\} \\
&= \min_{k\geqslant 0}\Big\{f(n+k) - \max_{l\geqslant 0}\big\{g(k+l) - h(l)\big\}\Big\} \\
&= \min_{k\geqslant 0}\Big\{f(n+k) + \min_{l\geqslant 0}\big\{-g(k+l) + h(l)\big\}\Big\} \\
&= \min_{\substack{k\geqslant 0\\ l\geqslant 0}}\big\{f(n+k) - g(k+l) + h(l)\big\} \\
&= \min_{\substack{k\geqslant 0\\ l\geqslant 0}}\big\{f(n+k) + h(l) - g(k+l)\big\}
\end{aligned}
$$

notice that for all k and l such that $k+l$ would remain the same, $g(k+l)$ does not change its value, hence the minimum above for that fixed value of $k+l$ will occur for the minimum value of $f(n+k)+h(l)$ over all such k's and l's; that is, we have

$$
\begin{aligned}
&= \min_{k+l\geqslant 0}\Big\{\min_{0\leqslant u\leqslant n+k+l}\big\{f(n+k+l-u)+h(u)\big\} - g(k+l)\Big\} \\
&\geqslant \min_{k+l\geqslant 0}\Big\{\min_{u\leqslant n+k+l}\big\{f(n+k+l-u)+h(u)\big\} - g(k+l)\Big\} \\
&= \min_{k+l\geqslant 0}\big\{(f\,\nabla h)(n+k+l) - g(k+l)\big\} \\
&= \min_{u\geqslant 0}\big\{(f\,\nabla h)(n+u) - g(u)\big\} \\
&= ((f\,\nabla h)\,\overline{\Delta}\,g)(n)\,. \qquad \blacksquare
\end{aligned}
$$

The next result is a corollary of theorem 1.

Corollary 3 *Let Z_1 and Z_2 be two processes that*

$$\frac{1}{n}\sum_{k<i\leqslant n+k}[Z_j(i)>\sigma] \leqslant U_j(\sigma,n) \quad \textit{for all } k,\ n>0,\ \sigma, \textit{ and for } j \textit{ equals to both } 1 \textit{ and } 2$$

where functions U_j's are as specified after definition 6. The following relation holds

$$\frac{1}{n}\sum_{k<i\leqslant n+k}[(Z_1+Z_2)(i)>\sigma]\leqslant(U_1\triangledown U_2)(\sigma,n)\quad \textit{for all } k,\ n>0,\ \textit{and } \sigma,$$

where the convolution above is carried out over the first arguments (i.e. σ) of U_j's.

Proof: The proof follows immediately by theorem 1. This could be seen from its proof if we were to shovel the terms involving service curves appearing in the definitions of the statements A and A_j's to the other side of the inequalities, and view the left-hand-side of those inequalities as Z_1+Z_2 and Z_j's, respectively. ■

We now give the main result of this section, stated next as a corollary.

Corollary 4 *Let a flow $R_1\sim(S^*,U)$ be fed into an S-server with service curve S_1. And let the output R_2 of this first server be fed into another S-server with service curve S_2. The following statements hold:*

1. *The output flow R_3 of the S-server with service curve S_2 is bursty with service curve S_3 and level-crossing function U, where*

$$S_3(n)=\begin{cases}0 & \textit{if } n\leqslant 0\\ (S^*\,\overline{\triangledown}\,(S_1\triangledown S_2)) & \textit{else.}\end{cases}$$

   In other words, $R_3\sim(S_3,U)$.

2. *The total number of packets, $Q_1(n)+Q_2(n)$, stored in the tandem network satisfies*

$$\frac{1}{n}\sum_{k<i\leqslant n+k}[(Q_1+Q_2)(i)>\sigma]\leqslant(g\triangledown h)(\sigma,n)$$

   for all k, $n>0$, and σ, where

$$g(\sigma,n)=U\big((S_1\,\overline{\vartriangle}\,S^*)(0)+\sigma,n\big)$$

$$\begin{aligned}h(\sigma,n)&=U\big((S_2\,\overline{\vartriangle}\,(S^*\,\overline{\triangledown}\,S_1))(0)+\sigma,n\big)\\&=U\big(((S_1\triangledown S_2)\,\overline{\vartriangle}\,S^*)(0)+\sigma,n\big)\,.\end{aligned}$$

3. *The total virtual-delay, $D_1(n)+D_2(n+\delta(n))$, experienced by a packet arriving at any time n at the first network element, and at time $n+\delta(n)$ for some $\delta(n)\geqslant 0$ at the second network element, satisfies for any $\delta(n)\geqslant 0$*

$$\frac{1}{n}\sum_{k<i\leqslant n+k}[D_1(i)+D_2(i+\delta(i))>\sigma]\leqslant(g\triangledown h)(\sigma,n)$$

   for all k, $n>0$, and σ, where

$$g(\sigma,n)=U\big((S_1\,\overline{\vartriangle}\,S^*)(\sigma),n\big)$$

$$\begin{aligned}h(\sigma,n)&=U\big((S_2\,\overline{\vartriangle}\,(S^*\,\overline{\triangledown}\,S_1))(\sigma),n\big)\\&=U\big(((S_1\triangledown S_2)\,\overline{\vartriangle}\,S^*)(\sigma),n\big)\,.\end{aligned}$$

Proof: The proof of

- statement 1 follows by theorem 7 and lemma 2, where notice for lemma 2 that the inequality in the lemma becomes an equality since S_1 and S_2 here are service curves;

- statement 2 follows by corollary 3, theorem 6, and lemma 3, where notice for lemma 3 that the inequality in the lemma becomes an equality since S_1 and S_2 here are service curves;
- statement 3 follows by corollary 3, theorem 8, and lemma 3, where again notice for lemma 3 that the inequality in the lemma becomes an equality since S_1 and S_2 here are service curves. ■

We could actually slightly improve the above results in items 2 and 3 in corollary 4, and further emphasize the message that we are trying to convey by corollary 4. It is not difficult to show that the bounds that would be obtained by replacing $(g \triangledown h)(\sigma, n)$ in items 2 and 3 in corollary 4 by $h(\sigma, n)$, also hold. This is given as a corollary next, for which we give the following lemmas first.

Both of the lemmas are on the properties of min-+ convolution. The first one is on the monotonicity of the min-+ convolution, and is given below.

Lemma 4 *Let f, g, and h be any there functions that $f(n) \leqslant h(n)$ for all n. There holds*

$$(f \triangledown g)(n) \leqslant (h \triangledown g)(n) \quad \textit{for all } n.$$

Proof: The proof follows immediately from the definition of min-+ convolution, and is given below for the sake of completeness; it holds for all n that

$$\begin{aligned}(f \triangledown g)(n) &= \min_{k \leqslant n}\{f(k) + g(n-k)\} \\ &\leqslant \min_{k \leqslant n}\{h(k) + g(n-k)\} \\ &= (h \triangledown g)(n)\,.\end{aligned}$$

■

The second one is on the associativity of the min-+ convolution, and is given below.

Lemma 5 *Let f, g, and h be any there functions. There holds*

$$((f \triangledown g) \triangledown h)(n) \geqslant (f \triangledown (g \triangledown h))(n) \quad \textit{for all } n.$$

Proof: The proof follows immediately from the definition of min-+ convolution, and is given below for the sake of completeness; it holds for all n that

$$\begin{aligned}((f \triangledown g) \triangledown h)(n) &= \min_{k \leqslant n}\{(f \triangledown g)(k) + h(n-k)\} \\ &= \min_{k \leqslant n}\Big\{\min_{l \leqslant k}\{f(l) + g(k-l)\} + h(n-k)\Big\} \\ &= \min_{k \leqslant n}\Big\{\min_{l \leqslant k}\{f(l) + g(k-l) + h(n-k)\}\Big\} \\ &= \min_{\substack{k \leqslant n \\ l \leqslant k}}\Big\{f(l) + g(k-l) + h(n-k)\Big\}\end{aligned}$$

notice that for a fixed l, $f(l)$ does not change its value, hence the minimum above for that value of l will occur for the minimum of $g(k-l) + h(n-k)$ over all k's; that is, we have

$$\begin{aligned}&= \min_{l \leqslant n}\Big\{f(l) + \min_{l \leqslant k \leqslant n}\{g(k-l) + h(n-k)\}\Big\} \\ &= \min_{l \leqslant n}\Big\{f(l) + \min_{0 \leqslant k-l \leqslant n-l}\{g(k-l) + h(n-k)\}\Big\} \\ &= \min_{l \leqslant n}\Big\{f(l) + \min_{0 \leqslant u \leqslant n-l}\{g(u) + h(n-l-u)\}\Big\} \\ &\geqslant \min_{l \leqslant n}\Big\{f(l) + \min_{u \leqslant n-l}\{g(u) + h(n-l-u)\}\Big\} \\ &= \min_{l \leqslant n}\{f(l) + (g \triangledown h)(n-l)\} \\ &= (f \triangledown (g \triangledown h))(n)\,.\end{aligned}$$

■

We would like to note that the relation in lemma 5 is in general not an equality due to the way the min-+ convolution is defined in this study (more specifically, due to the subscript of the minimum in definition 1 as being '$k \leqslant n$').

Corollary 5 *Let a flow $R_1 \sim (S^*, U)$ be fed into an S-server with service curve S_1. And let the output R_2 of this first server be fed into another S-server with service curve S_2. The following statements hold:*

1. *The total number of packets, $Q_1(n) + Q_2(n)$, stored in the tandem network satisfies*

$$\frac{1}{n} \sum_{k<i\leqslant n+k} [(Q_1 + Q_2)(i) > \sigma] \leqslant h(\sigma, n)$$

   for all k, $n > 0$, and σ, where

$$h(\sigma, n) = U\big(((S_1 \triangledown S_2) \,\underline{\vartriangle}\, S^*)(0) + \sigma, n\big)\,.$$

2. *The total virtual-delay, $D_1(n) + D_2(n + \delta(n))$, experienced by a packet arriving at any time n at the first network element, and at time $n + \delta(n)$ for some $\delta(n) \geqslant 0$ at the second network element, satisfies for any $\delta(n) \geqslant 0$*

$$\frac{1}{n} \sum_{k<i\leqslant n+k} [D_1(i) + D_2(i + \delta(i)) > \sigma] \leqslant h(\sigma, n)$$

   for all k, $n > 0$, and σ, where

$$h(\sigma, n) = U\big(((S_1 \triangledown S_2) \,\underline{\vartriangle}\, S^*)(\sigma), n\big)\,.$$

Proof: The proofs essentially follow by theorems 6 and 8, and lemmas 4 and 5.

First, notice by lemmas 4 and 5 that we have the following inequality which holds for all n for the output flow R_3 of the S-server with service curve S_2;

$$\begin{aligned} R_3(n) &\geqslant (R_2 \triangledown S_2)(n) \\ &\geqslant ((R_1 \triangledown S_1) \triangledown S_2)(n) \quad \text{by lemma 4} \\ &= (R_1 \triangledown (S_1 \triangledown S_2))(n) \quad \text{by lemma 5} \end{aligned}$$

note that the last relation above is an equality since S_1 and S_2 here are service curves.

- (Proof for item 1) The proof follows by replacing $Q(i)$ by $(Q_1 + Q_2)(i)$, $R(i)$ by $R_1(i)$, $G(i)$ by $R_3(i)$, and using the above lower-bound on $R_3(n)$ (and hence effectively replacing $S(i - j)$ by $(S_1 \triangledown S_2)(i - j)$), in the proof of theorem 6.

- (Proof for item 2) The proof follows by replacing $D(i)$ by $D_1(i) + D_2(i + \delta(i)$, $G(i + \sigma)$ by $R_3(i + \sigma)$, $R(i)$ by $R_1(i)$, and using the above lower-bound on $R_3(n)$ (and hence effectively replacing $S(i - j + \sigma)$ by $(S_1 \triangledown S_2)(i - j + \sigma)$), in the proof of theorem 8. ∎

Hence, it is now clear that the bounds in the last two items in corollaries 4 and 5 could actually be replaced by

$$\min\{h(\sigma, n), (g \triangledown h)(\sigma, n)\}$$

for all σ and $n > 0$. It is also clear by the first item in corollary 4 and the whole corollary 5 that we could view that tandem network as a single S-server with service curve $S_1 \triangledown S_2$, and obtain valid characterizations.

We could obtain similar results for any number of network elements in tandem, by a repeated application of corollaries 4 and 5. Thus, it has now become clear that the burstiness characterization 6 does indeed facilitate a systematic treatment of performance guarantees over a tandem of network elements, analytically.

Finally, one might want to demonstrate the results stated in corollaries 4 and 5 by an example set below, which is left as an exercise.

Example 8 A flow $R_1 \sim (S^*, U)$ is fed into an S-server with service curve S_1. The output R_2 of this first server is fed into another S-server with service curve S_2. The service curves S^*, S_1, and S_2 are as given below

$$\begin{aligned} S^*(n) &= \max\{0, \rho^*(n - D^*)\} \\ S_1(n) &= \max\{0, \rho_1(n - D_1)\} \\ S_2(n) &= \max\{0, \rho_2(n - D_2)\} \end{aligned}$$

where we assume $\rho^* \leqslant \min\{\rho_1, \rho_2\}$ for the practical reasons given in examples 5, 6, and 7. We are interested in finding the characterizations of the output flow R_3, the total queue size of the tandem network, and the total virtual-delay experienced through the tandem network, in the framework of our definition, as given by corollaries 4 and 5. One can work out the example for any choices of D^*, D_1, and D_2, again facilitated by the examples 5, 6, and 7.

Average performance guarantees for a tandem of network elements could also be easily calculated via the results in this section, as we have presented in section 4.2.1, similarly.

5 Discussions About The New Characterization

Note that the burstiness characterization provided by definition 6 facilitates analyses for both average and scalar worst-case performance guarantees at the same time. Average performance guarantees could be obtained in general as exemplified by the results in section 4.2.1. Scalar worst-case performance guarantees, on the other hand, are provided by the respective level-crossing functions whose second arguments (i.e. n) are all set equal to 1. A relevant scalar worst-case performance bound is given by the the minimum of the first argument of the corresponding level-crossing function $U(*, 1)$ that $U(*, 1) = 0$.

Secondly, note that the burstiness definition that we propose here also facilitates a systematic analytical framework for measurement based analysis probabilistic performance guarantees that would be inferred via sample-path analyses. In this sense, definition 6 is directly applicable to such measurement studies. This is made possible, since the implications of definition 6 (such as theorem 6 and 8) could also be viewed from the stand point of relative frequency interpretation of probability. However, we would like to note that the real starting point in getting definition 6 is the discussions given for the notion of burstiness via the concept of level-crossing, as presented in section 3.

Lastly, we would like to point out another contrast between the characterization that we propose here and the characterization in [1]: Consider feeding a flow $R \sim (S, U)$ into an S-server with service curve S. By theorem 3, we know that the queue size Q at the server satisfies

$$\frac{1}{n} \sum_{k < i \leqslant n+k} [Q(i) > \sigma] \leqslant U(\sigma, n) \qquad \text{for all } k,\ n > 0, \text{ and } \sigma.$$

Multiplying both sides by n, we get

$$\sum_{k < i \leqslant n+k} [Q(i) > \sigma] \leqslant n \cdot U(\sigma, n) \quad \text{for all } k,\ n > 0, \text{ and } \sigma.$$

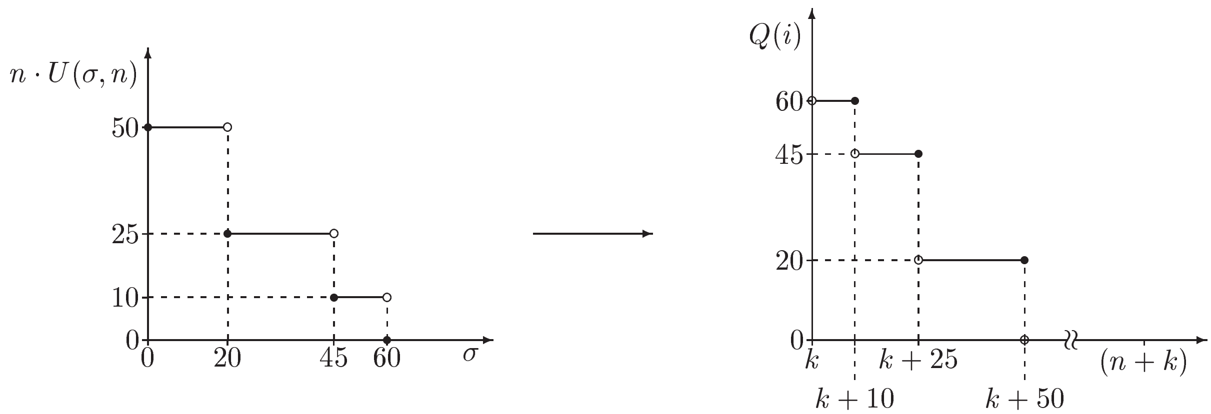

Figure 2: An "extremal arrival pattern" for the queue size, for some k and $n > 50$.

Now, if we rotate $n \cdot U(\sigma, n)$ +90 degrees (in the plane on which it is drawn, with respect to the $(0, 0)$ point) and flip it with respect to the vertical axis after the rotation (in which case this will be the axis for σ), and after same appropriate changes in order to have a corresponding queue size function, we would get an "extremal arrival pattern" for the queue size as exemplified in Figure 2 where time-slots are drawn very close to each other for convenience.

Scalar worst-case performance bounds in analysis such as those presented in [1], are achieved for extremal *input* flows whose arrival patterns, in principal, look like the figure on the right in Figure 2. Whereas, in our analysis we consider such "extremal arrival patterns" for the *queue sizes*. Although, one could always obtain a corresponding extremal input flow for an "extremal arrival pattern" of a queue size that we consider here, we believe that it would help if this contrast between the characterization that we propose here and the characterization in [1] has been pointed out.

6 Conclusions

In this study, we have proposed a deterministic definition of burstiness for network traffic characterization, based on service curves. The proposed definition facilitates performance analyses for both average and scalar worst-case performance guarantees at the same time, with ease. Average performance guarantees could be obtained as exemplified by the results in section 4.2.1. Scalar worst-case performance guarantees, on the other hand, are provided by the respective level-crossing functions whose second arguments (i.e. n) are all set equal to 1. A relevant scalar worst-case performance bound is given by the the minimum of the first argument of the corresponding level-crossing function $U(*, 1)$ that $U(*, 1) = 0$.

We have shown that the traffic characterization provided by the proposed definition 6 satisfies all the properties of a traffic characterization that we have sought to have as stated earlier in the introduction. Specifically, we have shown that the new characterization facilitates a systematic analytical approach to performance guarantees as also presented by the characterization in [1] and its companion service model (*service curve* model) that we have mentioned earlier.

The characterization of a flow according to this definition is also measurable by constructing an S-server with equality, and observing the queue size upon feeding a flow into that server.

The burstiness definition that we propose here also facilitates a systematic analytical framework for measurement based analysis of probabilistic performance guarantees that would be inferred via sample-path analyses. In this sense, definition 6 is directly applicable to such measurement studies. This is made possible, since the implications of definition 6 (such as theorem 6 and 8) could also be viewed from the stand point of relative frequency interpretation of probability.

We have also tried to clarify the notion of burstiness. We have pointed out that one might want to perceive the burstiness of a flow from the perspective of a network element; specifically, based on the queue size behavior that it induces on a network element of interest.

To this end, we have indicated that it is the decay rate of the tail of the queue size distribution that we care about in deciding the degree of burstiness of a flow with respect to another one, after some appropriate normalizations of the flows as we have indicated earlier. The faster the decay rate is the less bursty the traffic is, and vice versa. More specifically, we could decide the burstiness of a traffic source A with respect to another one B, from the perspective of a network element of interest, as follows: Normalize the sources such that the average rate of traffic coming out of source A is equal to that of source B. Feed the traffic coming out of each source A and B into identical network elements α and β, respectively, where the average rate that each network element serves packets is grater than that of its input. Observe the queue size distribution in each network element α and β. If there exists a queue size level σ_0 beyond which the decay rate of the tail of the queue size distribution in network element α is slower than that of network element β, then source A is more bursty than source B.

As a future study, it would be interesting to examine a similar traffic characterization and its implications, where the new characterization would also have a lower-bound, as well as an upper-bound, on the quantity given in definition 6. It would be also interesting to incorporate *maximum service curves* into our analysis, in which case a service curve that we have referred to in this study so far would be called as *minimum service curve*. A network element is said to deliver a *maximum service curves* S to an input flow R if the corresponding output flow G satisfies

$$G(n) \leqslant (R \triangledown S)(n) \quad \text{for all } n.$$

References

[1] R. L. Cruz. *A Calculus For Network Delay, Part I: Network Elements In Isolation*, IEEE Transactions on Information Theory, vol. 37, no. 1, pp. 114–131, Jan. 1991.

[2] D. Ferrari, D. Verma. *A Scheme for Real-Time Channel Establishment in Wide-Area Networks*, IEEE Journal on Selected Areas in Communications, vol. 8, pp. 368–379, April 1990.

[3] Z. Wang, J. Crowcroft. *Analysis of Burstiness and Jitter in Real-Time Communications*, Proceedings of SIGCOMM, pp. 13–19, 1993.

[4] A. K. Parekh, R. G. Gallagher. *A generalized processor sharing approach to flow control in integrated services networks: the single-node case*, IEEE/ACM Transaction on Networking, vol. 1, pp. 344–357, 1993.

[5] R. L. Cruz. *Quality of Service Guarantees in Virtual Circuit Switched Networks*, IEEE Journal of Selected Areas in Communication, 13(6): 1048–1056, 1995.

[6] C.-S. Chang. *Performance Guarantees in Communication Networks*, Springer Verlag, April 2000.

[7] J.-Y. Le Boudec, P. Thiran. *Network Calculus: A Theory of Deterministic Queuing Systems for the Internet*, Lecture Notes in Computer Science-2050, Springer Verlag, January 2002.

[8] G. Kesidis, T. Konstantopoulos. *Extremal shape-controlled traffic patterns in high-speed networks*, IEEE Transactions on Communications, vol. 48, no. 5, pp. 813–819, May 2000.

[9] F. Guillemin, N. Likhanov, R. Mazumdar, C. Rosenberg. *Extremal traffic and bounds on the mean delay of multiplexed regulated traffic streams*, Proceedings of the INFOCOM 2002, N.Y., June 2002.

[10] E. Reich. *On the Integrodifferential Equation of Takasc, I*, Ann. Math. Stat., vol. 29, pp. 563–570, 1958.

[11] S. Ayyorgun, W.-C. Feng. *Notes on Burstiness and Bursty Traffic Characterization*, Los Alamos National Laboratory, Technical Report.

[12] S. Ayyorgun, W.-C. Feng. *A Probabilistic Definition of Burstiness Characterization: A Systematic Approach*, Los Alamos National Laboratory, Technical Report LA-UR-03-3668, May 2003.

[13] R. L. Graham, D. E. Knuth, O. Patashnik. *Concrete Mathematics: A Foundation for Computer Science*, Addison-Wesley, 2nd ed., 1994.