Queuing Processes in GPS and PGPS with LRD Traffic Inputs

Xiang Yu†, Ian Li-Jin Thng‡, Yuming Jiang* and Chunming Qiao†
†Department of Computer Science and Engineering
State University of New York at Buffalo
‡Department of Electrical and Computer Engineering
National University of Singapore
*Institute for Infocomm Research, Singapore

Abstract— **Long range dependent (LRD) traffic whose single server queue process is Weibull Bounded (WB) is first analyzed. Two upper bounds on the individual session's queue length of LRD traffic under the generalized processor sharing (GPS) scheduling discipline are then contributed. It is shown that the index parameter in the upper bound of one LRD flow, (in addition to the decay rate and the asymptotic constant), may be affected by other LRD flows. A new concept, called LRD isolation, is subsequently contributed and accompanying it, a new technique is contributed to check whether a flow, with a given GPS weight assignment, can be guaranteed to be LRD isolated. This technique is also amenable for use in an online call admission control (CAC) scenario. When existing flows have already been assigned contract weights that cannot be changed, our technique can be used to determine minimum contract weights to be assigned to a new flow in order to guaranteed the flow to be LRD isolated. The results are also extended to a PGPS (Packet-based GPS) scheduler and relevant numerical results are provided to show the usefulness of our bounds and LRD isolation technique.**

I. Introduction

Scheduling disciplines used by switches or routers are important to provide QoS (quality of service) support for integrated applications comprising of voice, video and data traffic. An ideal scheduling discipline should satisfy two requirements: (a) Provide isolation between sessions (where isolation means that the queuing process behaves no worse than its Single Server Queue (SSQ) process with a comparable service rate). This guarantees that the scheduling discipline is able to protect an individual flow[1] against misbehavior from other flows. (b) Realize statistical multiplexing gain. This suggests a flow can utilize excess service rate allocated to others.

Perhaps the most widely studied non-FCFS scheduling discipline is the Generalized Processor Sharing (GPS) discipline. GPS has two attractive characteristics: (1) each backlogged session is guaranteed a minimum service rate. This ensures that the misbehavior of other flows has a limited effect on an individual session, and provides the foundation of isolation between sessions. Achieving isolation further enables GPS to guarantee differentiated QoS for individual sessions. (2) it is work-conserving, and any excess service rate can be redistributed amongst backlogged flows. The second characteristic enables GPS to obtain statistical multiplexing gain between input flows. Because of these two characteristics, GPS is deemed an ideal scheduling discipline that can satisfy the two requirements (a) and (b) mentioned above, and thus has attracted a lot of research interests. When GPS is extended to packet switched networks, it is usually referred to as Weighted Fair Queuing (WFQ) or Packet-based GPS (PGPS) [2].

Long range dependent (LRD) traffic is an increasingly important class of traffic in modern day networks because long range dependency is exhibited in Ethernet traffic [3], WWW traffic [4], compressed video traffic [5,6], TCP traffic [7] and so on. Since LRD traffic has burstiness extending over various time scales, a Weibull bound rather than a conventional exponential bound is usually associated with LRD traffic's SSQ process [8]. Achieving differentiated QoS in a GPS queue remains a challenging task, especially for LRD traffic. Prior analysis of GPS are often based on input traffic streams which satisfy some well known burstiness constraints that are either deterministic [9] or stochastic [10, 11] in nature. In particular, the conventional stochastic traffic, whose associated queue length distribution has an exponential form, has been demonstrated to have an Exponential Bounded Burstiness (EBB) arrival process [12]. Based on such constraints on the arrival process, an upper bound has been developed for the individual session queue length in a GPS system [10, 11]. However, *few analysis of GPS based on LRD arrival processes have been carried out.* This paper contributes new insight into the behavior of the GPS server when scheduling several LRD flows with arrival process exhibiting what we call *Weibull Bounded Burstiness* (WBB) constraints.

The relationship between LRD traffic and the Weibull bound expression stems much from empirical studies demonstrating that the queuing process of a work conservative SSQ, whose input traffic exhibits LRD behavior, can be commonly characterized by a Weibull Bounded (WB) expression [13]. The work presented here is applicable to LRD traffic that are WB in its SSQ process (although not all LRD traffic can be WB in its SSQ process). Note that the term "WB" refers to the queuing process while the term

Work supported by MOE Research Project R-263-000-078-112 and Singaren Research Project R-263-000-202-305. A preliminary version of this paper appeared in IEEE Globecom [1].

[1] which will be used interchangeably with the term "session".

"WBB" we introduced earlier refers to the arrival process. The relationship between a WBB arrival process and a WB queuing process will be established in later sections.

In order to achieve QoS in a GPS system whose inputs are LRD flows, issues related to isolation between LRD flows in GPS systems need to be carefully characterized. Recent research contributions by Borst [14–16] have shown that under certain conditions, a LRD (or heavy tailed) traffic flow can be well isolated from other input flows in a GPS system [14,16] while in some other scenarios, it may inherit burstiness from other flows [15]. *However, good theoretical bounds for an individual session queue length in the GPS system remain to be found.* More specifically, while Borst's work is valuable in providing a method to judge whether an individual session can be isolated from other input sessions under the GPS scheduling discipline, the problem of how to guarantee the isolation of an individual session by tuning GPS parameters, or how to determine the QoS performance of an individual session that is not isolated from other sessions remains unsolved.

In this paper, we derive two bounds for individual queuing processes in a GPS system with LRD traffic inputs, each with its own advantages and disadvantages (to be spelt out in detail later). These bounds provide valuable insights into the isolation between multiple GPS sessions. More specifically, it is shown that the index parameter in the upper bound of one LRD flow, (in addition to the decay rate and the asymptotic constant), may be affected by other LRD flows. In addition, we introduce a useful notion called LRD isolation which differs from the usual notion of flow isolation in that a flow is guaranteed to be LRD isolated from other flows as long as its index parameter is not affected by other flows. We also develop a necessary and sufficient condition for a flow i being guaranteed to be LRD isolated from other flows. Based on this condition, a technique that can be used to quickly check if flow i can be guaranteed to be LRD isolated from other flows with a given (GPS service) weight assignment is proposed. When some flows have already been assigned contract weights (according to some Service Level Agreement or SLA) that cannot be changed, the proposed technique can also be used to determine the minimum contract weight to be assigned to flow i in order for it to be guaranteed to be LRD isolated from other flows, and thus is especially useful for on-line call admission control (CAC). The results are also extended to a PGPS (Packet-based GPS) scheduler so that numerical simulations can be presented to illustrate the usefulness of the bounds and the proposed technique.

Our contributions are different from those presented in [17, 18] which considered short range dependent models. The work in [19] studied Stochastically Bounded (SB) SSQ and Stochastically Bounded Burstiness (SBB) arrival processes. One of the *differences* is that while SB and SBB are discrete time processes, WB and WBB processes studied in this work are continuous time processes. In addition, the WB and WBB bounds developed in this paper are specialized bounds compared to SB and SBB bounds as the former two have an index parameter, while the latter two do not. Accordingly, not only are the results contributed in [19] *not* directly applicable, it is also nontrivial to derive the WB and WBB bounds (as well as other results in this paper).

Reference [20] did provide statistical bounds for queuing delay in GPS systems for the LRD traffic scenario. However, its focus was on the admission region of GPS with statistical resource sharing rather than the isolation between individual queuing processes. The bounds provided were not derived from the queue length distribution but were obtained via certain input/output traffic envelope approximation and a Gaussian distribution for each class of traffic.

Another significant divergence from conventional LRD traffic analysis which usually begins with associated queuing process is that *we begin our derivations from the arrival process* of LRD traffic. This brings more validity in all subsequent queue analysis, since only when the arrival process is bounded, the stability of the queuing process can be ensured.

The paper is organized as follows. In Section II, we review fundamental knowledge related to conventional LRD traffic analysis and the GPS discipline. In Section III, we present the relationship between a WB SSQ and a WBB arrival process, and then derive several upper bounds on the aggregate WB SSQ processes. In Sections IV, the concept of feasible ordering is reviewed, and the notion of LRD isolation is introduced. Two upper bounds related to the individual session queue length in a GPS system are also derived. In Section V, a necessary and sufficient condition to guarantee a flow i to be LRD isolated from other flows is contributed, and a technique for determining a minimum contract weight to be assigned to a flow i to ensure its LRD isolation is also proposed and illustrated via numerical examples. Results in GPS are further extended to PGPS system in Section VI, and in Section VII, we demonstrate, via numerical simulation of a two-input PGPS system, the usefulness of the two bounds. Section VIII concludes the paper.

II. Preliminaries

In this section, we briefly review the fundamental knowledge on the queue length distribution of LRD traffic and the GPS scheduling discipline.

A. LRD Traffic and WB SSQ

LRD traffic is often characterized by heavy traffic bursts that extend over a wide range of time scales [21, 22]. The LRD traffic backlog, buffered within a SSQ, often possesses a tail distribution that decays slower than that of traditional (e.g., Poisson) traffic. More specifically, the queue length distribution of traditional traffic obeys a certain exponential form. For the case of LRD traffic, the Weibull distribution is *often* used to characterize the slower decaying SSQ distribution [5, 23, 24]. This relationship between a WB SSQ distribution and LRD traffic is usually based on empirical observations. The queue length distribution which is Weibull Bounded (WB), has been defined as follows [25]:

Definition 2.1. A stochastic SSQ process, denoted by $W^{SSQ,\gamma}(t)$, where γ is the service rate of the queue, is WB(C, η, υ) with parameters $C > 0$ (which denotes the asymptotic constant), $\eta > 0$ (which denotes the decay rate), and $0 < \upsilon \leq 1$ (which denotes the index parameter), if it satisfies

$$Pr\{W^{SSQ,\gamma}(t) > w\} < Ce^{-\eta w^{\upsilon}} \tag{1}$$

for all $w \geq 0$ and all $t \geq 0$.

Remark: The quantity $Pr\{W^{SSQ,\gamma}(t) > w\}$ represents the probability that the backlog of the SSQ with service rate γ will exceed a certain queue size w. In other words, $Pr\{W^{SSQ,\gamma}(t) > w\}$ represents the queue length distribution of the SSQ. In addition, the decay rate η increases with γ because when the service rate increases, the likelihood that the queue length exceeding w will decrease. Also, the index parameter υ can be further expressed in terms of the Hurst parameter H (which is commonly used to characterize the degree of long range dependence [5, 23, 24]), and more specifically, $\upsilon = 2(1-H)$, where $0.5 \leq H < 1$. A traffic process with $H = 0.5$ corresponds to conventional traffic with a queue length distribution that decays exponentially. A larger H, or a smaller υ, corresponds to heavier tailed LRD traffic.

B. GPS Fundamentals

Generalized Processor Sharing (GPS) is a scheduling discipline defined under the assumption that sources are described by fluid models [11]. Consider a GPS server with rate γ serving N sessions. Let each session i be assigned a weight parameter, which is a fixed real-valued positive number ϕ_i. The set $\{\phi_1, \phi_2, ..., \phi_N\}$ thus represents the GPS assignment. The N sessions share the server in the following way [26]:

1. It is work conserving, i.e., as long as there are packets backlogged in any of the GPS queues, the server is never idle.

2. The excess service rate, if any, is redistributed among the backlogged sessions in proportion to their weight parameters.

3. Let $S_k(s,t)$ denote the amount of traffic served in the time interval $[s,t]$ for session k. If session i is backlogged in the system during the entire interval $[s,t]$, i.e., there is always traffic queued for session i, then

$$\frac{S_i(s,t)}{S_j(s,t)} \geq \frac{\phi_i}{\phi_j}\ ,\ j = 1, 2, ..., N. \tag{2}$$

¿From (2), it is clear that when session i is backlogged, it is guaranteed a backlog clearing rate (or equivalently, a guaranteed service rate) of at least

$$\gamma_i = \frac{\phi_i}{\sum_{j=1}^{N} \phi_j}\gamma. \tag{3}$$

III. Analysis of LRD Traffic

While analysis on LRD traffic usually starts with its SSQ process, it is still important to formally characterize the arrival process because, as mentioned previously, the stability of the SSQ process depends on whether the arrival process is bounded. Stochastically Bounded (SB) SSQ and Stochastically Bounded

A. Single WBB Arrival Process and Single WB SSQ

In this subsection, we establish the relationship between a WB SSQ and a Weibull Bounded Burstiness (WBB) arrival process. We begin with our definition of the burstiness constraint qualifier to describe the arrival process of LRD traffic as follows:

Definition 3.1. A traffic arrival process $A(t)$ is WBB(ρ, C, μ, υ) with parameters ρ, C, μ and υ, if it satisfies

$$Pr\{\int_s^t A(u)du > \rho(t-s) + w\} < Ce^{-\mu w^{\upsilon}} \tag{4}$$

for all $w \geq 0$ and all $0 \leq s \leq t$. Similar to the notation in Definition 2.1, C denotes the asymptotic constant, υ is the index parameter. Here, μ is the decay rate[2] and ρ is the long term "upper rate" of the arrival process, which will be further elaborated in Lemma 3.1.

Remark: $\int_s^t A(u)du$ is the amount of arrival traffic accumulated in time interval $[s,t]$. In addition, the decay rate μ will increase with ρ, just as η will increase with γ in a WB SSQ process (see Definition 2.1). However, there is a subtle difference between ρ and the parameter γ. In the former case, ρ is applied continuously from s to t while in the later case, γ is only applied when the queue is not idle in the interval $[s,t]$.

Several lemmas and theorems, useful to the objectives of this paper, are now presented.

Lemma 3.1. An arrival process $A(t)$ that is WBB(ρ, C, μ, υ) possesses the property that its parameter ρ is always larger than or equal to its long term average rate

$$\rho \geq \lim_{t-s\to\infty} \frac{E[\int_s^t A(u)du]}{t-s}. \tag{5}$$

The proof of Lemma 3.1, which intuitively explains why ρ is called the "upper rate", is in Appendix A.

Remark: The long term upper rate ρ is useful for the purpose of bounding the entire ensemble of sample time observations that constitute the stochastic arrival process $A(t)$. In particular, let $A_n(t)$ be the nth sample observation of $A(t)$ in $[s,t]$, and let $\lambda_n = \lim_{t-s\to\infty} \frac{\int_s^t A(n,u)du}{t-s}$ be the corresponding average arrival rate for this sample. If we were to repeat the observation of $A(t)$ infinitely many times using different start times, so that n approaches infinity, then we would have a corresponding list of average arrival rates $\lambda_1, \lambda_2, ..., \lambda_{n\to\infty}$. This long term upper rate ρ ranges between the lower limit $E[\lambda_n]$ and the higher limit $\rho_{max} = \max[\lambda_1, \lambda_2, ..., \lambda_{n\to\infty}]$. For a conservative (loose)

[2]Not to be confused with symbols η, which denotes the decay rate of a WB SSQ process instead.

WBB bound on $A(t)$, one may set ρ to the higher limit ρ_{max}. However, notice that the long term upper rate, defined in (4), is applied continuously even if the arrival process is inactive. Therefore, a lower value of ρ where $\rho_{max} \geq \rho \geq E[\lambda_n]$ may suffice to produce a tighter WBB bound on $A(t)$. In summary, the use of the long term upper rate ρ in (4) is essential for a general stochastic process which may not be stationary (i.e., $\lambda_1 \neq \lambda_2 \neq ... \neq \lambda_{n\to\infty}$). However, in practical arrival processes, stationary is an implicit property for a flow that has some fixed arrival rate λ. This means that if this flow is presented to the queue at different start times, the same average rate λ applies. Hence, for the case of practical flows, $\lambda_1 = \lambda_2 = ... = \lambda_{n\to\infty} = \lambda$ and therefore $\rho = \lambda$. Although many of our later derivations following this definition are still based on ρ, readers should be aware that for practical considerations, ρ ought to be replaced by λ since practical arrival processes are by default implicitly stationary in property. In fact, in the consideration of the GPS and PGPS discipline in Sections IV, VI and VII, we consider λ instead of ρ. Finally, it is also noted that besides ρ, the WBB expression in (4) also contains other parameters like the decay rate μ, the index υ and the asymptotic constant C. These parameters can similarly be modified to obtain either loose or tight WBB bounds.

In the following theorem, the relationship between a WBB arrival process that we have defined, and a WB SSQ process is established.

Theorem 3.1. Consider a work conserving SSQ that transmits at rate γ. Suppose the queue is fed with a single arrival process $A(t)$, and let $W^{SSQ,\gamma}(t)$ be the amount of workload stored in the queue at time t. Then:

(i)If $W^{SSQ,\gamma}(t)$ is WB, then $A(t)$ is WBB, with long term upper rate $\rho = \gamma$.

(ii)If $A(t)$ is WBB with long term upper rate $\rho = \gamma - \varepsilon$ for some $\varepsilon > 0$, then $W^{SSQ,\gamma}(t)$ is WB.

Proof: For (i), since the workload $W^{SSQ,\gamma}(t)$ can be expressed as follows:

$$W^{SSQ,\gamma}(t) = \max_{s<t}\left(\int_s^t A(u)du - \gamma(t-s)\right) \tag{6}$$

where $0 \leq s \leq t$, we have:

$$W^{SSQ,\gamma}(t) \geq \int_s^t A(u)du - \gamma(t-s)$$

for all $0 \leq s \leq t$, therefore,

$$\begin{aligned} Pr\{\int_s^t A(u)du > \gamma(t-s) + w\} &\leq Pr\{W^{SSQ,\gamma}(t) > w\} \\ &< Ce^{-\eta w^{\upsilon}} \end{aligned}$$

Hence, $A(t)$ is a WBB(γ, C, η, υ) process.

To prove (ii) we begin with the condition that $A(t)$ is WBB with long term upper rate ρ, i.e.

$$Pr\{\int_s^t A(u)du > \rho(t-s) + w\} < Ce^{-\eta w^{\upsilon}} \tag{7}$$

we prove by demonstrating that the SSQ process of any given sample observation of $A(t)$ is WB. Now, from basic stochastic theory, we can apply the bound in (1) to a particular sample observation of $A(t)$, say $A_n(t), t \in [0, t_n]$, where 0 is the start time of the observation and t_n is the end time of observation. Note that the subscript n is not only used to represent a particular time sample but also the start time and end time of that sample observation. Thus,

$$\begin{aligned} &Pr\{\int_s^t A(u)du > \rho(t-s) + w\} < Ce^{-\eta w^{\upsilon}}, 0 \leq s \leq t \\ &\Rightarrow Pr\{\int_x^{t_n} A_n(u)du > \rho(t_n - x) + w\} < Ce^{-\eta w^{\upsilon}}, \\ &0 \leq x \leq t_n \end{aligned} \tag{8}$$

The relation in (8) only has a one way implication. However, if $A(t)$ is an ergodic process, then a double implication can be used in (8). It is also noted that the observation time of the sample process $A_n(t)$ ranges from 0 to t_n. This is a finite time range since by definition, an observation requires a finite start time and a finite end time. Let $W_n^{SSQ,\gamma}$ be the maximum queue size value arising out of the $A_n(t)$ arrival sample, such that

$$W_n^{SSQ,\gamma} = \max_{x\in[0,t_n]}\left(\int_x^{t_n} A_n(u)du - \gamma(t_n - x)\right) \tag{9}$$

Since $A_n(t)$ is a continuous time process between finite time limits 0 and t_n, there must exist a value x_n^* that maximizes (9), so that we can proceed on to:

$$W_n^{SSQ,\gamma} = \int_{x_n^*}^{t_n} A_n(u)du - \gamma(t_n - x_n^*) \text{ where } 0 \leq x_n^* \leq t_n \tag{10}$$

Therefore,

$$\begin{aligned} &Pr\{W_n^{SSQ,\gamma}(t) > w\} \\ &= Pr\{\int_{x_n^*}^{t_n} A_n(u)du > \gamma(t_n - x_n^*) + w\} \\ &= Pr\{\int_{x_n^*}^{t_n} A_n(u)du > \rho(t_n - x_n^*) + \varepsilon(t_n - x_n^*) + w\} \\ &< Ce^{-\eta[w+\varepsilon(t_n - x_n^*)]^{\upsilon}} \\ &< Ce^{-2^{-1}\eta w^{\upsilon}} e^{-2^{-1}\eta\varepsilon^{\upsilon}[t_n - x_n^*]^{\upsilon}} \text{(See (47) in Appendix)} \end{aligned} \tag{11}$$

It is noted that the above bound on $Pr\{W_n^{SSQ,\gamma}(t) > w\}$ depends on the observed $t_n - 0$ duration. If the duration were to change, say for example, in the end time t_n, this will affect the value x_n^* and hence the bound will change accordingly. However, a generic (but looser) bound that is independent of the observed duration can be obtained as follows:

$$\begin{aligned} Pr\{W_n^{SSQ,\gamma}(t) > w\} &< Ce^{-2^{-1}\eta w^{\upsilon}} e^{-2^{-1}\eta\varepsilon^{\upsilon}[t_n - x_n^*]^{\upsilon}} \\ &< Ce^{-2^{-1}\eta w^{\upsilon}} \end{aligned} \tag{12}$$

A closer look at the Weibull bounded expression $Ce^{-2^{-1}\eta w^{\upsilon}}$ in (12) also shows an absence of dependency on the sample index n. This means, that given any sample observation $A_n(t)$ of $A(t)$ of any duration, the resulting SSQ process is WB($C, \frac{\eta}{2}, \upsilon$), or in other words, the SSQ process of $A(t)$ is WB($C, \frac{\eta}{2}, \upsilon$) $\square$.

Remark: Although conventionally, LRD traffic is usually described in terms of some WB SSQ process, it is still insufficient to proceed on to GPS analysis since in GPS, we are concerned with multiple arrival processes rather than a single arrival process. If there is no burstiness constraint on a single arrival process, there is not much that can be deduced on the stability of a GPS server that is serving a number of these arrival processes. With the introduction of Theorem 3.1, we can now proceed further, since it is now known that any LRD arrival process resulting in a WB SSQ process must satisfy the WBB constraint with some long term upper rate ρ. This means that for a GPS server serving a number of LRD sources, as long as the sum of the long term upper rates of these LRD sources does not exceed the service capacity of the GPS server, the GPS queue will be stable and further analysis can proceed.

B. Bounds on Aggregate WB SSQ

In this subsection, several bounds on the aggregate WB SSQ process are derived. These bounds will be used frequently in Section IV, VI and VII where GPS and PGPS analysis are presented.

Lemma 3.2. Let $W_1(t)$ be WB(C_1, η_1, υ_1) and $W_2(t)$ be WB(C_2, η_2, υ_2). The two processes can either be dependent or independent. Then, $W_1(t)+W_2(t)$ is WB($C_1 + C_2 + C^*$), η, υ) satisfying

$$Pr\{W_1(t) + W_2(t) > w\} < (C_1 + C_2 + C^*)e^{-\eta w^{\upsilon}} \quad (13)$$

where $\eta = \frac{\eta_1\eta_2}{\eta_1+\eta_2}$, $\upsilon = min(\upsilon_1, \upsilon_2)$ and C^* is a constant obtained using the method in Appendix C.
The proof of Lemma 3.2 is in Appendix C.

Lemma 3.2 can be easily extended to multiple WB processes as follows.

Theorem 3.2. Let $W_i(t)$, $1 \le i \le N$ be N WB processes with parameters (C_i, η_i, υ_i) respectively. These processes can either be dependent or independent. Then, $W_1(t)+W_2(t)+...+W_N(t)$ is WB(($\sum_{i=1}^{N} C_i+C^*$), η, υ) satisfying

$$Pr\{\sum_{i=1}^{N} W_i(t) > w\} < (\sum_{i=1}^{N} C_i + C^*)e^{-\eta w^{\upsilon}} \quad (14)$$

where $\eta = \frac{1}{\sum_{i=1}^{N}\frac{1}{\eta_i}}$, $\upsilon = min(\upsilon_1, \upsilon_2, ..., \upsilon_N)$ and C^* can be obtained using the method Appendix C by applying Lemma 3.2 step by step.
Remark: Given Lemma 3.2, the proof of Theorem 3.2 is straight forward and hence omitted. For N WB queuing processes with the same LRD degree (i.e. the same υ), Appendix C shows that $\upsilon = min(\upsilon_1, \upsilon_2, ..., \upsilon_N)$ is the tightest lower bound on the index parameter. But for N WB queuing processes with different LRD degrees, it is a loose bound because the index parameter of multiplexed LRD flows is in general heavier tailed than the individual flows due to multiplexing gain.

The following Lemma 3.3 and Theorem 3.3 present alternate bounds to those bounds obtained in Lemma 3.2 and Theorem 3.2 respectively. The alternate bounds are useful since in certain cases they are tighter (see Fig. 4(b)). However, these alternate bounds are for large queue situations. For this, we now present Lemma 3.3 and Theorem 3.3 as follows.

Lemma 3.3. Let $W_1(t)$ and $W_2(t)$ be two independent WB(C_1, η_1, υ_1) and WB(C_2, η_2, υ_2) respectively. If $\eta_2 \le \eta_1$ and $\upsilon_2 \le \upsilon_1$, then for $\forall w > 2$, $W_1(t) + W_2(t)$ has an upper bound of the form

$$Pr\{W_1(t) + W_2(t) > w\} < C_2^{WB}(w)e^{-\eta w^{\upsilon}} \quad (15)$$

where $\eta = min\{\eta_1, \eta_2\} = \eta_2$, $\upsilon = min\{\upsilon_1, \upsilon_2\} = \upsilon_2$ and

$$C_2^{WB}(w) = C_2 h(C_1) + C_1$$

where $h(C_1) = 1 + C_1\upsilon\eta(e^{\eta} - 1) + C_1 w^{\upsilon}\eta$ The proof of Lemma 3.3. is in Appendix D.

Remark: From Appendix C, we know that η in Lemma 3.2 is $\frac{\eta_1\eta_2}{\eta_1+\eta_2}$, which is always less than or equal to $\min(\eta_1, \eta_2)$, which is the η in (15). Hence Lemma 3.3 yields a larger (thus better) decay rate. In fact, if $\eta_2 \approx \eta_1$, then η in Lemma 3.3 is almost twice the value of η in Lemma 3.2. However, the asymptotic constant in Lemma 3.3 increases with w, which is a trade off. For a heavy tailed arrival process where $\upsilon \to 0$ so that for practical and finite values of w, $w^{\upsilon} \to 1$ and thus $C_2^{WB}(w)$ approaches a constant, the penalty for using Lemma 3.3 is insignificant. Conversely, if η_2 differs significantly from η_1 (i.e., $\eta_2 \ll \eta_1$), then $\eta \to min(\eta_1, \eta_2)$, making Lemma 3.2 more attractive.

Table 1 summarizes the preferences (in terms of which Lemma to use to obtain the bound) assuming that in all the scenarios, the queue size of interest is larger than or equal to 2.

Scenario	Preference
$\eta_2 \approx \eta_1$ & υ_2 is small	Lemma 3.3
$\eta_2 \ll \eta_1$ & υ_2 is large	Lemma 3.2
All other cases	Either is ok

TABLE I
PREFERENCE FOR LEMMA 3.3 OR LEMMA 3.2 IN DIFFERENT SCENARIOS

Similarly to the way in which Lemma 3.2 is extended to Theorem 3.2, we now extend Lemma 3.3 to the following Theorem 3.3.

Theorem 3.3. Let $W_i(t)$, $1 \le i \le N$, be N independent WB processes with parameters (C_i, η_i, υ_i) respectively. If the queuing processes can be rearranged such that the Nth queuing process has the property that $\eta_N \le \eta_j$ and $\upsilon_N \le \upsilon_j$ for $1 \le j \le N-1$, then for $\forall w > 2$,

$W_1(t)+W_2(t)+...+W_N(t)$ has an upper bound of the form

$$Pr\{\sum_{i=1}^{N} W_i(t) > w\} < C_N^{WB}(w)e^{-\eta w^{\upsilon}} \tag{16}$$

where $\eta = \min\{\eta_1, \eta_2, ..., \eta_N\} = \eta_N$, $\upsilon = \min\{\upsilon_1, \upsilon_2, ..., \upsilon_N\} = \upsilon_N$ and

$$C_N^{WB}(w) = \sum_{j=1}^{N} [C_j \prod_{l=1}^{j-1} h(C_l)] \tag{17}$$

where $h(C_l) = 1+C_l\upsilon\eta(e^{\eta}-1)+C_l w^{\upsilon}\eta$, and by convention, $\prod_{l=1}^{j-1} h(C_l) = 1$ when $j = 1$.
Proof: see Appendix D, particularly its last paragraph.

IV. Sample Path Behavior of LRD Traffic in a GPS System

Recall from Theorem 3.1 that any LRD traffic input whose queue length distribution is characterized by a WB distribution has an arrival process that satisfies the WBB constraint with some long term upper rate ρ. Hereafter, we consider N stationary flows that maintain the same long term average rate λ_i, $i = 1, 2, ..., N$ irrespective of the start time of the flow. As mentioned earlier (in the remark for Lemma 3.1), the long term upper rate ρ reduces to the more familiar λ.

A. Feasible Ordering and LRD Isolation of Flows

Let the arrival process for a stationary WBB LRD session i be $A_i(t)$ with long term average rate λ_i, such that $\sum_{i=1}^{N} \lambda_i < \gamma$. In order to characterize the effect of backlogs from a set of sessions, we use the following definition of the so-called *feasible ordering* of the sessions that will be frequently referred to hereafter, according to their arrival rates and (GPS service) weight parameters.

Definition 4.1. For a given set of input traffic flows in a GPS server, whose long term average rate is λ_i, an ordering is called a feasible ordering among the sessions with respect to $\{\lambda_1, \lambda_2, ..., \lambda_N\}$ and (GPS service) weight parameters $\{\phi_1, \phi_2, ..., \phi_N\}$ [2] if and only if:

$$\lambda_i < \varphi_i(\gamma - \sum_{j=1}^{i-1} \lambda_j) \text{ , } 1 \le i \le N \tag{18}$$

where $\varphi_i = \frac{\phi_i}{\sum_{j=i}^{N} \phi_j}$, is a constant associated with weight parameters. And by convention, $\sum_{j=1}^{i-1} \lambda_j = 0$ when $i = 1$.

We have the following result for feasible ordering, which has been proved in [11]:

Lemma 4.1. For a given set of input traffic flows in a GPS server with $\sum_{i=1}^{N} \lambda_i < \gamma$, there always exists one or more than one feasible ordering that satisfies (18) after being relabelled.

Remark: The right-hand side of (18) can be considered as the service rate available to flow i. It is clear, by definition, that those flows ordered earlier than flow i will affect the service rate available to flow i. However, they will not affect the index parameter of the queuing process of a heavier tailed flow i as to be explained in subsection IV.C and IV.D in more detail.

Note that the index parameter is what differentiates a LRD flow from a SRD flow. Although the decay rate and constant parameters also define the queuing process, however, these parameters form the exponential bound parameters commonly associated with an SRD flow. Hence their presence, by definition, is for the purpose of describing the SRD property of the flow. The index parameter, found in the Weibull bound formula, was introduced to bound flows exhibiting LRD behavior which cannot be suitably bounded by just the constant parameter and the decay rate. Hence, the LRD property of a flow, by definition, is primarily due to its index parameter. Accordingly, we introduce the following notion of "LRD isolation".

Definition 4.2. We say that a flow, when multiplexed with other flows in a queue system, is *LRD isolated* (from other flows) in that queue system if and only if its resulting queue process has the same or larger index parameter (i.e. less heavy tailed) as the index parameter associated with its SSQ process with equivalent service rate as that guaranteed in the queue system.

This notion of "LRD isolation" is different from the conventional understanding of flow isolation. In flow isolation, the major concern is the flow's service rate, and a flow is said to be isolated from other flows if this flow is not adversely affected by these flows [27]. Based on this, a LRD flow is flow isolated if and only if its queue process is not adversely affected after it is multiplexed and served with other flows in the GPS server.

It can be shown that flow isolation is guaranteed for a flow in a GPS server if the flow can be ordered first in a feasible ordering. The reason is under this case, the flow is always guaranteed a service rate greater than its long term average rate based on (18), which is not affected by other flows. In addition, Lemma 4.2, which will be presented later, also shows that the flow's queue process in the GPS system is not adversely affected (with respect to its SSQ process) by other flows. However, if the flow cannot be ordered first in any of the feasible ordering, the guaranteed service rate to the flow may depend on the arrival rates of some other flows. In other words, it may vary over time and hence the queue process of the flow in the GPS system could be affected adversely.[3] As a result, if a flow cannot be ordered first in any of the feasible ordering, the flow may or may not be guaranteed to be flow isolated from other flows. However, a flow can still be guaranteed to be LRD isolated (from heavier tailed flows) even if some lighter tailed flows have to be ordered before this flow in *all* feasible orderings, as to be discussed later in this section. Clearly, flow isolation implies LRD isolation but not vice versa. Since the index parameter is the most important measure of the LRD property (heaviness or lightness of the tail) of a flow, the notion of LRD isolation as defined above is useful when studying LRD flows.

[3]Note that, when $\lambda_i = \phi_i\gamma$, flow i cannot be ordered first according to (18) although it is flow isolated.

B. GPS Decomposition

Now let A_i denote a sample path (or a single realization) of the random arrival process $A_i(t)$, and $Q_i^{GPS,\gamma}$ denote the corresponding sample path of the GPS queue backlog due to the sample arrival process A_i. To obtain relevant bounds on $Q_i^{GPS,\gamma}$, we use a method similar to that in [11] to decompose the GPS system into N fictitious WB single server queues (SSQs), denoted by $\delta_i^{SSQ,\gamma_i}(t)$, with individual rates $\gamma_1, \gamma_2, ..., \gamma_N$, where $\gamma_i > \lambda_i$, $\sum_{i=1}^{N}\gamma_i \leq \gamma$, and $\gamma_i \leq \varphi_i(\gamma - \sum_{j=1}^{i-1}\gamma_j)$. Now, the reason for considering the N fictitious WB SSQs is that their bounds are easier to obtain and would surely bound $Q_i^{GPS,\gamma}$ as well. This is because the N fictitious WB SSQs do not consider multiplexing gain while the $Q_i^{GPS,\gamma}$ queue process does.

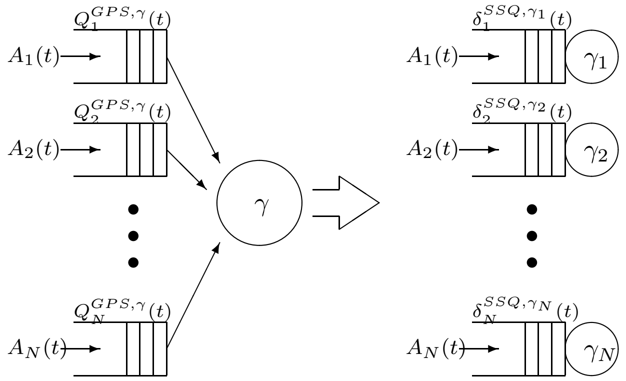


Fig. 1. Decompose a GPS system into N fictitious SSQs

Without loss of generality, let 1,2,...,N be a feasible ordering of the fictitious processes with respect to γ_i's. From Lemma 3 of [11], the following Lemma 4.2 can be derived:

Lemma 4.2. For any t,

$$Q_i^{GPS,\gamma}(t) \leq \varphi_i \sum_{j=1}^{i-1} \delta_j^{SSQ,\gamma_j}(t) + \delta_i^{SSQ,\gamma_i}(t) \tag{19}$$

where each δ_i^{SSQ,γ_i} SSQ process is independent.

Remark: Lemma 4.2 provides an upper bound on the queue length $Q_i^{GPS,\gamma}(t)$ of an individual session in the GPS system in terms of the queue length $\delta_i^{SSQ,\gamma}(t)$ in the fictitious system. It is clear from Lemma 4.2 that to bound the distribution of $Q_i^{GPS,\gamma}(t)$, it suffices to bound the following aggregate of fictitious queue length processes:

$$\varphi_i\delta_1^{SSQ,\gamma_1}(t)+\varphi_i\delta_2^{SSQ,\gamma_2}(t)+...+\varphi_i\delta_{i-1}^{SSQ,\gamma_{i-1}}(t)+\delta_i^{SSQ,\gamma_i}(t) \tag{20}$$

In what follows, we will provide two bounds on (20), i.e., the right hand side of (19).

C. A General Bound on Individual Session Queue Length

For N individual LRD flows sharing a GPS server on the condition of queue stability, i.e., $\sum_{i=1}^{N}\lambda_i < \gamma$, and under the assumption that 1,2,...N is a feasible ordering with respect to ϕ_i and λ_i, $\lambda_i < \gamma_i$ for $i = 1, 2, ..., N$, we present a GPS bound in the following Theorem 4.1 that is based on Theorem 3.2.

Theorem 4.1. Each individual queue length distribution in the GPS system has an upper bound as follows:

$$Pr\{Q_i^{GPS,\gamma}(t) > q\} < C_i^{GPS} e^{-\eta_i^{GPS} q^{\upsilon_i^{GPS}}} \tag{21}$$

where

$$\upsilon_i^{GPS} = \min_{1\leq j\leq i}\{\upsilon_j\}\ , \tag{22}$$

$$\eta_i^{GPS} = \frac{1}{\sum_{j=1}^{i}\frac{1}{\bar{\eta}_j}}\ , \tag{23}$$

$$C_i^{GPS} = (\sum_{j=1}^{i} C_j + C^*)e^{-\eta_i^{GPS}} \tag{24}$$

Note that, in the above,

$$\bar{\eta}_j = \begin{cases} \frac{\eta_j}{\varphi_i^{\upsilon_j}} & 1 \leq j < i \\ \eta_i & j = i \end{cases}$$

and C^* can be obtained similarly as in Theorem 3.2.

Proof: First, because all the input flows are LRD flows, we have

$$Pr\{\delta_j^{SSQ,\gamma}(t) > q\} < C_j e^{-\eta_j q^{\upsilon_j}}\ ,\ j = 1, 2, ..., N. \tag{25}$$

Secondly, let

$$\delta_{j,eqv}^{SSQ,\gamma_j}(t) = \varphi_i \delta_j^{SSQ,\gamma_j}(t)\ ,\ j < i \tag{26}$$

we have

$$Pr\{\delta_{j,eqv}^{SSQ,\gamma_j}(t) > q\} = Pr\{\delta_j^{SSQ,\gamma_j}(t) > \frac{q}{\varphi_i}\} < C_j e^{-\bar{\eta}_j q^{\upsilon_j}} \tag{27}$$

for $1 \leq j < i$, where $\bar{\eta}_j = \frac{\eta_j}{\varphi_i^{\upsilon_j}}$.

Finally, since (20) can now be written as

$$\delta_{1,eqv}^{SSQ,\gamma_1}(t)+\delta_{2,eqv}^{SSQ,\gamma_2}(t)+...+\delta_{i-1,eqv}^{SSQ,\gamma_{i-1}}(t)+\delta_i^{SSQ,\gamma_i}(t) \tag{28}$$

by combining (25), (27) and Lemma 4.2, one can easily verify the result in (21) based on Theorem 3.2. □

Remark: Theorem 4.1 gives a general upper bound on queue length distribution in a GPS system. It is important to note that the GPS upper bound on flow i is not affected by the flows that are ordered after flow i (because they do not factor in the upper bound expression for flow i). It is affected only by flows 1 to $i-1$, but the impact on the bound is *negligible* as long as flow i is heavier tailed (i.e., has a smaller index parameter υ_i) than any of the flows 1 through $i-1$. In fact, the index parameter in the bound for flow i is not affected at all as long as flows 1 to $i-1$ are lighter tailed than flow i.

D. An Alternate Upper Bound

In the following Theorem 4.2, an alternate upper bound on individual session queue length in GPS with LRD traffic is provided based on Theorem 3.3. Such a bound may be better than the bound previously given in Theorem 4.1 but can only be applied under the condition that for any given i, there exists a $1 \leq k \leq i$ such that both $\bar{\eta}_k$ and υ_k are minimal, in addition to the conditions stated for Theorem 4.1.

Theorem 4.2. If there exists a k, $1 \leq k \leq i$, such that $\bar{\eta}_k = \min_{1\leq j\leq i}\{\bar{\eta}_j\}$ and $\upsilon_k = \min_{1\leq j\leq i}\{\upsilon_j\}$, then for $\forall q > 2$, the upper bound for individual session queue length is:

$$Pr\{Q_i^{GPS,\gamma}(t) > q\} < C_i^{GPS}(q) e^{\eta_i^{GPS} q^{\upsilon_i^{GPS}}} \tag{29}$$

where

$$\upsilon_i^{GPS} = \min_{1\leq j\leq i}(\upsilon_j) = \upsilon_k \ , \tag{30}$$

$$\eta_i^{GPS} = \min_{1\leq j\leq i}(\bar{\eta}_j) = \bar{\eta}_k \ , \tag{31}$$

$$\begin{aligned} C_i^{GPS}(q) \ &= \ C_k \prod_{l=1,l\neq k}^{i} h_i^{GPS}(C_l) + \\ & \sum_{j=1,j\neq k}^{i} [C_j \prod_{l=1,l\neq k}^{j-1} h_i^{GPS}(C_l)] \end{aligned} \tag{32}$$

where $h_i^{GPS}(C_l) = 1 + C_l \upsilon_i^{GPS} \eta_i^{GPS} (e^{\eta_i^{GPS}} - 1) + C_l w^{\upsilon_i^{GPS}} \eta_i^{GPS}$, and by convention, $\prod_{l=1,l\neq k}^{j-1} h_i^{GPS}(C_l) = 1$ when $j = 1$.

Proof: Without loss of generality, assume that $k < i$. The aggregate process in (28) can be re-written such that the kth process with the minimum decay rate as well as with the minimum index parameter appears last in the sequence as follows:

$$\begin{aligned} &\delta_{1,eqv}^{SSQ,\gamma_1}(t) + \delta_{2,eqv}^{SSQ,\gamma_2}(t) + ... + \delta_{k-1,eqv}^{SSQ,\gamma_{k-1}}(t) \\ &+ \delta_{k+1,eqv}^{SSQ,\gamma_{k+1}}(t) + ... + \delta_{i}^{SSQ,\gamma_i}(t) + \delta_{k,eqv}^{SSQ,\gamma_k}(t) \end{aligned} \tag{33}$$

Hence, by applying Theorem 3.3, Theorem 4.2 can be easily verified. □

Remark: Theorem 4.2 provides an upper bound on an actual session i's backlog $Q_i^{GPS,\gamma}(t)$ in the GPS system when there exists a very heavy tailed LRD flow with the smallest index parameter (as well as the smallest decay rate). One implication of Theorem 4.2, similar to Theorem 4.1, is that it is desirable to order the flows that are heavier tailed as close to the end of a feasible ordering as possible, again since the index parameter in the upper bound for the individual queue length of flow i will not be affected if and only if the flows 1 through $i-1$ are all lighter-tailed than flow i.

V. A Technique to Check and Ensure LRD Isolation

Recall from Lemma 4.1 that in a stable GPS system where $\sum_{j=1}^{N} \lambda_j < \gamma$, there exists at least one feasible orderings for a given weight assignment. Before we discuss LRD isolation, it is useful to revisit the concept of flow isolation with the following Lemma 5.1.

Lemma 5.1: In a stable GPS system, if flow i satisfies the following condition:

$$\lambda_i < \gamma \frac{\phi_i}{\sum_{j=1}^{N} \phi_j} \tag{34}$$

then the flow is flow isolated.

Proof: The proof to Lemma 5.1 is straight-forward since the right hand side of (34) is the minimum guaranteed rate, however, we still provide the required proof for Lemma 5.1 since several intermediate results of this proof will be used later to prove newer results pertaining to LRD isolation.

Relabel flow i as flow 1, and all other $N-1$ flows to be flows 2 to N. Note that flow 1 now satisfies (18), and hence all we need to show is that the remaining $N-1$ flows can be feasibly ordered after flow 1. To this end, consider a new GPS system with service rate $\gamma' = \gamma - \lambda_1$. Since $\gamma' > \sum_{j=2}^{N} \lambda_j$, the new GPS system is also stable, and hence, there always exists a feasible ordering such that (after relabelling the flows 2 to N, we have for any flow $2 \leq i \leq N$:

$$\begin{aligned} \lambda_i \ &< \ \frac{\phi_i}{\sum_{j=i}^{N} \phi_j} (\gamma' - \sum_{j=2}^{i-1} \lambda_j) \\ &= \ \frac{\phi_i}{\sum_{j=i}^{N} \phi_j} (\gamma - \sum_{j=1}^{i-1} \lambda_j). \square \end{aligned} \tag{35}$$

Note that the above becomes the same as (18), which means that if flow 1 is ordered first, the remaining $N-1$ flows can also be ordered to yield a feasible ordering. Therefore, flow 1 is flow isolated.

Remark: It should be noted that (34) is only a sufficient condition for flow isolation, not a necessary condition. In fact, it is a sufficient condition to guarantee a flow to be flow isolated. However, as mentioned earlier, a flow can still be isolated even if it cannot be "guaranteed" to be flow isolated, or even if it does not satisfy (34).

Based on Lemma 5.1, an obvious method to guarantee the flow isolation of every flow is to assign weight of every flow according to (34) such that every flow i can be ordered in the first place in a feasible ordering. As mentioned earlier, being able to order a flow first in any feasible ordering is the most applicable condition to guarantee a flow to be flow isolated for CAC purposes. That, however, is not necessary to guarantee just LRD isolation of a flow, which is less strict than flow isolation, as to be discussed in a later subsection.

A. Limitations of Existing Methods

In this subsection, we discuss the shortcomings of the existing methods for assigning weights to achieve flow isolation and testing whether a flow can be flow isolated for a given weight assignment.

A GPS system may support the following three types (classes) of flows (services): A Type 1 flow requires a higher QoS than that provided by flow isolation, so it requires a contract weight that is much larger than $\frac{\lambda_i}{\gamma}\sum_{j=1}^{N}\phi_j$; A Type 2 flow requires flow isolation, and thus needs a contract weight that is a little larger than $\frac{\lambda_i}{\gamma}\sum_{j=1}^{N}\phi_j$; A Type 3 flow only requires LRD isolation (but not flow isolation), and thus can have a contract weight less than $\frac{\lambda_i}{\gamma}\sum_{j=1}^{N}\phi_j$. Note that the contract weight cannot be changed as long as the service level agreement (SLA) is in effect. On the other hand, a (lightly loaded) GPS system may assign a flow an *extra* weight (if available) to provide the flow with better service, and such extra weight can be adjusted (e.g., transferred to other flows) by the GPS system.

The method of assigning weights based on (34) has a limited applicability in supporting both Types 1 and 2 but is not applicable to Type 3 flows. From users or applications' viewpoint, having Type 3 flows is useful because certain applications may require less strict performance guarantee than that given by flow isolation, and such flows can be admitted into a GPS system and with less costs to the users or aplications. In addition, from the GPS system's viewpoint, supporting Type 3 flows allows it to admit more flows than otherwise possible, thus increasing its utilization and potential revenues.

For example, (hereafter referred to as Example 1), consider a GPS system with $\gamma = 16$ and five flows numbered 1 through 5 in the descending order of their index parameters, whose $\lambda_i = i$ where $1 \le i \le 5$. Assume that the total weight is $\sum_{j=1}^{5}\phi_j = 16$, and in addition flows 1 and 2 have been assigned contract weights of $\phi_1 = 1.1$ and $\phi_2 = 4$, respectively. Since the remaining weight for flows 3, 4 and 5 is 10.9 but the sum of their arrival rates is 12, it is clear that the (34) cannot be used to assign the weights to all the three remaining flows to guarantee their flow isolation.

In general, due to the existence of Type 1 flows (e.g., flow 2 in Example 1), flow isolation may not always be achievable by every flow, and accordingly, the existing approach based on (34) may not be useful. Note that, even if flow i does not satisfy (34), it may still be LRD isolated. In the above Example 1, one can assign 2.6 to flow 3 to ensure its LRD isolation (which can be verified using the technique to be proposed later), even though such a weight violates (34).

As another example (hereafter called Example 2) showing the deficiency of the existing approaches, assume that the weight assignment for the same five flows as in Example 1 is now $\{1.1, 2.1, 1, 4, 7.8\}$. It is clear that (34) cannot be used to test if flows 3 and 4 (both violate (34)) are LRD isolated or not. *In addition*, (18) in Definition 4.1 is not effective either. More specifically, in order to use it to test whether flow 4 can be guaranteed to be LRD isolated or not, a naive approach will test if the ordering of 1,2,3,4,5 is feasible, and because it is not, it will have to examine the ordering of 1,3,2,4,5 and then the ordering of 2,3,1,4,5 and so on. In the worst case, to test if flow i can be guaranteed to be LRD isolated or not, all possible orderings involving j lighter tailed flows, where $0 \le j \le (i-1)$, have to be tested. Thus, the (worse case) time complexity of the testing process is $O(i!)$. When the number of flows is large, such an approach is clearly infeasible.

B. A Necessary and Sufficient Condition for the Guarantee of LRD Isolation

We now determine, for a given flow i, not only the set of lighter-tailed flows, denoted by f_i, that can be ordered before flow i in a feasible ordering, but also the minimum contract weight to ensure the LRD isolation of flow i. To this end, we first initialize f_i to be empty. Then, if there exists a flow k, where $1 \le k < i$, which satisfies

$$\frac{\lambda_k}{\phi_k} < \frac{\gamma - \sum_{j\in f_i}\lambda_j}{\sum_{j=1}^{N}\phi_j - \sum_{j\in f_i}\phi_j} \tag{36}$$

we add flow k to f_i, and update the right-hand side of (36) which will be denoted by $R(f_i)$. We repeat the above process until no such flow k exists, and denote the resulting set by F_i, and accordingly, the final value of $R(f_i)$ by $R(F_i)$. Note that this process of obtaining F_i has the worst-case time complexity of $O(i^2)$.

One can easily verify that when a flow k that satisfies (36) is added to f_i, the resulting $R(f_i)$ increases. i.e. $R(f_i) < R(f_i \cup k) \le R(F_i)$ if $f_i \subseteq F_i$. Conversely, if we were to add a flow k' that does not satisfy (36) to f_i, then $R(f_i \cup k') \le R(f_i)$. In other words, $R(F_i)$ is the maximum value that flow i can obtain from all flows that are lighter-tailed than flow i. This observation is important for proving the following theorem which provides both a necessary and sufficient condition for the LRD isolation guarantee of flow i.

Theorem 5.1: Suppose there are N flows in a GPS system which are numbered in the descending order of their index parameters as $1, 2, ..., N$, and their contract weights are ϕ_1, ϕ_2, ..., ϕ_N, respectively. Then flow i is guaranteed to be LRD isolated from other flows if and only if:

$$\frac{\lambda_i}{\phi_i} < \frac{\gamma - \sum_{j\in F_i}\lambda_j}{\sum_{j=1}^{N}\phi_j - \sum_{j\in F_i}\phi_j} = R(F_i) \tag{37}$$

Proof: (i) To show that (37) is a sufficient condition, we note that flow i also satisfies (36), just as any flow $k < i$ in F_i does. Accordingly, if we let $F_i' = F_i \cup \{i\}$, and note that when f_i is empty, $R(f_i) = \frac{\gamma}{\sum_{j=1}^{N}\phi_j}$, we have the following (based on the observation drawn preceding the Theorem):

$$\frac{\gamma - \sum_{j\in F_i'}\lambda_j}{\sum_{j=1}^{N}\phi_j - \sum_{j\in F_i'}\phi_j} > \frac{\gamma - \sum_{j\in F_i}\lambda_j}{\sum_{j=1}^{N}\phi_j - \sum_{j\in F_i}\phi_j} > \frac{\gamma}{\sum_{j=1}^{N}\phi_j}$$

Accordingly, we can easily conclude that

$$\frac{\sum_{j\in F_i'}\lambda_j}{\sum_{j\in F_i'}\phi_j} < \frac{\gamma}{\sum_{j=1}^{N}\phi_j}$$

The above means that if we treat the flows in F_i' as one *big* flow with arrival rate $\sum_{j\in F_i'}\lambda_j$ and weight $\sum_{j\in F_i'}\phi$, it satisfies (34). Hence, according to Lemma 5.1, there exists a feasible ordering with this *big* flow ordered first. In other words, flow i can be feasibly ordered before any heavier tailed flow. Note that the exact ordering of the flows within F_i will not affect the LRD isolation of flow i. In fact, the flows in F_i can be feasibly ordered according to the order they are added to F_i in (36) with flow i being ordered right after them.

(ii) we now prove that (37) is necessary by contradiction. Suppose (37) does not hold for flow i, but there still exists a feasible ordering with flow i ordered before any heavier tailed flows. Denote the set of all the (lighter tailed) flows that are feasibly ordered before flow i by F_i^* (which may be empty). According to (18), we should have:

$$\frac{\lambda_i}{\phi_i} < \frac{\gamma - \sum_{j\in F_i^*}\lambda_j}{\sum_{j=1}^{N}\phi_j - \sum_{j\in F_i^*}\phi_j} = R(F_i^*)$$

However, since F_i^* contains zero or more flows in F_i, and zero or more flows not in F_i, we have $R(F_i^*) \leq R(F_i)$ based on the discussion preceding the Theorem. Or in other words,

$$\frac{\lambda_i}{\phi_i} < R(F_i^*) \leq R(F_i)$$

which contradicts with the assumption that (37) does not hold for flow i. □

Remark: If a flow satisfies (34), it will satisfy (36) but not vice versa. With (37), whether a flow is guaranteed to be LRD isolated or not depends only on the weights assigned to the flows in F_i, and flow i itself. In previous Example 1, one can easily verify that $F_3 = \{1, 2\}$, and $R(F_3) = (16 - 3)/(16 - 5.1) = 1.19$. Hence, if $\phi_3 = 2.6$, flow 3 satisfies (37) and thus is guaranteed to be LRD isolated. On the other hand, in previous Example 2 (where the weight assignment for five flows is $\{1.1, 2.1, 1, 4, 7.8\}$), one can easily verify that since $F_3 = F_4 = \{1, 2\}$, and $R(F_3) = R(F_4) = 13/12.8$, (37) cannot be satisfied by flow 3, and thus flow 3 is not guaranteed to be LRD isolated. On the other hand, flow 4 satisfies (37), and thus is guaranteed to be LRD isolated.

C. Weight Adjustment and Assignment to Ensure LRD Isolation

Theorem 5.1 is also useful for weight assignment and adjustment to guarantee flow's LRD isolation. More specifically, the observation drawn preceding the theorem, i.e., $R(F_i)$ is maximum with respect to flow i, serves as the base for determining a minimal ϕ_i to guarantee the LRD isolation of flow i.

For instance, consider again previous Example 2 but now assume that only the weights assigned to flows 1, 2 and 4 are contract weights (i.e., non-adjustable). If we want to ensure LRD isolation of flow 3, we must increase ϕ_3 to above 13/12.8. Such an increase can be accomplished if ϕ_5 has an *extra* weight of 2 that can be transferred to flow 3 (and as a result, ϕ_5 is reduced to 5.8 from 7.8).

The above technique to adjust the weight of a single flow to ensure its LRD isolation can certainly be extended to ensure LRD isolation of more than one flows provided that there are extra weights in the GPS system that can be adjusted/transferred. As a slightly different example from those above (Example 3), consider five flows numbered in the descending order of their index parameters whose arrival rates are more or less randomly distributed as $\{2, 4, 5, 1, 3\}$. Suppose that $\gamma = 17$ (which is sufficient to make the system stable) and the total weight is a constant 17. In addition, suppose that flow 2 (which is a Type 1 flow) has been assigned a contract weight of 7 (and thus the method based on (34) cannot be used for weight assignment to guarantee flow isolation of all the other flows as discussed earlier). If all other four flows are Type 3 flows that only require LRD isolation, we can use Theorem 5.1 to assign contract weights to them to guarantee their LRD isolation as follows (note that one can easily verify that flow 2 can be ordered first in any feasible ordering so it is already flow isolated).

For the first flow, from Theorem 5.1, we need to have $\phi_1 > \lambda_1 = 2$, so we set $\phi_1 = 2.1$ (theoretically speaking, we can set $\phi_1 = 2 + \epsilon$ where $\epsilon > 0$ can have a very small value). For flow 3, we first obtain $F_3 = \{1, 2\}$, and then from (37), we have

$$\begin{aligned}\phi_3 &> \lambda_3 \frac{\sum_1^5 \phi - \phi_1 - \phi_2}{\gamma - \lambda_1 - \lambda_2} \\ &= 5\frac{17 - 2.1 - 7}{17 - 2 - 4} = 3.59\end{aligned}$$

Accordingly, we set $\phi_3 = 3.6$ to flow 3. Similarly, we set $\phi_4 = 0.72$ and $\phi_5 = 0$[4]. The extra weight available in the system is $17 - 2.1 - 7 - 3.6 - 0.72 = 1.58$, which may be distributed among the five flows in an arbitrary manner.

To further illustrate the usefulness of the proposed technique, let us consider the following Corollary of Theorem 5.1 which is applicable to the case of on-line CAC.

Corollary 5.1: If a flow i is provided a contract weight ϕ_i which guarantees it to be either flow isolated or just LRD isolated, it will be guaranteed to be flow isolated or just LRD isolated after a new flow j is admitted as long as the system remains stable (note that flow j cannot take away any existing contract weights already assigned to the other flows so its contract weight can only come from the extra weight available in the system before it is admitted).

Proof: If flow i was guaranteed to be flow isolated before flow j is admitted, flow i must satisfy (34). Hence, flow i is guaranteed to be flow isolated after flow j is admitted.

Now assume flow i was only guaranteed to be LRD isolated but not guaranteed to be flow isolated before flow j is admitted. Under this assumption, there existed a feasible ordering in which the set of flows ordered in front of flow i, F_i, are all lighter-tailed than flow i. Now treat $F_i \cup \{i\}$ as one big flow F_i'. Just as shown in the Part (i) of the proof for Theorem 5.1, this big flow F_i' satisfies (34) and thus can be ordered in the first place in a feasible ordering. Thus, flow i is still guaranteed to be LRD isolated. □

[4] Note that with $\phi_5 = 0$, flow 5 gets best effort service

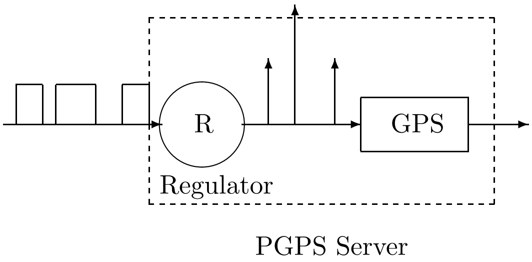


Fig. 2. PGPS Server

Let us continue the above Example 3 by assuming the online CAC receives a request for a new flow (flow 6). Suppose that its arrival rate is λ_6=1, and its index parameter is in between those of flows 2 and 3. To ensure its LRD isolation, we first obtain $F_6 = \{1, 2\}$ and then conclude we need a contract $\phi_6 > 0.718$. Since we have an extra weight of 1.58, we can assign $\phi_6 = 0.72$ and re-distribute the remaining extra weight $1.58 - 0.72 = 0.86$ among all six flows.

Note that from Corollary 5.1, admitting flow 6 as done in the above case will not affect either the flow isolation or LRD isolation of any flows, or in other words, their guaranteed (contracted) performance. There are, however, cases where a heavy tailed flow has been assigned a weight more than its arrival rate, and hence, the remaining weight is not enough to ensure the LRD isolation of the newly arrived flow. An example is that for the same set of 5 flows described in Example 3, but this time flow 4, instead of flow 2, is a Type 1 flow that requires a contract weight of 5. To ensure every other four flows are LRD isolated, we need the following weight assignment: $\{2.1, 4.1, 5.1, 5, 0\}$, which leaves an extra weight of only 0.7. Hence, when flow 6 arrives, it needs $\phi_6 > 1$ to ensure its LRD isolation. In such a case, the system may decide not to admit flow 6 or admit it without ensuring its LRD isolation.

VI. Sample Path Behavior of LRD Traffic in a PGPS System

The results obtained so for the GPS system are now extended to the PGPS system. While the GPS discipline assumes that the input traffic behaves like a fluid such that multiple sessions can be served bit by bit, the Packet based GPS (PGPS) is a more practical discipline in that only one packet at a time may be served. In other words, a PGPS server considers the arrival of a packet only after its last bit has been received. To manage this difference, the PGPS server is often taken to consist of two parts, a regulator and a PGPS core which is a GPS scheduler (see Chapter 4 in [28]), as illustrated in Figure 2. Partially-complete (or partially arrived) packets are queued in the regulator which passes only complete (or arrived) packets to the PGPS core. The output of this regulator, which is the input to the PGPS core, is a series of impulses, whose heights represent the sizes of the packets.

Let A_i be the session i input traffic to the PGPS server, which is also the input to the regulator, $A_{i,reg}$ be the output traffic from the regulator, which is the input traffic to the PGPS core, and finally $A(s,t)$ be the total amount of traffic arrived in time interval $[s,t]$. From Corollary 1 in [2], the queuing process of $A_{i,reg}(s,t)$ is also bounded by the queuing process of $A_i(s,t)$ with an extra length L, i.e., $Q_i^{PGPS}(s,t) \leq Q_i^{GPS}(s,t) + L$, where L is the maximum length of all arrived packets.. From the queuing process Q_i of A_i, which is WB(C, η, υ), we obtain the queuing process Q_i^{PGPS} of $A_{i,reg}$ as follows:

$$\begin{aligned} Pr\{Q_i^{PGPS}(s,t) > q\} &\leq Pr\{Q_i^{GPS}(s,t) + L > q\} \\ &= Pr\{Q_i^{GPS}(s,t) > q - L\} \\ &< C_i e^{-\eta_i (q-L)^{\upsilon_i}} \\ &\leq^5 C_i e^{\eta_i L^{\upsilon_i}} e^{-\eta_i q^{\upsilon_i}} \end{aligned} \quad (38)$$

which is WB$(Ce^{\eta L^{\upsilon_i}}, \eta, \upsilon)$. In other words, the two GPS upper bounds derived in Theorems 4.1 and 4.2 in the previous section can be extended to the PGPS domain via a simple transformation of the asymptotic constant $C_i \to C_i e^{\eta_i L^{\upsilon_i}}$ provided that the queue length or backlog is large enough to exceed the maximum packet length L, i.e., $q > L$. Note that this assumption ($q > L$) is reasonable because in practice, the buffer size B is much larger than L, i.e. $B \gg L$, and in addition, since the main concern is whether the backlog is about to exceed B, the values of q that are of interest should be close to B and thus is larger than L.

For completeness, we now present Theorem 6.1 and Theorem 6.2 which are derived from Theorem 4.1 and Theorem 4.2 respectively via the use of the simple transformation $C_i \to C_i e^{\eta_i L^{\upsilon_i}}$ as follows:

Theorem 6.1: Let $Q_i^{PGPS,\gamma}$, $1 \leq i \leq N$ represent the *ith* queuing process of the PGPS system with N LRD arrival processes, then at any time t, for any queue length $q > L$ where L is the maximum packet length of all the N sessions, we have:

$$Pr\{Q_i^{PGPS,\gamma} > q\} < C_i^{PGPS} e^{-\eta_i^{GPS} q^{\upsilon_i^{GPS}}} \quad (39)$$

where υ_i^{GPS}, η_i^{GPS} have already been defined in (22) and (23), and

$$C_i^{PGPS} = \left(\sum_{j=1}^{i} C_j e^{\bar{\eta}_j L^{\upsilon_j}} + C^*\right)$$

where C^* can be obtained similarly as in Theorem 3.2.

An alternate bound is given below, which, like the bound in Theorem 4.1, applies only when there exists a $1 \leq k \leq i$ for any given i, such that both $\bar{\eta}_k$ and υ_k are minimal.

Theorem 6.2: Under the same assumptions used for Theorem 4.2 except that the server is now a PGPS server, at any time t, for any $q > L > 2$ where L is the maximum packet length of all the N sessions:

$$Pr\{Q_i^{PGPS,\gamma} > q\} < C_i^{PGPS}(q) e^{-\eta_i^{GPS} q^{\upsilon_i^{GPS}}} \quad (40)$$

where υ_i^{GPS}, η_i^{GPS} have already been defined in (30) and

[5] see (48) in the Appendix.

(31), and

$$C_i^{PGPS}(q) = C_k e^{\bar{\eta}_k L^{\upsilon_k}} \prod_{l=1, l\neq k}^{i} h_i^{GPS}(C_l e^{\eta_l L^{\upsilon_l}}) + \sum_{j=1, j\neq k}^{i} [C_j e^{\bar{\eta}_j L^{\upsilon_j}} \prod_{l=1, l\neq k}^{j-1} h_i^{GPS}(C_l e^{\eta_l L^{\upsilon_l}})]$$

where

$$h_i^{GPS}(C_l e^{\eta_l L^{\upsilon_l}}) = (1 + C_l e^{\bar{\eta}_l L^{\upsilon_l}} \upsilon_i^{GPS} \eta_i^{GPS} (e^{\eta_i^{GPS}} - 1) + C_l e^{\bar{\eta}_l L^{\upsilon_l}} q^{\upsilon_i^{GPS}} \eta_i^{GPS})$$

and by convention, $\prod_{l=1,l\neq k}^{j-1} h_i^{GPS}(C_l e^{\eta_l L^{\upsilon_l}}) = 1$ when $j = 1$.

Remark: Note that the above bounds shed light on the LRD isolation among LRD sources sharing a PGPS server. To illustrate this, consider a simple case of two independent LRD sources with a feasible ordering of 1, 2. From Theorem 4.1, the source which appears first in the feasible ordering is always guaranteed to be LRD isolated. Therefore, the queuing process that is of interest is the last queuing process in the feasible ordering, i.e., $Q_2^{PGPS,\gamma}$. By applying Theorems 6.1 and 6.2, three possible sets of bounds can be obtained as follows:

(i)If $\eta_1 \le \eta_2$ and $\upsilon_1 \le \upsilon_2$ then from Theorem 6.2, we have:

$$Pr\{Q_2^{PGPS,\gamma} > q\} < C_{2'}^{PGPS}(q) e^{-\eta_1 q^{\upsilon_1}} \tag{41}$$

where

$$C_{2'}^{PGPS}(q) = C_1 e^{\eta_1 L^{\upsilon_1}} h(C_2 e^{\eta_2 L^{\upsilon_2}}) + C_2 e^{\eta_2 L^{\upsilon_2}}$$

(ii)Else if $\eta_2 \le \eta_1$ and $\upsilon_2 \le \upsilon_1$ then from Theorem 6.2:

$$Pr\{Q_2^{PGPS,\gamma} > q\} < C_2^{PGPS}(q) e^{-\eta_2 q^{\upsilon_2}} \tag{42}$$

where

$$C_2^{PGPS}(q) = C_2 e^{\eta_2 L^{\upsilon_2}} h(C_1 e^{\eta_1 L^{\upsilon_1}}) + C_1 e^{\eta_1 L^{\upsilon_1}}$$

(iii)In general, regardless of the relationship between η_1 and η_2 and that between υ_1 and υ_2, from Theorem 6.1, we have

$$Pr\{Q_2^{PGPS,\gamma} > q\} < (C_1 e^{\eta_1 L^{\upsilon_1}} + C_2 e^{\eta_2 L^{\upsilon_2}}) \times e^{-\eta(q_0^{\upsilon_{max}} - q_0^{\upsilon})} e^{-\eta q^{\upsilon}} \tag{43}$$

where

$$\eta = \frac{\eta_1 \eta_2}{\eta_1 + \eta_2} \text{ and } \upsilon = \min\{\upsilon_1, \upsilon_2\}$$

The index parameter (as well as the decay rate parameter) of the three bounds shown in (41)-(43) indicate the influence of source 1 on source 2. In the first case, the bound on $Q_2^{PGPS,\gamma}$ decays slower, and in fact, it adopts the same index parameter as that in the bound on the more heavier tailed queuing process δ_1^{SSQ,γ_1}. This means that source 2 is not guaranteed to be LRD isolated from source 1. In the second case, source 2 is not much affected by source 1 since the bound on $Q_2^{PGPS,\gamma}$ adopts the same index parameter as the bound on δ_2^{SSQ,γ_2}. Finally in the third case, which is useful when neither of the first two cases is applicable, the bound on $Q_2^{PGPS,\gamma}$ decays slower than both the bound on δ_1^{SSQ,γ_1} and the bound on δ_2^{SSQ,γ_2}.

VII. Experimental Study of a Two Queue PGPS System

In this section, we present numerical results on a PGPS system with two LRD flows to demonstrate the usefulness of the bounds given by Theorem 6.1 and Theorem 6.2. Two scenarios are considered, each involving a HT (heavier tailed) flow and a LT (lighter tailed) flow. The HT flow is a bursty portion of a Star War MPEG trace (ftp://thumper.bellcore.com/pub/vbr.video.trace), while the LT flow is an artificially generated LRD flow using aggregated Pareto On/Off processes [13].

In the first scenario, called "Non-LRD-Isolated", the HT flow appears first in the feasible ordering while the LT (lighter tailed) flow appears next (last). In the second scenario, called "LRD-Isolated", the order of the feasible ordering is reversed. The aim of the graphical presentations is to illustrate the LRD isolation properties of the two flows in the PGPS server and to demonstrate that, given the WB parameters of the individual flows, the derived PGPS bounds will always bound the actual PGPS $Pr\{Q > B\}$ distribution.

A. Non-LRD-Isolated Scenario

The "Non-LRD-Isolated" scenario is described as follows, where PGPS capacity $\gamma = 1100kbps$ and all other parameters are shown in the table below. Note that to simplify calculation, we introduce the normalized long term average rate λ_i' based on λ_i).

Also, the index parameters are obtained via least-squares matching of WB models to the numerically generated SSQ process. This method of obtaining the Hurst parameter is not recommended for real-time scenarios since the method requires the luxury of simulating the sample LRD sources countless times in order to obtain a reasonably stable SSQ tail probability curve. This method is nonetheless sufficient for us since the focus of this paper is on the multiplexing of LRD sources and is not concerned with the nitty-gritty of obtaining the index parameter of individual LRD sources. For a more professional treatment in regard to a real-time estimation of the index parameter of an individual LRD source, the method suggested in [29] can always be incorporated.

	HT Flow	LT Flow
λ_i	655kbps	345kbps
λ_i'	0.5954	0.3136
υ_i	0.6	0.8
η_i	0.00025	0.00055
$C_i e^{\eta_i L^{\upsilon_i}}$	0.44	1
ϕ_i	0.7	0.3
γ_i	0.655	0.345

Given the above weight assignment of $\phi_{HT} = 0.7$ and $\phi_{LT} = 0.3$, the feasible ordering is $\{HT, LT\}$, which in fact is, according to (18), the only feasible ordering. As a result, the bounds for the LT flows is of interest. Since

$\eta_{HT} < \eta_{LT}$, $\upsilon_{HT} < \upsilon_{LT}$, the bounds given in (42) and (43) for the LT flow are applicable. We recall from the previous discussion that with the use of bound in (42), the LT flow is expected to be severely influenced by the HT flow in this "Non-LRD-Isolated" scenario.

The WB parameters of the two flows are obtained via matching whereby the HT flow and the LT flow are simulated in SSQ conditions with service rate $\gamma_{HT} = 0.655$ and $\gamma_{LT} = 0.345$ respectively, and the resulting SSQ $Pr\{Q > B\}$ distribution is plotted out to find a matching WB curve. It is noted that the SSQ service rate, i.e., γ_{HT} and γ_{LT}, associated with the fictitious SSQ queue should be chosen as large as possible while satisfying the constraints $\gamma_i \le \varphi_i(\gamma - \sum_{j=1}^{i-1} \gamma_j)$. This is because larger γ_{HT} and γ_{LT} translate to bounds that decay faster, leading eventually to PGPS bounds that are tighter (i.e. not overly conservative).

B. LRD-Isolated Scenario

Consider the case where all parameters remain the same except the weight parameters have now been changed to $\phi_{HT} = 0.4$ and $\phi_{LT} = 0.6$. Accordingly, the feasible ordering becomes $\{LT, HT\}$, which in fact is the only feasible ordering according to (18). Since $\eta_{HT} < \eta_{LT}$, $\upsilon_{HT} < \upsilon_{LT}$, the PGPS bounds given in (41) and (43) are applicable to the HT flow. We recall from the previous discussion that with the use of the bound in (41), both flows are expected to be well LRD isolated in this "LRD-Isolated" scenario.

C. Numerical Results and Further Discussion

Figure 3 ("Non-LRD-Isolated" scenario) and Figure 4 ("LRD-Isolated" scenario) illustrate the actual queue length distribution of HT and LT flows in comparison with the applicable PGPS bounds given by (41), (42) and (43). More specifically, Figure 3(a) and Figure 4(a) illustrate the numerical results for first flow in the feasible ordering. Since the first flow in the feasible ordering is always fully LRD isolated, the PGPS bound for the first flow is also the flow's SSQ Weibull bound (found by matching). On the other hand, Figure 3(b) and Figure 4(b) are the plots of interest as they illustrate the applicable PGPS bounds for the second (i.e., last) flow in the feasible ordering, and provide information on whether the flow is affected by the first flow preceding it in the feasible ordering.

It is clear from Figure 3(b) and Figure 4(b) that the PGPS bounds are effective in predicting whether the decay slope of the last flow will be affected by the first flow because they match the decay slope of the actual PGPS queue length distribution. In the "Non-LRD-Isolated" scenario, it is noted that the decay slope of the second flow, the LT flow, has adopted the decay slope of the first flow, the HT flow, meaning that the HT flow has induced LRD burstiness into the LT flow. Both bounds from (41) and (43) are fairly close in predicting that there will be induced burstiness in the LT flow. In the "LRD Isolated" scenario, the PGPS bound given by (42) provides a more accurate prediction of the decay slope of the HT flow compared to the PGPS bound given by (43), but both bounds from (42) and (43) indicate that the decay rate slope of the HT flow will not be affected by the LT flow. Finally, it is noted that in both scenarios, the asymptotic constant of the PGPS bound is rather conservative. This is because in the formulation of the GPS and PGPS bounds, multiplexing gain was not considered.

VIII. Conclusion

In this paper, we have established the relationship between what we call a Weibull Bounded Burstiness (WBB) arrival process and a Weibull Bounded (WB) queuing process, which brings more validity in the analysis of the upper bounds on the queuing process with long range dependent (LRD) traffic inputs.

As a major contribution, we have developed, via analysis, several upper bounds on the queue length distribution of the Generalized Processor Sharing (GPS) scheduling discipline with LRD traffic inputs. The GPS bounds have also been extended to a Packet-based GPS (PGPS) system. These explicit bounds show that the long range dependency and queue length distribution of an LRD source in a GPS system will in general not be adversely affected despite the presence of other admitted sources as long as it can be feasibly ordered before other heavier tailed flows. This observation has been verified via simulation involving two LRD flows in two different PGPS scenarios as well as from other independent research efforts.

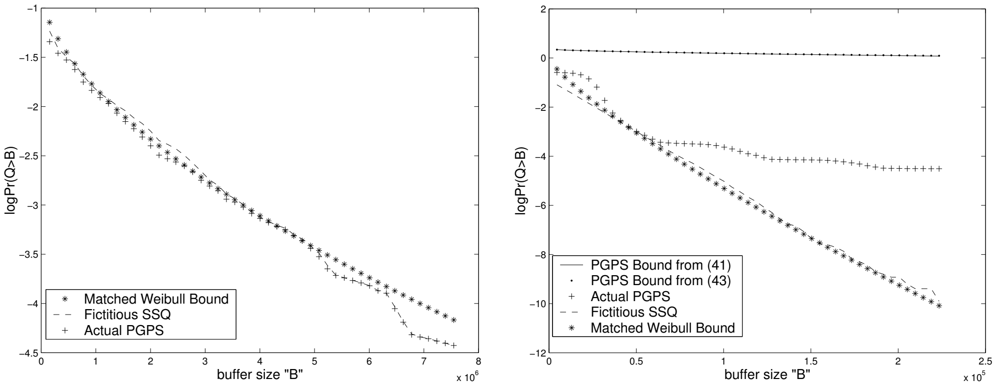
In addition, we have introduced a useful notion called LRD isolation, which is different from the conventional understand of flow isolation. This notion broadens the range of services that can be offered by GPS systems by admitting flows requiring less strict performance guarantee (i.e., LRD isolation) than flow isolation that cannot be admitted otherwise. Based on this notion, we have provided a fast and effective method to test whether a flow can be guaranteed to be LRD isolated with a given GPS service weight assignment. When the LRD isolation of a flow cannot be ensured with the current weight assignment, the proposed method can also be used to determine the minimum weight required by the flow to ensure its LRD isolation, and thus may be used to adjust its weight to the minimum required value without affecting other flows' LRD isolation. This is particularly useful for on-line call admission control where the QoS of both the new and existing flows need to be satisfied. As LRD isolation could be useful for individual flow performance in GPS servers throughout multi-hop networks, and this study will remain as our future work.

References

[1] X. Yu, I. Thng, Y. Jiang, and C. Qiao, "Upper bounds for individual queue length distribution in GPS with LRD traffic inputs," in *Proceedings of GLOBECOM*, 2001, vol. 4, pp. 2336–2340.

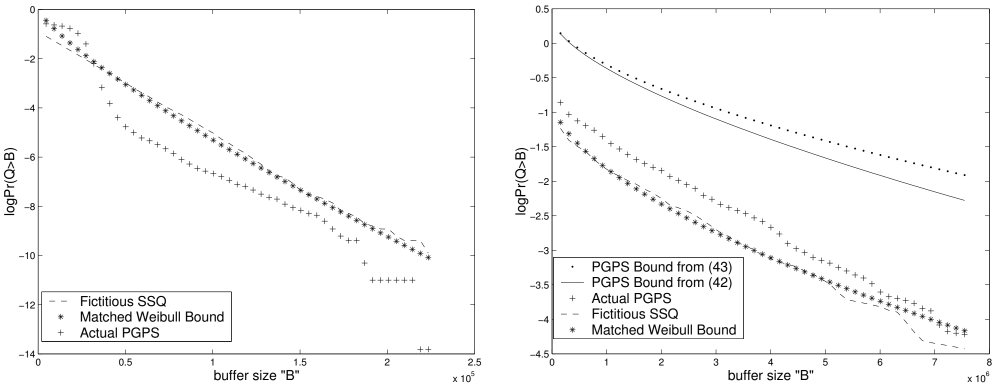
[2] A. Parekh and R. Gallager, "A generalized processor sharing approach to flow control in integrated services networks: The single node case," *IEEE/ACM Transactions on Networking*, vol. 1, no. 3, pp. 344–357, 1993.

[3] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of ethernet traffic (extended version)," *IEEE/ACM Transactions on Networking*, vol. 2, pp. 1–15, 1994.



(a) HT Flow (1st in Feasible Ordering) (b) LT Flow (Last in Feasible Ordering)

Fig. 3. Queue length distribution of HT flow and LT flow in Scenario "Non-LRD-Isolated"



(a) LT Flow (1st in Feasible Ordering) (b) HT Flow (Last in Feasible Ordering)

Fig. 4. Queue length distribution of LT flow and HT flow in Scenario "LRD-Isolated"

[4] M. Crovella and A. Bestavros, "Self-similarity in world wide web traffic: Evidence and possible causes," *IEEE/ACM Transactions on Networking*, vol. 5, no. 6, pp. 835–846, 1997.

[5] J. Beran, R. Sherman, M. Taqqu, and W. Willinger, "Long-range dependence in variable-bit-rate video traffic," *IEEE Transaction on Communication*, vol. 43, pp. 1566 – 1579, 1995.

[6] D. Heyman and T. Lakshman, "What are the implications of long- range dependence for vbr-video traffic engineering," *IEEE/ACM Transactions on Networking*, vol. 4, no. 3, pp. 301–317, 1996.

[7] Y. Gao G. He and J. Hou, "A case for exploiting self-similarity of network traffic in TCP congestion control," in *Proceedings of IEEE International Conference on Network Protocols*, 2002, pp. 34–43.

[8] A. Greenberg and N. Madras, "How fair is fair queuing?," *Journal of ACM*, vol. 39, no. 3, pp. 568 – 598, 1992.

[9] J. Turner, "New directions in communications," *IEEE Communications Magazine*, vol. 42, no. 1, 1986.

[10] O. Yaron and M. Sidi, "Generalized processor sharing networks with exponentially bounded burstiness arrivals," in *Proceedings of INFOCOM*, 1994.

[11] Z. Zhang, D. Towsley, and J. Kursose, "Statistical analysis of the generalized processor sharing scheduling discipline," *IEEE Journal on Selected Areas in Communications*, vol. 13, pp. 1071–1080, 1995.

[12] O. Yaron and M. Sidi, "Performance and stability of communication networks via robust exponential bounds," *IEEE/ACM Transactions on Networking*, vol. 1, pp. 372–385, 1993.

[13] R. Addie, M. Zukerman, and T. Neame, "Broadband traffic modeling: Simple solutions to hard problems," *IEEE Communications Magazine*, vol. 36, no. 8, pp. 88–95, 1998.

[14] S. Borst, O. Boxma, and P. Jelenkovic, "Asymptotic behavior of generalized processor sharing with long-tailed traffic sources," in *Proceedings of INFOCOM*, 2000, vol. 2, pp. 912–921.

[15] S. Borst, O. Boxma, and P. Jelenkovic, "Induced burstiness in generalized processor sharing queues with long-tailed traffic flows," in *Proceedings of the 37th Annual Allerton Conference on Communication, Control and Computing*, 1999.

[16] M. Uitert and S. Borst, "Generalized processor sharing networks fed by heavy-tailed traffic flows," in *Proceedings of INFOCOM*, 2001.

[17] A. Elwalid, D. Heyman, and T. Lakshman, "Fundamental bounds and approximations for ATM multiplexers with applications to video teleconferenceing," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 6, pp. 1004– 1016, 1995.

[18] G. Veciana and G. Kesidis, "Bandwidth allocation for multiple qualities of service using generalized processor sharing," *IEEE/ACM Transactions on Networking*, vol. 42, no. 1, pp. 268–272, 1996.

[19] D. Starobinski and M. Sidi, "Stochastically bounded burstiness for communication networks," *IEEE Transactions on Information Theory*, vol. 46, no. 1, pp. 206–216, 2000.

[20] J. Qiu and E. Knightly, "Inter-class resource sharing using statistical service envelopes," in *Proceedings of INFOCOM*, 1999, vol. 3, pp. 1404–1411.

[21] V. Paxson and A. Floyd, "Wide area traffic: The failure of poisson modeling," *IEEE/ACM Transactions on Networking*, vol. 3, no. 3, pp. 226–244, 1995.

[22] W. Willinger, M. Taqqu, W. Leland, and D. Wilson, "Self-

similarity in high-speed packet traffic: Analysis and modeling of ethernet traffic measurements," *Statistical Science*, vol. 10, pp. 67–85, 1995.

[23] A. Elwalid, D. Mitra, and R. Wentworth, "Design of generalized processor sharing schedulers which statistically multiplex heterogeneous QoS classes," in *Proceedings of INFOCOM*, 1999.

[24] N. Rananand, "Upper-bounds for tail probability of a queue with long-range dependent input," in *Proceedings of IEEE ICC*, 1998, vol. 3, pp. 1466–1472.

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

[26] A. Parekh and R. Gallager, "A generalized processor sharing approach to flow control in integrated services networks: The multiple node case," *IEEE/ACM Transactions on Networking*, vol. 2, no. 2, pp. 137–150, 1994.

[27] J. Kurose and K. Ross, "Computer networking: A top-down approach featuring the internet," in *Addison Wesley*, 2001.

[28] A. Parekh, "A generalized processor sharing approach to flow control in integrated services networks," *Ph.D. thesis, Department of Electrical Engineering and Computer Science, MIT*, 1992.

[29] T. Karagiannis, M. Faloutsos, and R. Riedi, "Long-range dependence: Now you see it, now you don't!," in *Global Internet Symposium*, 2002.

Appendix A

The derivation of Lemma 3.1 is provided in this appendix. First we note that

$$\begin{aligned}
&E[\int_s^t A(u)du] \\
&= \int_{x=0}^{\infty} Pr\{\int_s^t A(u)du > x\}dx \\
&= \int_{x=0}^{\rho(t-s)} Pr\{\int_s^t A(u)du > x\}dx + \\
&\int_{x=0}^{\infty} Pr\{\int_s^t A(u)du > \rho(t-s)+x\}dx \\
&< \rho(t-s) + \int_{x=0}^{\infty} Ce^{-\mu x^{\upsilon}}dx. \qquad (44)
\end{aligned}$$

Secondly, as long as $\upsilon > 0$, we have

$$\begin{aligned}
\lim_{t-s\to\infty} \frac{\int_{x=0}^{\infty} Ce^{-\mu x^{\upsilon}}dx}{t-s} &= \lim_{t\to\infty} \frac{\int_{x=0}^{t} Ce^{-\mu x^{\upsilon}}dx}{t} \\
&= \lim_{t\to\infty} Ce^{-\mu t^{\upsilon}} = 0
\end{aligned} \qquad (45)$$

Therefore,

$$\rho \geq \lim_{t-s\to\infty} \frac{E[\int_s^t A(u)du]}{t-s} \square$$

Appendix B

Appendix B provides a list of inequalities which are heavily utilized throughout the paper. The derivations for these inequalities are also provided.

Inequality 1: For all $y_1, y_2, ..., y_N > 0$ and $0 \leq x < 1$:

$$y_1^x + y_2^x + ... + y_N^x \geq (y_1 + y_2 + ... + y_N)^x \qquad (46)$$

Inequality 2: For all $y_1, y_2, ..., y_N > 0$ and $0 \leq x < 1$:

$$y_1^x + y_2^x + ... + y_N^x \leq N(y_1 + y_2 + ... + y_N)^x \qquad (47)$$

Inequality 3: For all $y_1 > y_2 > 0$ and $0 \leq x < 1$:

$$y_1^x - y_2^x \leq (y_1 - y_2)^x \qquad (48)$$

Derivations: Let $f_s(x) = s^x$ where $0 \leq x < 1$, $s > 0$. Using the single prime to denote the derivative of a function, we obtain

$$f'_s(x) = (s^x)' = (e^{x \ln s})' = e^{x \ln s} \times \ln s \qquad (49)$$

Since

$$\ln s > 0 \text{ for } s > 1 \text{ and } \ln s < 0 \text{ for } s < 1 \qquad (50)$$

hence

$$f'_s(x) > 0 \text{ for } s > 1 \text{ and } f'_s(x) < 0 \text{ for } s < 1 \qquad (51)$$

It is clear for any $s < 1$, when x increases, $f_s(x)$ will decrease, i.e.

$$f_s(1) < f_s(x) < f_s(0) \text{ as long as } s < 1 \text{ and } 0 \leq x < 1 \qquad (52)$$

Let us define a set of variables $s_i = \frac{y_i}{y_1+y_2+...+y_N}$ where all $y_i's > 0$ and corresponding function

$$f_{s_i}(x) = s_i^x = (\frac{y_i}{y_1 + y_2 + ... + y_N})^x$$

It is clear that $0 < s_i < 1$, therefore $f_{s_i}(1) < f_{s_i}(x) < f_{s_i}(0)$, $0 < x < 1$.

Now, the proof for Inequality 1:

$$\begin{aligned}
&\sum_{i=1}^{N} f_{s_i}(x) \geq \sum_{i=1}^{N} f_{s_i}(1) \Leftrightarrow \frac{y_1^x + y_2^x + ... + y_N^x}{(y_1 + y_2 + ... + y_N)^x} \geq 1 \\
&\Leftrightarrow y_1^x + y_2^x + ... + y_N^x \geq (y_1 + y_2 + ... + y_N)^x \qquad (53)
\end{aligned}$$

Now, the proof for Inequality 2:

$$\begin{aligned}
&\sum_{i=1}^{N} f_{s_i}(x) \leq \sum_{i=1}^{N} f_{s_i}(0) \Leftrightarrow \frac{y_1^x + y_2^x + ... + y_N^x}{(y_1 + y_2 + ... + y_N)^x} \leq N \\
&\Leftrightarrow y_1^x + y_2^x + ... + y_N^x \leq N(y_1 + y_2 + ... + y_N)^x \qquad (54)
\end{aligned}$$

Now, the proof for Inequality 3: Given the condition that $y_1 > y_2$, therefore, from Inequality 1,

$$\begin{aligned}
(y_1 - y_2)^x + y_2^x &\geq [(y_1 - y_2) + y_2]^x = y_1^x \\
&\Leftrightarrow (y_1 - y_2)^x \geq y_1^x - y_2^x \square \qquad (55)
\end{aligned}$$

Appendix C

The derivation of Lemma 3.2 is provided in this appendix. Set $w \geq 0$, and let $0 < p < 1$ be some constant. Then we have

$$\begin{aligned}
&\{W_1(t) + W_2(t) > w\} \\
&\subset \overline{\{W_1(t) \leq pw\} \bigcap \{W_1(t) \leq (1-p)w\}} \\
&= \overline{\{W_1(t) \leq pw\}} \bigcup \overline{\{W_1(t) \leq (1-p)w\}} \\
&= \{W_1(t) > pw\} \bigcup \{W_1(t) > (1-p)w\}
\end{aligned}$$

Thus,

$$\begin{aligned}&Pr\{W_1(t)+W_2(t)>w\}\\&\le Pr\{W_1(t)>pw\}+Pr\{W_2(t)>(1-p)w\}\\&< C_1e^{-\eta_1 p w^{\upsilon_1}}+C_2e^{-\eta_2(1-p)w^{\upsilon_2}}\end{aligned}$$

Choose p according to [19] such that $\eta_1 p=(1-p)\eta_2$, i.e. $p=\frac{\eta_2}{\eta_1+\eta_2}$. Define $\eta=\frac{\eta_1\eta_2}{\eta_1+\eta_2}$ and $\upsilon=\min(\upsilon_1,\upsilon_2)$, we obtain

$$Pr\{W_1(t)+W_2(t)>w\}<C_1e^{-\eta w^{\upsilon_1}}+C_2e^{-\eta w^{\upsilon_2}} \tag{56}$$

If $0<w<1$, then (56) can reduce to

$$Pr\{W_1(t)+W_2(t)>w\}<(C_1+C_2)e^{-\eta w^{\upsilon_{max}}} \tag{57}$$

where $\upsilon_{max}=max[\upsilon_1,\upsilon_2]$;

If $w>1$, then (56) can reduce to

$$Pr\{W_1(t)+W_2(t)>w\}<(C_1+C_2)e^{-\eta w^{\upsilon}} \tag{58}$$

where $\upsilon=min[\upsilon_1,\upsilon_2]$;

It is noted that both (57) and (58) have the Weibull bound form except with different index parameters. The next strategy is to combine (57) and (58) so that the same index parameter, namely υ rather than υ_{max}, can also be used for the case where $0<w<1$. Although this case where the queue length is less than 1 is unimportant, for completeness of the bound, we still consider how to provide a bound this range in the following discussions. Firstly, we notice that the bound using the index υ_{max} in (57) is always larger than the bound using the index in (58) for the range $0<w<1$. Once $w>1$, the bound in $0<w<1$ is always larger than the bound in (57). At $w=0$ and at $w=1$, the bounds in (57) and (58) have exactly the same values. Hence in order to extend (58), which uses the index υ, to provide a bound for the case where $0<w<1$, we can always add an additional asymptotic constant factor C_2^* to raise the bound of (58). This additional asymptotic constant C_2^* can be easily obtained and is related to the maximum displacement between (57) and (58) when $0<w<1$. More specifically, let

$$f(w)=(C_1+C_2)e^{-\eta w^{\upsilon_{max}}}-(C_1+C_2)e^{-\eta w^{\upsilon}} \tag{59}$$

Notice that $f(w)$ is zero at $w=0$ and $w=1$, and $f(w)>0$ only for $0<w<1$, where both $e^{-\eta w^{\upsilon_{max}}}$ and $e^{-\eta w^{\upsilon}}$ monotonically decreases with w, therefore, there exists a unique maximum point of $f(w)$ for $w\in(0,1)$. Let w_0 maximize $f(w)$ for $0<w<1$. Specifically, w_0 is the solution to the following non-algebraic equation:

$$\frac{e^{-\eta w^{\upsilon_{max}}}}{w^{\upsilon_{max}}}\upsilon_{max}=\frac{e^{-\eta w^{\upsilon}}}{w^{\upsilon}}\upsilon \tag{60}$$

Hence the additional asymptotic constant C_2^* is given by:

$$\begin{aligned}C_2^* &= f(w_0)\times e^{\eta w_0^{\upsilon}}\\&=(C_1+C_2)[e^{-\eta w_0^{\upsilon_{max}}}-e^{-\eta w_0^{\upsilon}}]\times e^{\eta w_0^{\upsilon}}\\&=(C_1+C_2)[e^{-\eta(w_0^{\upsilon_{max}}-w_0^{\upsilon})}-1]\end{aligned} \tag{61}$$

Therefore,

$$Pr\{W_1(t)+W_2(t)>w\}<(C_1+C_2+C_2^*)e^{-\eta w^{\upsilon}}$$

which can be further simplified to:

$$Pr\{W_1(t)+W_2(t)>w\}<(C_1+C_2)e^{-\eta(w_0^{\upsilon_{max}}-w_0^{\upsilon})}e^{-\eta w^{\upsilon}} \tag{62}$$

where $w\ge 0$.□

It is noted that if $\upsilon_1=\upsilon_2$ (i.e. the LRD queuing processes have identical index parameters or Hurst parameters), Hence the bound in (62) is a tighter bound in the index parameter for the case of two aggregate LRD queuing processes with identical index parameters compared to the case of two aggregate LRD queuing processes with different index parameters. In the latter case, the bound adopts the index parameter of the more heavier tailed queuing process making the bound conservative.

Finally, without loss of generality, let $\eta_2\le\eta_1$. Hence the range of values for η is as follows:

$$\frac{\eta_2}{2}\le\eta\le\eta_2 \text{ , i.e. } \frac{\min(\eta_1,\eta_2)}{2}\le\eta\le\min(\eta_1,\eta_2) \tag{63}$$

The lower bound for η is obtained by considering the situation where $\eta_1\approx\eta_2$ while the upper bound is obtained by considering the situation where η_2 is far smaller than η_1.

Appendix D

The derivation of Lemma 3.3 and Theorem 3.3 is provided in this appendix.

To derive Lemma 3.3, set $t>0$ and $w>2$. Let w_1^t denote the probability density function of $W_1(t)$. We have:

$$\begin{aligned}&Pr\{W_1(t)+W_2(t)>w\}\\&=\int_{q=0}^{\infty}Pr\{W_1(t)\in[q,q+dq]\}\times Pr\{W_2(t)>w-q\}\\&=\int_{q=0}^{\infty}w_1^t(q)\times Pr\{W_2(t)>w-q\}dq\end{aligned} \tag{64}$$

The integral in (64) can be decomposed into 2 portions, i.e., the first goes from $q=0$ to $q=w$ and the second from $q=w$ to $q=\infty$.

For portion 1, as we consider $w\ge 2$, we have:

$$\begin{aligned}&\int_{q=0}^{w}w_1^t(q)\times Pr\{W_2(t)>w-q\}dq\\&<\int_{q=0}^{w}w_1^t(q)C_2e^{-\eta_2(w-q)^{\upsilon_2}}dq\\&=(-\int_q^{\infty}w_1^t(u)du)\times C_2e^{-\eta_2(w-q)^{\upsilon_2}}|_{q=0}^{w}+\\&\int_{q=0}^{w}C_2\eta_2\upsilon_2(w-q)^{\upsilon_2-1}e^{-\eta_2(w-q)^{\upsilon_2}}\int_q^{\infty}w_1^t(u)dudq\\&\le C_2e^{-\eta_2 w^{\upsilon_2}}\int_0^{\infty}w_1^t(q)dq-C_2\int_w^{\infty}w_1^t(q)dq\end{aligned}$$

$$+C_1C_2\upsilon_2\eta_2\int_{q=0}^{1}(w-q)^{\upsilon_2-1}e^{-\eta_2(w-q)^{\upsilon_2}-\eta_1q^{\upsilon_1}}dq$$
$$+C_1C_2\upsilon_2\eta_2\int_{q=1}^{w}(w-q)^{\upsilon_2-1}e^{-\eta_2(w-q)^{\upsilon_2}-\eta_1q^{\upsilon_1}}dq$$
$$\leq C_2e^{-\eta_2w^{\upsilon_2}}+$$
$$C_1C_2\upsilon_2\eta_2\int_{q=0}^{1}e^{-\eta_2(w-q)^{\upsilon_2}-\eta_1q^{\upsilon_1}}dq+$$
$$C_2C_1\upsilon_2\eta_2e^{-\eta_2w^{\upsilon_2}}\int_{q=1}^{w}(w-q)^{\upsilon_2-1}e^{-\eta_1q^{\upsilon_1}}e^{\eta_2q^{\upsilon_2}}dq$$
$$\leq C_2e^{-\eta_2w^{\upsilon_2}}+C_1C_2\upsilon_2\eta_2\int_{q=0}^{1}e^{-\eta_2(w-q)^{\upsilon_2}}dq+$$
$$C_2C_1\upsilon_2\eta_2e^{-\eta_2w^{\upsilon_2}}\int_{q=1}^{w}(w-q)^{\upsilon_2-1}e^{-[\eta_1q^{\upsilon_1}-\eta_2q^{\upsilon_2}]}dq \tag{65}$$

Since the second integration in (65) can be simplified to:

$$\int_{q=0}^{1}e^{-\eta_2(w-q)^{\upsilon_2}}dq=$$
$$e^{-\eta_2(w-q)^{\upsilon_2}}|_{q=0}^{1}-\int_{q=0}^{1}d\left[e^{-\eta_2(w-q)^{\upsilon_2}}\right]$$
$$=e^{-\eta_2(w-1)^{\upsilon_2}}-e^{-\eta_2(w)^{\upsilon_2}}-$$
$$\int_{q=0}^{1}\left[\eta_2\upsilon_2(w-q)^{\upsilon_2-1}\right]e^{-\eta_2(w-q)^{\upsilon_2}}dq \tag{66}$$

Since $0<(w-q)^{\upsilon_2-1}\leq 1$, we further have

$$\begin{aligned}\int_{q=0}^{1}e^{-\eta_2(w-q)^{\upsilon_2}}dq &\leq e^{-\eta_2(w-1)^{\upsilon_2}}-e^{-\eta_2(w)^{\upsilon_2}}\\ &\leq (e^{\eta_2}-1)e^{-\eta_2w^{\upsilon_2}}\end{aligned} \tag{67}$$

Since it is assumed that $\eta_1\geq\eta_2$ and $\upsilon_1\geq\upsilon_2$, we substitute (67) into (65) and(65) can be transformed into

$$\int_{q=0}^{w}w_1^t(q)\times Pr\{W_2(t)>w-q\}dq$$
$$<C_2e^{-\eta_2w^{\upsilon_2}}+C_2C_1\upsilon_2\eta_2(e^{\upsilon_2}-1)e^{-\eta_2w^{\upsilon_2}}$$
$$+C_2C_1\upsilon_2\eta_2e^{-\eta_2w^{\upsilon_2}}\int_{q=1}^{w}(w-q)^{\upsilon_2-1}dq$$
$$\leq C_2[1+C_1\upsilon_2\eta_2(e^{\eta_2}-1)+C_1w^{\upsilon_2}\eta_2]e^{-\eta_2w^{\upsilon_2}} \tag{68}$$

For portion 2, we have:

$$\int_{q=w}^{\infty}w_1^t(q)\times Pr\{W_2(t)>w-q\}dq$$
$$\leq\int_{q=0}^{w}w_1^t(q)dq<C_1e^{-\eta_1w^{\upsilon_1}} \tag{69}$$

By plugging expressions (68) and (69) into (64) we obtain

$$Pr\{W_1(t)+W_2(t)>w\}$$
$$\leq(C_2(1+C_1\upsilon\eta(e^{\eta}-1)+C_1w^{\upsilon}\eta)+C_1)e^{-\eta w^{\upsilon}}$$
$$=(C_2h(C_1)+C_1)e^{-\eta w^{\upsilon}} \tag{70}$$

where $\eta=\min\{\eta_1,\eta_2\}$, $\upsilon=\min\{\upsilon_1,\upsilon_2\}$ and

$$h(x)=1+x\upsilon\eta(e^{\eta}-1)+xw^{\upsilon}\eta.\square$$

Based on Lemma 3.3, the proof of Theorem 3.3 can be recursively obtained but with a slight modification. Consider N queues $W_i(t)$, where $0\leq i\leq N$, that are being aggregated recursively according to the above procedure, starting with the aggregate queue $W'_{N-1}(t)=W_{N-1}(t)+W_N(t)$ first. By virtue of Lemma 3.3, it is easy to see that the decay rate and index parameter associated with $W'_{N-1}(t)$ are η_N and υ_N respectively. In addition, the asymptotic constant with parameter w associated with $W'_{N-1}(t)$ is $C_2^{WB}(w)=C_Nh(C_{N-1})+C_{N-1}$. Next, consider the aggregate queue $W'_{N-2}(t)=W_{N-2}(t)+W'_{N-1}(t)$. The decay rate and index parameter stay the same as η_N and υ_N. On the other hand, the asymptotic constant in (65) is no longer C_2 but $C_2^{WB}(w-q)$. Although the presence of $C_2^{WB}(q)$ breaks the recursive nature of (65), since $C_2^{WB}(q)$ is a monotonically increasing function of q, we can replace $C_2^{WB}(q)$ with $C_2^{WB}(w)$. This not only preserves the inequality in (65) but makes its usage recursive since the replaced asymptotic constant, i.e. $C_2^{WB}(w)$, is no longer part of the integration and can be treated in the same fashion as the constant C_2. The asymptotic constant with parameter w associated with $W'_{N-2}(t)$ now becomes $C_3^{WB}(w)=C_Nh(C_{N-1})h(C_{N-2})+C_{N-1}h(C_{N-2})+C_{N-2}$. If the aggregation is continued in this manner, i.e. $W'_{N-j}(t)=W_{N-j}(t)+W'_{N-j+1}(t)$ for $1\leq j\leq N-1$, then the proof will be completed. The asymptotic constant with parameter w associated with the last aggregate $W'_1(t)=W_1(t)+W'_2(t)=\sum_{i=1}^{N}W_i(t)$ is thus given in (17).$\square$