Available as Los Alamos Unclassified Report LA-UR 03-7268

A New Traffic Model and Its Implications over Multiplexers and Switches *

Sami Ayyorgun Wu-chun Feng

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

E-mail: {sami,feng}@lanl.gov

Abstract—We propose a new traffic model and show its implications over multiplexers and switches. The analysis assumes neither a scheduling policy employed in multiplexers or switches nor statistical independence of input flows. The proposed model $\mathcal{M}$ (i) facilitates a tractable and systematic analysis of probabilistic QoS guarantees in communication networks, (ii) is general enough to characterize traffic in any stochastic process model including self-similar models, (iii) allows the regulation of arbitrary traffic to any specified parameters in $\mathcal{M}$ by a *probabilistic leaky bucket scheme*, which contributes to the viability of the framework put forth by our results for QoS management in communication networks, and (iv) yields tighter bounds compared to the ones provided by some other approaches.

1 Introduction

About a decade ago, network traffic in various packet networks such as local-area networks was observed not to agree with the widely accepted and adopted Markovian traffic models. Instead, traffic was found to exhibit *self-similar* and/or *multi-fractal* behavior. This has been confirmed for various network systems and traffic, e.g., [1, 2, 3, 4].

These observations have led researchers to model network traffic as self-similar stochastic processes, such as *Fractional Gaussian Noise*, by fitting parameters (e.g., *Hurst* parameter) to observed traffic. Performance analyses with self-similar stochastic processes, however, have proved to be highly difficult and intractable, e.g., refer to [5, 6, 7].

The tractability of performance analyses is crucial for the deployment of services which require *provable* Quality of Service (QoS) guarantees. There is a prevailing agreement among network researchers and practitioners that such services are needed, as well as in demand, for broadband networks (e.g., refer to [8]). Moreover, it is well-known that QoS guarantees are needed to be expressed in probabilistic terms for an efficient utilization of network resources.

Recently, we have proposed a new traffic model [9]. In this model, the characterization of traffic is based *directly* on the queue-size behavior that the traffic induces on fundamentally important network elements (such as *work-conserving* servers), instead of first picking a particular stochastic process to model the traffic and later examining essentially its implications on the queue-size behavior. This way, we eliminate many of the mathematical complications forbidding tractable performance analyses.

Another advantage of our traffic model is that it does not depend on or adopt any particular stochastic process model. Thus, it could easily characterize traffic in any stochastic process model including self-similar models.

*Los Alamos National Laboratory, Technical Report LA-UR-03-7268, September 2003.

In [9, 10] together with this study, we have shown that the model we propose satisfies the following properties: Let $\mathcal{M}$ denote the traffic model that we propose, then

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

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

   (a) the output flow is also easily characterized according to $\mathcal{M}$, and

   (b) both the queue-size and the virtual delay are easily characterized in the same framework as $\mathcal{M}$, as well,

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

4. the characterization of traffic according to $\mathcal{M}$ is measurable, and easily measured,

5. the model $\mathcal{M}$ captures the behavior of traffic in real networks as closely as possible, without sacrificing either tractability or reality.

The first three properties facilitate a tractable and systematic analysis of probabilistic QoS guarantees in communication networks. Together with the last two properties, they make the framework that we propose a viable solution to provide probabilistic QoS guarantees in communication networks. The focus of this study is to show that property 2 holds too when network elements are multiplexers or switches.

To verify the above properties and to see how applicable our model is in real networks, we have carried out some experimental and simulation studies as well. The simulations that we have performed have utilized the traffic collected from the FDDI[1]-based wide-area network of the Los Alamos National Laboratory. A subset of these studies is presented in [10]; a comprehensive study will be made available in the near future. Due to space limitations, we are unable to present simulation results in this text.

In short, our experimental and simulation studies show agreement with the theoretical results and that the proposed traffic model is capable of addressing the network traffic with arbitrary *heavy-tailed* distributions, including the *power-tailed* distributions, in performance analyses.

Besides the above properties, a fundamental advantage of the model $\mathcal{M}$ that we propose is that it allows the regulation of arbitrary traffic to any specified parameters in $\mathcal{M}$ by a *probabilistic leaky bucket scheme*; see [10] for details. This contributes greatly to the viability of the framework put forward by our results for QoS management in networks.

Finally, the probabilistic bounds provided by our model are tighter compared to the ones provided by some other approaches, for example the ones in [11, 12].

The rest of the paper is organized as follows: Section 2 provides background. Section 3 introduces the new traffic model and reports two of its implications proven in [9]. Sections 4 and 5 show the implications of the new traffic model over multiplexers and switches, respectively. Finally, Section 6 gives the conclusions.

[1] Fiber Distributed Data Interface.

2 Background

We adopt a discrete-time formulation for simplicity. 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 discrete random process such that each sample-path 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/session*. A *network element* is an input-output device which accepts packets at its input and delivers them at its output. The capacity $c(n)$ of a network element at time n is the total number of packets that it can serve at that time. Finally, all functions are assumed to be defined from the integers to the integers, unless otherwise noted from here on.

We utilize the following definitions which essentially have been introduced in the literature previously (e.g., see [13, 14] and the references therein).

Definition 1 *Let f and g be any two functions. The* min-+ convolution $f \triangledown g$ *of f and 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".

Definition 2 *Let f and g be any two functions. The* min-+ deconvolution $f \underline{\triangledown} g$ *of f with 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".

Definition 3 *Let f and g be any two functions. The* max-+ convolution $f \vartriangle g$ *of f with 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".

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

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

The deconvolution $f \overline{\vartriangle} g$ 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, where $S(n) = 0$ for all $n \leqslant 0$.

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

A *work-conserving*[2] server with a constant integer capacity of serving ρ packets per time slot is an S-server with equality with service curve $S(n) = \max\{0, \rho \cdot n\}$, since the output flow G of a work-conserving server with an input flow R is given[3] by

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

which is equivalently represented as

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

An S-server could be viewed as a generalization of a work-conserving server with constant rate.

The concept of service curve is introduced in [17, 18]. Service curves together with an appropriate traffic characterization facilitate a systematic treatment of performance guarantees in communication networks; like in linear system theory. This has been demonstrated for some deterministic regimes, e.g., refer to [13, 14, 18, 19, 20]. We carry this much desired feature into a probabilistic setting. For various applications of the service curve model, such as multimedia smoothing, refer to [14], for example.

3 The New Traffic Model

We propose the following probabilistic characterization of network traffic [9].

Definition 6 *A traffic flow R is said to be bursty with service curve S and bounding function f, and denoted as $R \sim (S, f)$, if*

$$\mathrm{P}\Big(R(n) - R(k) > S(n-k) + \sigma, \text{ for some } k < n\Big) \leqslant f(\sigma)$$

for all σ and for all n, where f is defined from the integers to the non-negative real numbers.

We assume that the following properties hold for any bounding function f, without loss of generality:

1. f is non-increasing, since the probability corresponding to a σ in the above definition is non-increasing with σ.
2. $f(\sigma) \leqslant 1$ for all σ, since the probability of an event can not be larger than 1. We assume for mathematical convenience that $f(\sigma) = 1$ for all $\sigma < 0$, unless otherwise noted from here on.
3. $\lim_{\sigma\to\infty} f(\sigma) = 0$, since any cumulative distribution function F satisfies $\lim_{x\to\infty} F(x) = 1$.

The traffic model that we propose provided by definition 6 utilizes the concept of service curves.

As we have indicated before, the characterization provided by definition 6 satisfies the properties stated in the introduction; see [9, 10]. In this section, we report only two of the implications of definition 6, which we utilize in this study; the first one is on aggregate flows and the other is on output flows. For other results, see [9, 10].

The traffic characterization provided by definition 6 satisfies property 1 listed in the introduction. This is given by the following theorem.

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

[3] This is often referred to as Reich's result [15]; see [16], for example, for a simple derivation.

Theorem 1 *Given any two flows $R_1 \sim (S_1, f_1)$ and $R_2 \sim (S_2, f_2)$, the aggregate flow $R_1 + R_2$ is bursty with service curve $S_1 + S_2$ and bounding function $f_1 \triangledown f_2$. In other words, $R_1 + R_2 \sim (S_1 + S_2, f_1 \triangledown f_2)$.*

Proof: (See [9] or Appendix[4].)

By repeatedly applying theorem 1, we could easily find the characterization of an aggregate flow with any number of flows being aggregated.

The output flow of an S-server fed by a flow $R \sim (S^*, f)$ could also be easily characterized according to definition 6. This is given by the following theorem.

Theorem 2 *The output flow G of an S-server with service curve S, fed by a flow $R \sim (S^*, f)$, is bursty with service curve $S^* \underline{\triangledown} S$ and the bounding function f. In other words, $G \sim (S^* \underline{\triangledown} S, f)$.*

Proof: (See [9] or Appendix[4].)

We might actually 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 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 $S^* \underline{\triangledown} S$ for positive values of its argument (this could be seen clearly via 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.

4 Implications over Multiplexers

A *multiplexer* is a network element with more than one input and a single output. The function of a multiplexer is to combine flows at its input into a single flow at its output. Packets arriving at the input ports of a multiplexer could be stored in either individual buffers allocated for each input or a common buffer. In this study, we assume individual (receive) buffers for each input, in which case we also assume a buffer space for the output. The capacity $c(n)$ of a multiplexer at time n is the maximum number of packets that it could deliver at its output. Having made this abstraction for the capacity, we assume without loss of generality that a multiplexer pulls packets from its input buffers instantaneously for transmission at its output. With that, we can view a multiplexer as a single-input, single-output network element where the input flow is equal to the aggregate of the flows incident onto the multiplexer. In other words, a multiplexer operates on the aggregate flow $R(n) = R_1(n) + R_2(n) + \cdots + R_N(n)$ where R_i's are its input flows. Similar modeling descriptions to the one we have made here were effectively made in the literature previously, e.g., [11, 19].

A multiplexer is said to deliver a service curve S to the aggregate flow R of its input flows, *if and only if*

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

where G is the output flow. Note that for a work-conserving multiplexer with constant capacity rate C packets per time-slot, the output flow G is given by $G(n) = (R \triangledown S)(n)$ where $S(n) = \max\{0, C \cdot n\}$.

[4]Appendix is provided solely to help reviewers.

With the above explanations, it is clear that we could easily characterize the output flow of a multiplexer, as well as give probabilistic bounds on the total queue-size and the virtual-delay, by utilizing theorems 1, 2, and other results in [10].

For instance, the output flow G of a multiplexer delivering a service curve S to the aggregate of input flows $R_i \sim (S_i, f_i)$, $i \in I = \{1, 2, \ldots, N\}$, is bursty with service curve $S^* \,\underline{\triangledown}\, S$ where $S^* = \sum_{i \in I} S_i$ and bounding function $f = f_1 \triangledown f_2 \triangledown \cdots \triangledown f_N$. In other words, $G \sim (S^* \,\underline{\triangledown}\, S, f)$.

5 Implications over Switches

We adopt the *switch* model in [11] due to its analytical functionality. Results along the lines presented in this section could also be obtained for other modeling descriptions (e.g., output queueing) and switching methods, similarly.

A *switch* is a network element with more than one input and more than one output. The function of a switch is to redirect the input flows at its input to predetermined output ports. More than one input flow could enter a switch through an input port, as well as depart from an output port. Packets arriving at the input ports of a switch are temporarily stored in individual receive buffers allocated for each input, which then are instantaneously (i.e., with infinite capacity) copied down into a common buffer. Similarly, a packet to be transmitted from an output port is instantaneously copied up from the common buffer to the transmit buffer at the respective output port. The capacity $c(n)$ of a switch at time n is the maximum number of packets that it could deliver at its output.

In this section, we will only show the characterization of output flows of a switch, which is given by the following theorem. Results for the queue-size and virtual-delay could be derived similarly.

Theorem 3 *Given a set of flows $R_i \sim (S_i, f_i)$, $i \in \{1, 2, \ldots, N\}$, $N > 1$, input to a switch having a constant capacity C packets per time slot, the output flow G_i corresponding to input flow R_i, for each i in I, is bursty with service curve S_i and bounding function $(f_1 \triangledown f_2 \triangledown \cdots \triangledown f_N)(\sigma + A_i)$ where $A_i = \big(S \,\overline{\vartriangle}\, \big[(\sum_{k \in I - \{i\}} S_k) + (S_i \,\underline{\triangledown}\, S_i)\big]\big)(0)$ and $S(n) = \max\{0, C \cdot n\}$.*

Notice that if $S(n) \geqslant \sum_{i \in I} S_i(n)$ for all n, then the constants A_i's in the theorem all become 0.

Proof: The characterization of any output flow is found similarly. Thus, we present the result for flow 1.

In order to find the characterization of G_1, we need to appropriately upper-bound the probability of the following event E

$$E = \big\{G_1(n) - G_1(k) > S_1(n-k) + \sigma, \text{ for some } k < n\big\} \tag{1}$$

which is equivalently represented as

$$\bigcup_{k<n} \big\{G_1(n) - G_1(k) > S_1(n-k) + \sigma\big\}. \tag{2}$$

Let us define a random variable $z(k)$ for any time k as the last time before k (inclusive) that the buffer of the switch was empty. That is, if we denote the buffer-size at time n by $B(n)$ where $B(n) = \big(\sum_{i \in I} R_i(n)\big) - \big(\sum_{i \in I} G_i(n)\big)$, then $z(k)$ is given by

$$z(k) \triangleq \max\{u : u \leqslant k, B(u) = 0\}\,.$$

The events whose union is taken in (2) could be represented as

$$\big\{G_1(n) - G_1(k) > S_1(n-k) + \sigma\big\} = \bigcup_{l \leqslant k} \Big(\big\{G_1(n) - G_1(k) > S_1(n-k) + \sigma\big\} \cap \{z(k) = l\}\Big). \quad (3)$$

Furthermore, the following relations hold for the event whose union is taken in (3);

$$\begin{aligned}
&\big\{G_1(n) - G_1(k) > S_1(n-k) + \sigma\big\} \cap \{z(k) = l\} \\
&\subseteq \Big\{R_1(n) - R_1(l) + \sum_{2 \leqslant i \leqslant N} \big(R_i(k) - R_i(l)\big) > S_1(n-k) + \sigma + C \cdot (k-l)\Big\} \\
&= \Big\{R_1(n) - R_1(l) + \sum_{2 \leqslant i \leqslant N} \big(R_i(k) - R_i(l)\big) > S_1(n-l) + S_1(n-k) - S_1(n-l) + \\
&\qquad\qquad \sum_{2 \leqslant i \leqslant N} S_i(k-l) - \sum_{2 \leqslant i \leqslant N} S_i(k-l) + \sigma + C \cdot (k-l)\Big\} \\
&= \Big\{R_1(n) - R_1(l) + \sum_{2 \leqslant i \leqslant N} \big(R_i(k) - R_i(l)\big) > S_1(n-l) + \sum_{2 \leqslant i \leqslant N} S_i(k-l) + \sigma + \\
&\qquad\qquad C \cdot (k-l) - \big(S_1(n-l) - S_1(n-k)\big) - \sum_{2 \leqslant i \leqslant N} S_i(k-l)\Big\} \\
&\subseteq \Big\{R_1(n) - R_1(l) + \sum_{2 \leqslant i \leqslant N} \big(R_i(k) - R_i(l)\big) > S_1(n-l) + \sum_{2 \leqslant i \leqslant N} S_i(k-l) + \sigma + \\
&\qquad\qquad C \cdot (k-l) - \Big[(S_1 \mathbin{\underline{\triangledown}} S_1)(k-l) + \sum_{2 \leqslant i \leqslant N} S_i(k-l)\Big]\Big\} \\
&\subseteq \Big\{R_1(n) - R_1(l) + \sum_{2 \leqslant i \leqslant N} \big(R_i(k) - R_i(l)\big) > S_1(n-l) + \sum_{2 \leqslant i \leqslant N} S_i(k-l) + \sigma + A_1\Big\},
\end{aligned}$$

where A_1 is as defined in the body of the theorem.

With the above result, the event E could be bounded as

$$\begin{aligned}
E &\subseteq \Big\{R_1(n) - R_1(l) + \sum_{2 \leqslant i \leqslant N} \big(R_i(k) - R_i(l)\big) > S_1(n-l) + \sum_{2 \leqslant i \leqslant N} S_i(k-l) + \sigma + A_1, \text{ for some } l \leqslant k < n\Big\} \\
&\subseteq \big\{R_1(n) - R_1(l) > S_1(n-l) + u_1,\ \exists\, l \leqslant k < n\big\} \cup \\
&\qquad\qquad \Big(\bigcup_{2 \leqslant i \leqslant N} \big\{R_i(k) - R_1(l) > S_i(k-l) + u_i,\ \exists\, l \leqslant k < n\big\}\Big),
\end{aligned}$$

where $\sum_{i \in I} u_i = \sigma + A_1$ and $u_i \geqslant 0$ for all $i \in I - \{1\}$,

$$\subseteq \big\{R_1(n) - R_1(l) > S_1(n-l) + u_1,\ \exists\, l < n\big\} \cup \Big(\bigcup_{2 \leqslant i \leqslant N} \big\{R_i(k) - R_1(l) > S_i(k-l) + u_i,\ \exists\, l < k\big\}\Big).$$

Now, taking the probability of both sides of the last line above, using the Union Bound, and utilizing the characterizations of each flow R_i, we get

$$\mathrm{P}(E) \leqslant \sum_{i \in I} f_i(u_i)\,.$$

Since the above inequality holds for any choice of u_i's as specified earlier, we also have

$$\begin{aligned} \mathrm{P}(E) &\leqslant \min_U \left\{ \sum_{i \in I} f_i(u_i) \right\} \qquad (4) \\ &= (f_1 \triangledown f_2 \triangledown \cdots \triangledown f_N)(\sigma + A_1), \qquad (5) \end{aligned}$$

where U is the statement specifying u_i's. A few remarks follow for (4) and (5): (First) We have allowed a slight abuse of notation in (4). To be strict, we should have used infimum 'inf' above instead of minimum 'min', since (a) f_i's are real-valued, and (b) the minimum could potentially be taken over an infinite number of terms. However, note that the minimum (or infimum) is taken effectively over a set of finite number of elements due to the fact that $f_i(s) = 1$ for all $s < 0$ and $\lim_{s \to \infty} f_i(s) = 0$ for all $i \in I$. (Second) Note that the min-+ convolution the way we have defined it is in general not commutative for any two functions. However, due to the properties of the bounding functions mentioned in the previous remark, it becomes commutative for bounding functions. Thus, the convolution expressed in equality (5) is meaningful. (Third) Again due to the properties of the bounding functions mentioned in the first remark note that we have $(f_1 \triangledown f_2 \triangledown \cdots \triangledown f_N)(\sigma + A_1) = 1$ for all $\sigma < -A_1$. This completes the proof. ■

6 Conclusion

We have proposed a new traffic model provided by definition 6 and showed its implications over multiplexers and switches.

The proposed model $\mathcal{M}$ **(i)** facilitates a tractable and systematic analysis of probabilistic QoS guarantees in communication networks [9, 10], **(ii)** is general enough to characterize traffic in any stochastic process model including self-similar models, **(iii)** allows a regulation of arbitrary traffic to any specified parameters in $\mathcal{M}$ by a *probabilistic leaky bucket scheme*, see [10] for details, which contributes greatly to the viability of the framework put forward by our results for QoS management in communication networks, **(iv)** yields tighter bounds compared to the ones provided by some other approaches, for example the ones in [11, 12].

In addition to above properties, note that in getting the results in Sections 4 and 5, we assumed neither a scheduling policy employed in multiplexers or switches nor statistical independence of input flows. This is a fundamental advantage of the traffic characterization that we propose, as it facilitates a systematic approach to probabilistic QoS guarantees in a network environment by overcoming two past problems: (1) having to assume that all multiplexers or switches employ a common scheduling policy, and (2) having to assume that input flows to multiplexers or switches are independent. We note that a significant portion of the works in statistical QoS, e.g., [21, 22, 23], suffer from assuming independent input flows.

Future work includes (i) a comprehensive experimental and simulation study mentioned in the introduction, a small part of which is made available in [10], (ii) a study providing the details of the *probabilistic leaky bucket system* explained in [10], (iii) a study for statistical multiplexing of traffic that are characterized in the proposed model, which request separate QoS guarantees (e.g., delay)—the current work fits better into DiffServ, whereas the future study is for IntServ; refer to IETF website for information on IntServe and DiffServe.

Appendix

The Proof of Theorem 1: The proof follows by considering the following events for any n, σ, and $u \leqslant \sigma$;

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

Notice that $A \subseteq A_1 \cup A_2$, since the intersection of the complements of both of the events A_1 and A_2 is clearly a subset of the complement of the event A. Thus, by utilizing the Union Bound and the fact that $R_1 \sim (S_1, f_1)$ and $R_2 \sim (S_2, f_2)$, we get

$$\begin{aligned} \mathrm{P}(A) &\leqslant \mathrm{P}(A_1 \cup A_2) \\ &\leqslant \mathrm{P}(A_1) + \mathrm{P}(A_2) \\ &\leqslant f_1(u) + f_2(\sigma - u). \end{aligned}$$

Since the last inequality above holds for any $u \leqslant \sigma$, we also get

$$\begin{aligned} \mathrm{P}(A) &\leqslant \min_{u \leqslant \sigma}\{f_1(u) + f_2(\sigma - u)\} \\ &= (f_1 \triangledown f_2)(\sigma). \end{aligned}$$

We have allowed a slight abuse of notation in the last two lines above. To be strict, we should have used infimum 'inf' above instead of minimum 'min', since (1) f_i's are real-valued, and (2) the minimum could potentially be taken over an infinite number of terms. However, note that the minimum (or infimum) is taken effectively over a set of finite number of elements due to the fact that $f_i(s) = 1$ for all $s < 0$ and $\lim_{s \to \infty} f_i(s) = 0$ for i equals to both 1 and 2. More specifically, the min-+ convolution $(f_1 \triangledown f_2)(\sigma)$ is equal to the minimum of the terms corresponding to each integer u in $[0, \sigma]$ and the term 1.

Finally note that we also have $(f_1 \triangledown f_2)(\sigma) = 1$ for all $\sigma < 0$, again due to the properties of a bounding function mentioned above. This completes the proof. ∎

The Proof of Theorem 2: The proof follows by considering the following events for any n,

$k < n$, and σ;

$$
\begin{aligned}
\Big\{G(n) - G(k) > (S^* \overline{\triangledown} S)(n-k) + \sigma\Big\} &\subseteq \Big\{G(n) - (R \triangledown S)(k) > (S^* \overline{\triangledown} S)(n-k) + \sigma\Big\} \\
&\subseteq \Big\{R(n) - (R \triangledown S)(k) > (S^* \overline{\triangledown} S)(n-k) + \sigma\Big\} \\
&= \Big\{R(n) - \min_{l \leqslant k}\{R(l) + S(k-l)\} > (S^* \overline{\triangledown} S)(n-k) + \sigma\Big\} \\
&= \Big\{\max_{l \leqslant k}\{R(n) - R(l) - S(k-l)\} > (S^* \overline{\triangledown} S)(n-k) + \sigma\Big\} \\
&= \Big\{R(n) - R(l) - S(k-l) > (S^* \overline{\triangledown} S)(n-k) + \sigma, \quad \text{for some } l \leqslant k\Big\} \\
&= \Big\{R(n) - R(l) > (S^* \overline{\triangledown} S)(n-k) + S(k-l) + \sigma, \quad \text{for some } l \leqslant k\Big\} \\
&\subseteq \Big\{R(n) - R(l) > S^*(n-l) - S(k-l) + S(k-l) + \sigma, \quad \text{for some } l \leqslant k\Big\} \\
&= \Big\{R(n) - R(l) > S^*(n-l) + \sigma, \quad \text{for some } l \leqslant k\Big\} \\
&= \Big\{R(n) - R(l) > S^*(n-l) + \sigma, \quad \text{for some } l \leqslant k < n\Big\} \\
&\subseteq \Big\{R(n) - R(l) > S^*(n-l) + \sigma, \quad \text{for some } l < n\Big\}.
\end{aligned}
$$

Taking the union of both sides over all $k < n$ (note that the union of the right-hand-side is equal to itself), we obtain the desired result as shown below

$$
\begin{aligned}
\mathrm{P}\Big(G(n) - G(k) > (S \overline{\triangledown} S)(n-k) + \sigma, \quad \text{for some } k < n\Big) \leqslant& \\
&\mathrm{P}\Big(R(n) - R(l) > S(n-l) + \sigma, \quad \text{for some } l < n\Big) \\
&\leqslant f(\sigma)\,. \qquad \blacksquare
\end{aligned}
$$

References

[1] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson. *On the self-similar nature of Ethernet traffic*, Proc. of ACM Sigcomm, pp. 183–193, 1993. (Extended version: IEEE/ACM Transactions on Networking, vol. 2, no. 1, pp. 1–15, 1994.)

[2] J. Beran, R. Sherman, M. S. Taqqu, and W. Willinger. *Long-range dependence in variable bit-rate video traffic*, IEEE/ACM Transactions on Communications, vol. 43, pp. 1566–1579, 1995.

[3] M. Crovella, A. Bestavros. *Self-similarity in World Wide Web traffic: Evidence and possible causes*, Performance Evaluation Review, vol. 24, pp. 160–169, 1996.

[4] V. Paxson, S. Floyd. *Wide-area traffic: The failure of Poisson modeling*, IEEE/ACM Transactions on Networking, vol. 3, pp. 226–244, 1995.

[5] K. Park, W. Willinger (editors). *Self-Similar Network Traffic and Performance Evaluation*, John Wiley & Sons, 2000.

[6] I. Norros. *A storage model with self-similar input*, Queueing Systems, vol. 16, pp. 387–396, 1994.

[7] B. Tsybakov, N. D. Georganas. *Self-similar traffic and upper bounds to buffer overflow in an ATM queue*, Performance Evaluation, vol. 32, no. 1, pp. 57–80, Feb. 1998.

[8] A. K. Parekh. *Breaking the Great Internet Deadlock OR Why there isn't any Network QoS and what to do about it?* (invited talk), 11th International Workshop on Quality of Service (IWQoS), 2003. `http://iwqos03.cs.berkeley.edu/invited-talks/parekh.pdf`

[9] 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.

[10] S. Ayyorgun, W.-C. Feng. *A Systematic Approach to Probabilistic Quality of Service Guarantees in Communication Networks*, Los Alamos National Laboratory, Technical Report, Sept. 2003.

[11] O. Yaron, M. Sidi. *Performance and stability of communication networks via robust exponential bounds*, IEEE/ACM Transactions on Networking, 1(3): 372–385, 1993.

[12] D. Starobinski, M. Sidi. *Stochastically Bounded Burstiness for Communication Networks*, IEEE Trans. on Information Theory, vol. 46, no. 1, Jan. 2000. (Also appeared in the Proceedings of Infocom 1999, pp. 36–42.)

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

[14] 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.

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

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

[17] 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.

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

[19] 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.

[20] S. Ayyorgun, W.-C. Feng. *A Deterministic Definition of Burstiness For Network Traffic Characterization*, Los Alamos National Laboratory, Technical Report LA-UR-03-4477, June 2003.

[21] R. Boorstyn, A. Burchard, J. Liebeherr, C. Oottamakorn. *Effective Envelopes: Statistical Bounds on Multiplexed Traffic in Packet Networks*, in Proceedings of IEEE Infocom 2000, March 2000. (Also: *Statistical Multiplexing Gain of Link Scheduling Algorithms in QoS Networks-Short Version*, University of Virginia, Department of Computer Science, CS-99-23, July 1999.)

[22] N. Likhanov, R. R. Mazumdar. *Cell-loss Asymptotics in Buffers Fed with a Large Number of Independent Stationary Sources*, Infocom, 1998.

[23] H. S. Kim, N. B. Shroff. *Loss Probability Calculations and Asymptotic Analysis for Finite Buffer Multiplexers*, IEEE/ACM Trans. on Networking, vol. 9, no. 6, pp. 755–768, 2001.