A Survey of an Integrated Scheduling Scheme with Long-Range and Short-Range Dependent Traffic

A. Sakira Parveen
Assistant Professor, PSNA College of Engineering & Technology, Tamilnadu, India
sakasi09@gmail.com

Abstract—Multimedia applications in wireless networks are usually categorized into various classes according to their traffic patterns and differentiated Quality-of-Service (QoS) requirements. The traffic of heterogeneous multimedia applications often exhibits the Long-Range Dependent (LRD) and Short-Range Dependent (SRD) properties, respectively. The integrated scheduling scheme that combines Priority Queuing (PQ) and Generalized Processor Sharing (GPS) within a hierarchical structure, referred to as PQ-GPS, has been identified as an efficient mechanism for QoS differentiation in wireless networks and attracted significant research efforts. However, due to the high complexity and interdependent relationship among traffic flows, modelling of the integrated scheduling scheme poses great challenges. To address this challenging and important research problem, we develop an original analytical model for PQ-GPS systems under heterogeneous LRD and SRD traffic. A cost-effective flow decomposition approach is proposed to equivalently divide the integrated scheduling system into a group of Single-Server Single-Queue (SSSQ) systems. The expressions for calculating the queue length distribution and loss probability of individual traffic flows are further derived. After validating its accuracy, the developed model is adopted as an efficient performance tool to investigate the important issues of resource allocation and call admission control in the integrated scheduling system under QoS constraints.

Index Terms—Wireless multimedia, heterogeneous network traffic, integrated scheduling, priority queuing, generalized processor sharing, performance modelling.

I. INTRODUCTION

The past decade has witnessed a rapid increase in various advanced multimedia applications with diverse Quality of Service (QoS) requirements over wireless communication networks. In order to efficiently support these applications, the Differentiated Services (DiffServ) architecture has been proposed as a scalable and coarse-grained mechanism for managing heterogeneous wireless network traffic. The data packets entering a DiffServ domain are placed into a limited number of traffic classes, which are handled in differentiated manners, ensuring preferential treatment for the high priority traffic. In general, the DiffServ architecture can provide low delay for real-time applications, while offering best-effort services for non-real-time ones.

DiffServ can, for example, be used to provide low-latency to critical network traffic such as voice or streaming media while providing simple best-effort service to non-critical services such as web traffic or file transfers. It doesn’t transit state information across Internet but works on aggregates. It functions on both IPv4 and IPv6 and it uses one header field in an IP-packet. A packet receives forwarding treatment from network devices based on the value on the field. Network devices map the packet to some behavioural aggregate in which the packets receive uniform treatment. Different services are not standardised within DiffServ. Instead, the building blocks are standardised from which the services can be built.

Traffic scheduling plays a key role in the design and performance of the DiffServ architecture. The integrated scheduling mechanism which combines different fundamental scheduling schemes, such as Priority Queuing (PQ), and Generalized Processor Sharing (GPS) or its variants within a hierarchical structure, has attracted significant research interests from both academia and industry. For instance, Nikolouzou studied the definition and deployment issues of DiffServ networks, where the integrated scheduling mechanism is identified as a suitable traffic handling scheme. Maniatis proposed to adopt the integrated scheduling mechanism to deploy QoS-enabled end-to-end services in 3G wireless networks. More-over, the existing studies have applied the integrated scheduling schemes to investigate the issues of bandwidth sharing, buffer allocation and QoS differentiation in IEEE 802.16/WiMAX networks. In industry, Cisco has developed two hybrid schedulers, namely, IP Real-time Transport Proto-col (RTP) Priority and Low Latency Queuing (LLQ), which are essentially based on the PQ-GPS principle and have been deployed on a variety of Cisco gigabit switch routers, such as, Cisco 12000 series, Cisco 2600 and 3600 series.

In the integrated PQ-GPS scheduling system, one traffic flow is served with the strictly high priority. As a consequence, this flow experiences low loss, low delay and jitter, which are actually the QoS requirements desirable for real-time applications, e.g., the conversational applications in wireless networks. On the other hand, other traffic flows of the low priority in the PQ-GPS system are handled by the conventional GPS scheduling mechanism. Each traffic flow in GPS is assigned a weight in order to guarantee a minimum service rate and provide differentiated QoS, even though other flows may be greedy in demanding service. This property offers forwarding assurance to individual traffic flows and prevents them from experiencing service starvation.

Modelling and analysis of scheduling schemes is an important issue for efficient resource management and optimization of communication networks. In order to obtain a proper and deep understanding of the performance behaviors of the integrated scheduling scheme, it is essential to take the realistic characteristics of network traffic into account. Many high quality measurement studies have demonstrated that the real-world traffic in modern communication networks with multimedia applications exhibits heterogeneous properties and can be categorized into two significantly different classes.
The classes are Long-Range Dependent (LRD) and Short-Range Dependent (SRD) traffic, respectively. Specifically, the traffic generated by conversational applications usually exhibits short-range dependency and can be well characterized by Markovian stochastic processes, while the traffic of various streaming and background applications often reveals noticeable long-range dependency (i.e., large-lag correlation and scale-invariant burstiness). It is worth noting that both SRD and LRD traffic has been widely discovered in wireless networks. However, most existing performance studies on traffic scheduling have been confined to either SRD traffic or LRD traffic only, neither of which is able to solely capture the heterogeneous properties of the realistic wireless network traffic. Furthermore, the majority of the existing studies on analytical modelling of traffic scheduling schemes has been focused on the relatively simple systems with the fundamental PQ or GPS mechanisms involved, separately. Due to the high complexity and the interdependent relationships among heterogeneous traffic flows, analytical modelling of the integrated scheduling scheme poses more challenges.

In this paper, an analytical model is developed for investigating the queue length distribution and loss probability of the integrated PQ-GPS scheduling scheme in the presence of heterogeneous traffic. This approach is extended in presented in and model the heterogeneous traffic via LRD fractional Brownian motion (fBm) processes and SRD Markov-Modulated Poisson Processes (MMPPs). The fBm model has been identified as an efficient way for modelling and generating LRD traffic. On the other hand, the MMPP model has been extensively used for characterizing SRD bursty traffic owing to its ability of capturing the time-varying arrival rate and important correlation between inter-arrival times while still maintaining analytical tractability. The major contributions of the paper can be summarized as follows: (1) A Large Deviation Principle (LDP)-based method subject to Gaussian traffic is applied in a creative way to handle the heterogeneous multi-service fBm and MMPP traffic flows. The expression for calculating the total queue length distribution of the integrated PQ-GPS system is then derived. (2) In order to make the challenging issue of modelling the integrated system tractable, a flow decomposition approach is developed to divide the integrated scheduling system into a group of equivalent, but relatively simple, Single-Server Single-Queue (SSSQ) systems and obtain their service capacities. (3) By virtue of the equivalent relationship between these SSSQ systems and the PQ-GPS system, we derive the analytical expressions of the queue length distribution and loss probability of individual traffic flows in the integrated scheduling system. (4) The accuracy of the developed model is validated through the extensive comparison between analytical and simulation results under various working scenarios with typical parameter settings. (5) The developed model is used to investigate the important issues of resource management and call admission control of the integrated scheduling scheme in wireless networks. These contributions are simpler and inexpensive.

The rest of this paper is organized as follows. Section II presents the integrated PQ-GPS system and reviews the modelling issue of fBm and MMPP traffic. In Section III, the total queue length distribution of the PQ-GPS system is derived. Section IV presents the flow decomposition approach in detail, based on which we calculate the queue length distribution and loss probability of individual traffic flows in Section V. Section VI validates the effectiveness and accuracy of the model and Section VII demonstrates its applications. Finally, the paper is concluded in Section VIII.

II. PQ-GPS Systems with Heterogeneous Network Traffic

This section outlines the integrated PQ-GPS scheduling system under heterogeneous network traffic and then presents how to model LRD and SRD traffic.

A. System Model

Figure 1 shows a schematic diagram of the PQ-GPS system subject to heterogeneous MMPP and fBm traffic flows. This integrated scheduling system is motivated by the well-known DiffServ architecture where the traffic flows are classified into three classes, namely, Expedited Forwarding (EF), Assured Forwarding (AF), and Best-Efforts (BE), according to their distinct QoS requirements. The EF traffic class, e.g., real-time voice traffic, has strict QoS requirements in terms of low packet latency and jitter and thus is served by the PQ policy with high priority. The AF traffic class should be guaranteed with a given service capacity and be allowed to access extra service, if available, while the BE class is served with the remaining service capacity. Therefore, one queue in the GPS system can serve the AF traffic class if it is allocated a relatively high weight according to its service requirement, while the other queue can be left for the BE traffic class.

Since a two-state MMPP is often used to model the SRD bursty traffic generated by real-time conversational voice applications [4], [8], [9], the high-priority Flow 1 in Fig. 1 is denoted $MMPP_1$. The residual service capacity of the integrated scheduling system is supplied to other traffic flows that are handled by GPS at the low level, modelled by the LRD fBm process and denoted by $fBm_2$ and $fBm_3$, respectively. Note that in Fig. 1, parameters $\mu_2$ and $\mu_3$ denote the weights of $fBm_2$ and $fBm_3$ ($\mu_2 + \mu_3 = 1$) under the GPS scheduling policy.
B. Modelling of Heterogeneous fBm and MMPP Traffic

1) Fractional Brownian Motion (fBm): Let \( X = \{ X(t) \}_{t \in \mathbb{N}} \) be a covariance stationary stochastic process with variance \( \sigma^2 \) and autocorrelation function \( r(k) \) \( (k \geq 0) \). Process \( X \) is said to be LRD, if
\[
r(k) \sim \rho(k)k^{-\alpha}, \quad \text{as } k \to \infty,
\]
where \( \alpha \in (0,1) \) and \( \rho(k) \) is a slowly varying function, that is, for \( \forall x > 0, \rho(tx) \sim \rho(t) \) as \( t \to \infty \). Equation (1) denotes that \( r(k) \) is asymptotically equivalent to \( \rho(k)k^{-\alpha} \), i.e., \( r(k) / (\rho(k)k^{-\alpha}) \to 1 \) as \( k \to \infty \).

The innovative study of Leland, triggered an explosion of research on LRD traffic. Many analytical models and techniques have been developed to characterize or generate LRD traffic. Among these models, fBm has been identified as an efficient way for modelling LRD self-similar traffic in terms of both time and space complexity and is thus adopted in this study. A standard fBm process is a Brownian motion without independent increments. It is essentially a continuous-time Gaussian process with zero expectation at any time. In general, a traffic flow can be modelled as a stochastic process in a cumulative arrival form as \( A = \{ A(t) \}_{t \in \mathbb{N}}, \) where \( A(t) \) is the cumulative amount of traffic arrived up to time \( t \). The mean \( A(s, t) = A(t) - A(s) \) can denote the amount of traffic arrived in time interval \( (s, t) \). For an fBm traffic flow, \( A_f \), the corresponding \( A_f(t) \) can be expressed as
\[
A_f(t) = \lambda_f t + Z_f(t),
\]
where \( \lambda_f \) is the mean arrival rate and \( Z_f(t) = \sqrt{a_f}Z_H(t) \), Parameter \( a_f \) is the variance coefficient of \( A_f(t) \) and \( Z_f(t) \) is a standard fBm with variance \( \tilde{v}_f(t) = t^{2H_f} \), where \( H_f \in [\frac{1}{2}, 1] \) is the Hurst parameter. Based on \( \tilde{v}_f(t) \), the variance function of \( A_f(t) \) can be given as follows:
\[
v_f(t) = a_f \lambda_f \tilde{v}_f(t) = a_f \lambda_f t^{2H_f}.
\]

The increment process of \( A_f \), denoted as \( B_f = \{ B_f(t) \}_{t \in \mathbb{N}} \), where \( B_f(t) = A_f(t + 1) - A_f(t) \), is called fractional Gaussian noise. Its autocorrelation function is given by
\[
r(k) = \frac{1}{2}(|k + 1|^{2H_f} - 2|k|^{2H_f} + |k - 1|^{2H_f}).
\]

For \( H_f \in [\frac{1}{2}, 1] \), the autocorrelation function can be expressed as
\[
r(k) \sim H_f(2H_f - 1)k^{2H_f - 2}, \quad \text{as } k \to \infty.
\]

In other words, \( B_f \) is an LRD process while \( A_f \) is its cumulative arrival form.

2) Markov Modulated Poisson Process (MMPP): An MMPP is a doubly stochastic Poisson process with the arrival rate being modulated by a multi-state irreducible Markov process. In other words, a few Poisson processes in an MMPP are switched to be active according to an underlying Markov process. MMPPs can well characterize the properties of the time-varying arrival rate and correlation between in-terarrival times of a stochastic process. Therefore, they can be employed to model the burstiness of network traffic and are analytically tractable. It has been shown that MMPPs are fairly accurate in modelling voice traffic. In this paper, we use a two-state MMPP, denoted by \( MMPP_1 \), to characterize the high priority traffic generated by real-time conversational applications. Specifically, \( MMPP_1 \) can be parameterized by an infinitesimal generator matrix, \( \Omega \), and an arrival rate matrix, \( \Lambda \).

\[
\Omega = \begin{bmatrix} -\delta_1 & \delta_1 \\ \delta_2 & -\delta_2 \end{bmatrix}, \quad \text{and} \quad \Lambda = \begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix},
\]

where \( \delta_i (i = 1,2) \) represents the transition rate from state \( i \) to the other, and \( r_i (i = 1, 2) \) is the traffic arrival rate when the MMPP is at state \( i \). Similar to the representation of an fBm traffic flow, the cumulative amount of traffic arrived up to time \( t \), \( A_m(t) \), of an MMPP flow can be expressed as
\[
A_m(t) = \lambda_m t + Z_m(t),
\]
where \( \lambda_m = (r_1\delta_2 + r_2\delta_1)/(\delta_1 + \delta_2) \) is the mean arrival rate of \( A_m(t) \) and \( Z_m(t) \) is a stochastic process with mean \( E(Z_m(t)) = 0 \). The variance function of \( Z_m(t) \) and \( A_m(t) \) is given by [9]
\[
v_m(t) = \frac{r_1\delta_2 + r_2\delta_1}{\delta_1 + \delta_2}t + \frac{2(r_1 - r_2)^2\delta_1\delta_2}{(\delta_1 + \delta_2)^3} t \bigg[ 1 - e^{-(\delta_1 + \delta_2)t} \bigg].
\]

III. TOTAL QUEUE LENGTH DISTRIBUTION OF THE PQ-GPS SYSTEM

In this section, the expression is derived to calculate the total queue length distribution of the integrated PQ-GPS system by adopting an LDP-based method to handle the heterogeneous fBm and MMPP traffic. An LDP characterizes the asymptotic behavior of remote tails of sequences of random variables based on a non-negative rate function. It particularly examines the exponential decay of the probability measures of extreme or tail events, as the number of observations increases arbitrarily large. LDP has been successfully applied in information theory and risk management. Our work is built on the LDP specialized for Gaussian processes where the reproducing kernel Hilbert space is used to deal with the Gaussian case. This method requires that the variance \( v_k(t) \) of Gaussian traffic flow \( A_k \) fed into the queuing system satisfies: \( \exists \alpha < 2 \), such that \( \lim_{t \to \infty} v_k(t)/t^\alpha = 0 \). It has been successfully applied to study the queue length distribution of PQ and GPS systems subject to Gaussian traffic flows. The Gaussian characteristics of real network traffic has been demonstrated by measurement studies.

Since an fBm is essentially a Gaussian process, in order to adopt the LDP method to derive the queue length distribution of the PQ-GPS system under both MMPP and fBm traffic, we need to address how an MMPP can be approximated as a Gaussian process. We find that the following two properties of MMPP make such an approximation feasible. Firstly, according to Equation (8), the variance \( v_m(t) \) of an MMPP traffic flow is continuous and \( \forall \alpha (1 < \alpha < 2), \lim_{t \to \infty} v_m(t)/t^\alpha = 0 \). Secondly, an MMPP can be regarded as the superposition of two independent interrupted Poisson processes, which can be approximated as Gaussian processes as time tends to infinity. Since the superposition of independent Gaussian processes is still Gaussian, \( A_m(t) \).
can be reasonably approximated as a Gaussian process with \( A_m(t) \sim N(\lambda_m t, v_m(t)) \), as \( t \to \infty \).

Now let us consider the PQ-GPS system subject to MMPP and fBm traffic flows. Let \( A_k = \{A_k(t)\}_{t \in \mathbb{R}} (1 \leq k \leq 3) \) be the \( k^{th} \) traffic flow fed into the queuing system and \( A_k(s, t) \) be the amount of this traffic flow arrived during time interval \((s, t]\). Consequently, the total queue length, \( Q(t) \), of the system at time \( t \) can be denoted as

\[
Q(t) = \sup_{s \leq t} \left\{ \sum_{k=1}^{3} A_k(s, t) - C(t-s) \right\},
\]

where \( C \) represents the service capacity of the PQ-GPS system. Note that theoretically, as long as the sum of the mean arrival rates of all traffic flows is less than the service capacity, the PQ-GPS system is able to converge to a steady state. However, practically the service capacity, \( C \), of the PQ-GPS system should be much larger than the aggregated arrival rates of all traffic flows because of the bursty and dynamic nature of network traffic.

Based on the LDP method, we can derive the upper and lower bounds of the total queue length distribution, \( P(Q > x) \), of the PQ-GPS system in the presence of heterogeneous fBm and MMPP traffic as follows:

\[
\frac{\exp\left(-\frac{1}{2}\Theta(t_x)\right)}{\sqrt{2\pi\left(1+\sqrt{\Theta(t_x)}\right)^2}} \leq P(Q > x) \leq \exp\left(-\frac{1}{2}\Theta(t_x)\right),
\]

where \( \Theta(t) \) is referred to as the determinative function of the queue length distribution and is given by

\[
\Theta(t) = \frac{\left(-x + \left(C - \sum_{k=1}^{3} \lambda_k t\right) \right)^2}{3\sum_{k=1}^{3} v_k(t)}. \tag{11}
\]

Parameter \( t_x \) \((< 0)\) minimizes \( \Theta(t) \), i.e., \( t_x = \arg \min_s \Theta(s) \). Function \( v_1(t) \) is given by Equation (8), and functions \( v_2(t) \) and \( v_3(t) \) by Equation (3).

Observing Equation (10), we can find that the difference between the upper and lower bounds of the total queue length distribution is the coefficient of \( \exp\left(-\frac{1}{2}\Theta(t_x)\right) \). This fact motivates us to take a certain mean (e.g., arithmetic mean, geometric mean) of the upper and lower bounds to approximate the queue length distribution. In this paper, we adopt the geometric mean that has been proven effective in. As a result, the total queue length distribution can be given by

\[
P(Q > x) \approx \frac{\exp\left(-\frac{1}{2}\Theta(t_x)\right)}{\sqrt{2\pi\left(1+\sqrt{\Theta(t_x)}\right)^2}}. \tag{12}
\]

### IV. Flow Decomposition of PQ-GPS Systems

Generally speaking, modelling individual traffic flows in the PQ-GPS system is considerably complex and challenging due to their heterogeneity and interdependent relationships. In this section, we develop an efficient flow decomposition approach that can equivalently divide the complex PQ-GPS system into individual SSSQ systems. As a result, the task of performance modelling and analysis of the original PQ-GPS system can be transformed to analyzing these equivalent but relatively simple SSSQ systems. The flow decomposition of the PQ-GPS system can be carried out in two steps. Firstly, the PQ-GPS system is decomposed into an SSSQ system with MMPP\(_1\) and a GPS system subject to fBm\(_2\) and fBm\(_3\). Secondly, the GPS system is further decomposed into two SSSQ systems with fBm\(_2\) and fBm\(_3\) traffic flows, respectively. The key issue of flow decomposition is to derive the service capacities, \( c_k \) \((1 \leq k \leq 3)\), of the \( k^{th} \) SSSQ system, denoted by SSSQ\(_k\), and \( c_{25} \) of the GPS system.

#### A. Decomposition of the Integrated Scheduling System into PQ and GPS Systems

As MMPP\(_1\) is served with the strict high priority, it is handled in a manner as if the low-level GPS subsystem does not exist. Therefore, we can readily obtain the service capacity of SSSQ\(_1\) as \( c_1 = C \) by assigning the total system capacity to MMPP\(_1\) traffic flow.

The derivation of the service capacity, \( c_{25} \), of the GPS system is mainly based on the well-known Empty Buffer Approximation (EBA), which has been widely adopted to model the queue length of the low priority traffic in two-class PQ systems [3]. In this PQ system, since the traffic flow with high priority can be served in time, the total queue is almost exclusively composed of packets from the low priority traffic. Therefore, EBA suggests that the total queue length of the PQ system can be used to approximate reasonably the queue length distribution of its low priority traffic flow. In the PQ-GPS system, as compared to the high priority MMPP\(_1\), the aggregate traffic flow of fBm\(_2\) and fBm\(_3\) in the GPS system is served with low priority. Therefore, the total queue length distribution of the PQ-GPS system can be reasonably utilized to approximate that of its low priority GPS subsystem.

**Proposition 1:** The service capacity, \( c_{25} \), of the GPS system decomposed from the original PQ-GPS system can be calculated approximately by solving the following equation:

\[
\Theta(t_x) = \Phi(s_x), \tag{13}
\]

where \( t_x = \arg \min \Theta(t) \),

\[
\Theta(t) = \frac{\left(-x + (C - \sum_{k=1}^{3} \lambda_k t) \right)^2}{\sum_{k=1}^{3} v_k(t)}, \tag{14}
\]

and \( s_x = \arg \min \Phi(s) \),

\[
\Phi(s) = \frac{\left(-x + (c_{25} - \sum_{k=2}^{3} \lambda_k s) \right)^2}{\sum_{k=2}^{3} v_k(t)}, \tag{15}
\]

**Proof:** According to Equations (11) and (12), the total queue length distribution of the PQ-GPS system is determined by the minimum value of the determinative function given in Equation (14). Upon solving the differentiation equation, \( \Theta'(t) = 0 \), and substituting the root, denoted as \( t_x \), into Equation (14), we obtain the minimum value of \( \Theta(t) \) as \( \min \Theta(t_x) = \Theta(t_x) \).

On the other hand, the total queue length distribution of the GPS system decomposed from the original system can be obtained, similar to Equations (11) and (12), by setting the number of traffic flows fed into the queuing system to
be 2 and $C = c_{23}$. It is determined by the minimum value of the determinative function, as given in Equation (15). By differentiating Equation (15), solving $\Phi'(s) = 0$, and further substituting the root, $s_x$ ($<0$), into Equation (15), we get the minimum value of $\Phi(s)$ as $\min_x \Phi(s) = \Phi(s_x)$. 

According to the EBA, the total queue length of the original PQ-GPS system can be used to approximate that of its GPS subsystem. Meanwhile, due to the equivalent decomposition nature, the queue length distribution of the decomposed GPS system should have exactly the same queue length distribution as that of the GPS subsystem in the original PQ-GPS system. Therefore, since the minimum values of functions $\Theta(t)$ and $\Phi(s)$ uniquely determine the queue length distribution of the PQ-GPS and GPS systems, $\Theta(t)$ and $\Phi(s)$ should be approximately equal to each other. As a consequence, we have $\min_t \Theta(t) = \min_x \Phi(s)$, i.e., $\Theta(t_x) = \Phi(s_x)$. By solving this equation, we can finally obtain the service capacity, $c_{23}$, of the GPS system.

B. Decomposition of GPS Systems into SSSQ Subsystems

In the GPS system, the guaranteed service rates received by $fBm_2$ and $fBm_3$ can be given by $\mu_2c_{23}$ and $\mu_3c_{23}$, respectively. Let $e_k = \mu_kc_{23} - \lambda_k (k = 2, 3)$. If $e_k \geq 0$, it denotes the excess service of $fBm_k$. Otherwise, it means that $fBm_k$ is in need of additional service and thus $|e_k|$ denotes its service deficit. In order to obtain the service capacities of $SSSQ_k$ ($k = 2, 3$), without loss of generality we need to consider two situations with different combinations of $e_i$ and $e_j (i, j \in \{2, 3\}$ and $i \neq j$). Note that in the stable GPS systems $e_i + e_j > 0$.

a) Situation I ($e_i < 0$ and $e_j > 0$): Under this situation, $fBm_j$ is guaranteed excess service while $fBm_i$ is frequently in need of additional service. Consequently, $fBm_i$ cannot be served timely and becomes the dominating contributor of the aggregate queue of the GPS system. On the other hand, $fBm_j$ makes minor contribution to the aggregate queue because its arrivals can be handled in time. Owing to such a difference of $fBm_i$ and $fBm_j$, the aggregate queue length distribution of the GPS system can be used to approximate that of $fBm_i$.

Proposition 2: The service capacity, $c_i$, of $SSSQ_i$ can be calculated approximately by solving the following equation:

$$\Phi(s_x) = \Upsilon(x),$$

where $s_x = \arg \min_x \Phi(s)$ and $\Phi(s)$ is given in Equation (15).

$$\Upsilon(x) = \frac{\left(\frac{m-\frac{x}{m}}{m} - 1\right)^2}{a_i\lambda_i} \frac{H_i x}{(\epsilon_i - \lambda_i)(H_i - 1)}^{2H_i}.$$  

Proof: As $SSSQ_i$ has only one traffic flow, its queue length distribution can be obtained, similar to Equations (11) and (12), by setting the number of traffic flow to be one and $C = e_i$. Similarly, the queue length distribution of $SSSQ_i$ is determined by the minimum value of the determinative function which can be denoted as

$$\Gamma(u) = \frac{(-x + (\epsilon_i - \lambda_i)u)^2}{a_i\lambda_i u^{2H_i}}.$$  

Differentiating this equation and solving $\Gamma'(u) = 0$, we get the required value $u_x$ that minimizes $\Gamma(u)$ as follows:

$$u_x = \frac{H_i x}{(\epsilon_i - \lambda_i)(H_i - 1)}.$$  

Upon substituting $u$ in Equation (18) by $u_x$, we can express the minimum value of $\Gamma(u)$ as $\Upsilon(x)$ that is given in Equation (17).

Since the aggregate queue length distribution of the GPS system can be reasonably used to approximate that of $fBm_i$, and the minimum values of functions $\Phi(s)$ and $\Gamma(u)$ uniquely determine the queue length distribution of the GPS system and $SSSQ_i$, respectively, the minimum values of functions $\Phi(s)$ and $\Gamma(u)$ should be approximately equal to each other. Consequently, we get Equation (16). By solving this equation, we can obtain the service capacity $c_i$ of $SSSQ_i$. 

As the proposed flow decomposition approach aims to divide the integrated scheduling system into a group of equivalent SSSQ systems and ensure that, for any given queue length $x$ ($0 \leq x < \infty$), the queue length distribution of the resulting SSSQ systems is the same as that of their corresponding queuing subsystems in the original PQ-GPS system, $c_i$ is usually a function of queue length $x$ and it is difficult to derive a closed-form expression of $c_i$ under heterogeneous fBm traffic with different Hurst parameters. However, if both $fBm_i$ and $fBm_j$ have the same Hurst parameter, we can derive a closed-form expression, which has significantly low computational complexity. Specifically, substituting $H_i = H_j = H$ into Equations (15) and (18), we obtain the service capacity $c_i$ of $SSSQ_i$ as follows:

$$c_i = \lambda_i + \left(c_{23} - \sum_{k \in \{i,j\}} \lambda_k \right) \left(\frac{a_i\lambda_i}{\sum_{k \in \{i,j\}} a_k\lambda_k}\right)^{\frac{1}{2H_i}}.$$  

In this equation, we can rewrite terms $\lambda_i$ and $(c_{23} - \sum_{k \in \{i,j\}} \lambda_k)$ as $\mu_i c_{23} + |e_i|$ and $e_j - |e_i|$, respectively. We can then find that the additional service obtained by $fBm_i$ from the excess service of $fBm_j$ consists of two parts. Firstly, $fBm_j$ takes a portion of the excess service of $fBm_i$ to completely supplement its service deficit, i.e., $|e_i|$. Secondly, the residual excess service, i.e., $e_j - |e_i|$, is shared by $fBm_i$ and $fBm_j$ in proportion of $a_i\lambda_i$.

Under this situation, $fBm_j$ is guaranteed excess service and cannot obtain additional service from $fBm_i$. As a result, $fBm_j$ is served in a manner as if it is handled in a separated system with its guaranteed service rate as the service capacity. Therefore, the service capacity of $SSSQ_j$ is $c_j = \mu_j c_{23}$.

b) Situation II: $e_i \geq 0$ and $e_j \geq 0$: Under this situation, both traffic flows are guaranteed excess services. However, due to the stochastic and bursty nature of LRD traffic, there exist some time intervals, during which one flow has temporarily excess service, while the other is in need of additional service. As a result, during these time intervals $fBm_i$ shares its excess service with $fBm_j$, or vice versa. Therefore, since both traffic flows have excess service, following the analysis in Situation I regarding excess service, the service capacities $c_i$ and $c_j$ can be calculated as follows:

$$c_i = \mu_i c_{23} + (\mu_j c_{23} - \lambda_j) \left(\frac{a_i\lambda_i}{\sum_{k \in \{i,j\}} a_k\lambda_k}\right)^{\frac{1}{2H_i}}.$$  


and 
\[ c_j = \mu_j c_{23} + (\mu_i c_{23} - \lambda_i) \left( \frac{a_j \lambda_j}{\sum_{k \in \{i,j\}} a_k \lambda_k} \right)^{1/\gamma}. \]  

(22)

V. QUEUE LENGTH DISTRIBUTION AND LOSS PROBABILITY OF TRAFFIC FLOWS

In this section, we derive the queue length distribution, \( P(Q_k > x) \) (1 \( \leq \) k \( \leq \) 3), and loss probability, \( L_k(x) \), of MMPP and fBm traffic flows, based on the SSSQ systems obtained in Section IV.

Upon obtaining the service capacities, \( c_k \) (1 \( \leq \) k \( \leq \) 3), of the individual SSSQ systems, the analytical modelling and performance analysis of the original complex PQ-GPS system can be transferred to analyzing these equivalent but relatively simple SSSQ systems. More specifically, given the service capacity, \( c_k \), of SSSQ\( _k \) (1 \( \leq \) k \( \leq \) 3) as well as the equivalent relationship between the \( k^{th} \) traffic flow and SSSQ\( _k \), we can finally obtain the queue length distribution, \( P(Q_k > x) \), based on Equations (11) and (12) by setting \( N = 1 \) and \( C = c_k \) as follows:

\[ P(Q_k > x) \approx \frac{\exp \left( -\frac{1}{2} \Theta_k(t_x) \right)}{\sqrt{2\pi} \left( 1 + \sqrt{\Theta_k(t_x)} \right)^2}, \]  

(23)

where

\[ \Theta_k(t) = \frac{(-x + (c_k - \lambda_k)t)^2}{v_k(t)}. \]  

(24)

In what follows, we derive the loss probability, \( L_k(x) \) (1 \( \leq \) k \( \leq \) 3), of MMPP and fBm traffic flows when the buffer size is \( x \), based on the relationship between the loss probability and queue length distribution of SSSQ systems presented in. Given that each SSSQ system is stable (i.e., \( \lambda_k < c_k \)), the relationship between its loss probability and queue length distribution can be described as

\[ \frac{L_k(x)}{P(Q_k > x)} = \frac{L_k(b)}{P(Q_k > b)}, \]  

(25)

where \( b \) is an arbitrary non-negative constant. Let \( \beta_k = L_k(b)/P(Q_k > b) \). Equation (25) can be rewritten as \( L_k(x) = \beta_k P(Q_k > x) \). In order to obtain the loss probability of traffic flows MMPP\( _1 \), fBm\( _2 \) and fBm\( _3 \), we need to calculate \( \beta_k \) (1 \( \leq \) k \( \leq \) 3), which is directly based on the work of Ref. [20].

Specifically, for a Gaussian traffic flow A = \{A(t)\}\( _{t \in \mathbb{N}} \) with mean arrival rate \( \lambda \) and variance of increments \( \sigma^2 \) (i.e., \( \text{Var}(A(t + 1) - A(t)) = \sigma^2 \)), if setting \( b = 0 \), \( \beta \) can be calculated as follows [20]:

\[ \beta = \frac{1}{\lambda \sqrt{2\pi} \sigma} \exp \left( \frac{(c - \lambda)^2}{2\sigma^2} \right) \int_{c}^{\infty} (y - c) \exp \left( -\frac{(y - \lambda)^2}{2\sigma^2} \right) dy, \]  

(26)

where \( c \) is the service rate obtained by the Gaussian traffic flow. By deriving the definite integral of its last term, Equation (26) can be simplified and expressed as

\[ \beta = \frac{\sigma^2}{\sqrt{2\pi} \lambda (c - \lambda + \sigma)}. \]  

(27)

As discussed in Section III, traffic flow MMPP\( _1 \) can be approximated as a Gaussian process with mean arrival rate \( \lambda_1 \) and variance of increments

\[ \sigma_1^2 = \frac{r_1 \delta_2 + r_2 \delta_1}{\delta_1 + \delta_2} + 2 \left( \frac{r_1 - r_2}{\delta_1 + \delta_2} \right)^2 \left( \frac{1}{\delta_1 + \delta_2} \right)^\delta_2 \]  

(28)

Given it is served with rate \( \lambda_1 \), \( \beta_1 \) is given by

\[ \beta_1 = \frac{\sigma_1^2}{\sqrt{2\pi} \lambda_1 \lambda_1 (1 - \lambda_1 + \epsilon_1)}. \]  

(29)

Since fBm\( _i \) (\( i = 2, 3 \)) is essentially a Gaussian process with mean arrival rate \( \lambda_i \) and variance of increments \( a_i \lambda_i \), and served with rate \( c_i \), we can derive \( \beta_i \) (\( i = 2, 3 \)) by substituting of \( c = c_i \), \( \lambda = \lambda_i \), and \( \sigma = \sqrt{a_i \lambda_i} \) into Equation (27) as follows:

\[ \beta_i = \frac{a_i}{\sqrt{2\pi} (c_i - \lambda_i + \sqrt{a_i \lambda_i})}. \]  

(30)

VI. MODEL VALIDATION

In this section, the accuracy of the proposed model is validated via thorough comparison between the analytical performance results and those obtained from simulation experiments. To this end, we have developed a simulator for the PQ-GPS system using the C++ programming language. In order to reduce the computational complexity and simulation time, the conditionalized Random Midpoint Displacement algorithm (RMD\( _{3,3} \)) was adopted to generate fBm traffic traces in the simulation experiments.

Extensive simulation experiments have been conducted under various scenarios and consistent conclusions have been reached. Due to space limitation, in what follows we present the results of four typical scenarios (i.e., Scenarios A, B, C, and D) corresponding to representative parameter settings of MMPP and fBm traffic flows under different working conditions, as shown in Table I. Specifically, in Scenarios A and B, fBm\( _2 \) and fBm\( _3 \) dominate the input of the PQ-GPS system, while in Scenarios C and D MMPP\( _1 \) dominates the system input. In order to cover both situations being classified in Section IV-B for the model derivation, the parameters of fBm\( _2 \) and fBm\( _3 \) are set such that Scenarios A, B and D fall into Situation I, whilst Scenario C belongs to Situation II. Note that Reference [21] presents a lower bound for the queue length distribution of SSSQ systems subject to fBm traffic, which has been broadly adopted in the subsequent open literature. In order to investigate the precision gain of the proposed model, we compare the analytical queue length distribution of SSSQ systems derived in this paper to the lower bound obtained using the formulae in Reference [21] in the four typical scenarios.

The analytical and simulation results of the queue length distribution and loss probability of individual traffic flows in these scenarios are depicted in Figs. 2–5. In these figures, the solid, dashed, and dash-dotted thick curves represent the analytical results of MMPP\( _1 \), fBm\( _2 \), and fBm\( _3 \), while the corresponding thin curves show the lower bound of their queue length distribution. Their simulation results are denoted by signs ‘ o ’, ‘ ∗ ’, and ‘ + ’, for MMPP\( _1 \), fBm\( _2 \), and fBm\( _3 \) traffic flows, respectively.
A. Scenario A

This scenario intends to investigate the performance of the proposed model under the condition that the combination of $fBm_2$ and $fBm_3$ dominates the input of the integrated system and the settings of their mean arrival rate and weights make them falling into Situation I, as discussed in Section IV-B.

Figure 2(a) depicts the analytical and simulation results of the queue length distribution of individual traffic flows in this scenario, while Fig. 2(b) presents the results of their loss probability. Figure 2(a) shows that the simulation results of both traffic flows $fBm_2$ and $fBm_3$ well match their corresponding analytical queue length distribution obtained from the model derived in this paper. In addition, as compared to both the analytical and simulated results, the lower bounds of $fBm_2$ and $fBm_3$ obtained using the formulae in Reference [21] are slightly loose. Furthermore, Fig. 2(b) regarding the loss probability distribution reveals that the analytical results of both traffic flows $fBm_2$ and $fBm_3$ obtained in this paper closely match the corresponding simulation results. These observations indicate that the developed analytical model possesses a good degree of accuracy in predicting both queue length and packet loss.

In this scenario, as compared to the service capacity of the PQ-GPS system the mean arrival rate of traffic flow $MMPP_1$ is relatively small (specifically, 40 vs. 200). Meanwhile, in the integrated queuing system $MMPP_1$ is handled with the strict high priority and is served in a manner as if both $fBm_2$ and $fBm_3$ do not exist. Therefore, $MMPP_1$ can be served promptly: Packets from $MMPP_1$ do not need to wait for service in buffer and no packets of $MMPP_1$ are lost due to buffer overflow. As a result, no curves that depict the queue length distribution and loss probability of traffic flow $MMPP_1$ are presented in Fig. 2.

B. Scenario B

This scenario aims to examine the performance of the developed model in a scene where the states of $MMPP_1$ have significantly different transition rates, showing a high degree of traffic burstiness. In addition, traffic flows $fBm_2$ and $fBm_3$ have the same weights but their different mean arrival rates lead them to belong to Situation I in Section IV-B. Moreover, traffic flows $fBm_2$ and $fBm_3$ dominate the input of the integrated system.

Figures 3(a) and (b) present the analytical and simulation results of the queue length distribution and loss probability of individual traffic flows in this scenario. The figures reveal that the curves depicting the analytical results of both $fBm_2$ and $fBm_3$ are in close agreement with the corresponding simulation results of queue length distribution and loss probability of individual traffic flows. The above observations reveals that the model proposed in this study performs well in characterizing the QoS metrics of different traffic flows in the integrated scheduling system.

The same as in Scenario A, since traffic flow $MMPP_1$ has a remarkably small mean arrival rate as compared to the service capacity of the PQ-GPS system and it is served with strict high priority, no packets of $MMPP_1$ need to be buffered in the queue and hence no curve of traffic flow $MMPP_1$ is shown in Fig. 3.

C. Scenario C

Unlike the above two scenarios where traffic flows $fBm_2$ and $fBm_3$ dominate the input of the integrated queuing system, in what follows we will investigate the performance of the developed analytical model in Scenarios C and D where the system input is dominated by traffic flow $MMPP_1$.  

### Table I

<table>
<thead>
<tr>
<th>Scenario</th>
<th>$C$</th>
<th>$MMPP_1$</th>
<th>$fBm_2$</th>
<th>$fBm_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>200</td>
<td>0.3</td>
<td>0.6</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2</td>
<td>0.7</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.6</td>
<td>0.7</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>0.7</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>300</td>
<td>200</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.8</td>
<td>0.7</td>
</tr>
</tbody>
</table>

### Figures

- Fig. 2. The comparison between the analytical and simulation results of the queue length distribution and loss probability in Scenario A.
- Fig. 3. The comparison between the analytical and simulation results of the queue length distribution and loss probability in Scenario B.
Figures 4(a) and (b) present the performance results of the queue length distribution and loss probability, respectively, of individual traffic flows in Scenario C, which falls in Situation II of Section IV-B. Specifically, as shown in Table I, the mean arrival rate of MMPP1 in this scenario is 240, which is relatively large as compared to the service capacity 300 of the integrated system. As a result, the curves and signs of the analytical and simulation results corresponding to both queue length distribution and loss probability of MMPP1 are present in Fig. 4. However, since it is served with the strict high priority in the integrated queuing system, the queue length and loss probability of MMPP1 are much less than those of \( fBm_2 \) and \( fBm_3 \) although its mean arrival rate is much larger than those of \( fBm_2 \) and \( fBm_3 \). For example, from Fig. 4 it can be seen that if the buffer size is set as \( x = 20 \), the loss probability of MMPP1 is around \( 10^{-7} \) whilst that of \( fBm_2 \) and \( fBm_3 \) is higher than \( 10^{-3} \).

Figure 4 reveals that the signs representing the simulation results of MMPP and fBm traffic flows in Scenario C closely match the curves of the corresponding analytical results. This observation demonstrates again the accuracy of the analytical model. Regarding the lower bounds obtained using the formulæ in Reference [21], we can find that the bounds of traffic flows \( fBm_2 \) and \( fBm_3 \) are also loose. However, the lower bound of traffic flow MMPP1 is relatively tight compared to the cases of \( fBm_2 \) and \( fBm_3 \). This is mainly due to two reasons: First, since traffic flow MMPP1 is served with the strict high priority, it can build up a short queue only in its buffer with a very small probability in comparison with traffic flows \( fBm_2 \) and \( fBm_3 \). Therefore, when the results of all three traffic flows are depicted in the same figure, the gap between the lower bound and analytical queue length distribution of MMPP1 is very limited as compared to those corresponding to \( fBm_2 \) and \( fBm_3 \), respectively. Second, MMPP1 in the proposed model is approximated as a Gaussian traffic flow, which slightly underestimates the traffic burstiness and hence leads to a little bit small service capacity identified for the SSSQ system subject to MMPP1. Consequently, we can see in Fig. 4(a) that the results of the analytical queue length distribution of MMPP1 are slightly above the simulated results.

**D. Scenario D**

This scenario aims to investigate the performance behavior of the developed analytical model under a scene where traffic flow MMPP1 dominates the input of the integrated scheduling system and its two states have identical arrival rates, namely, \( r_1 = r_2 \). That is to say, in this scene traffic flow MMPP1 degenerates to a traditional Poisson arrival process with the mean arrival rate 1420.

Figure 5 depicts the analytical and simulation results of Scenario D, where the queue length distribution and loss probability of all traffic flows are presented. We can find again that the simulation results of MMPP and fBm traffic flows are in good agreement with the corresponding analytical results of queue length distribution and loss probability. Based on the comparison of the four representative scenarios, we can conclude that the analytical model developed in this paper performs very well in obtaining the queue length and packet loss distribution of individual traffic flows of the PQ-GPS system subject to heterogeneous SRD and LRD traffic.

**VII. PERFORMANCE ANALYSIS**

In this section, the developed model is used to investigate the important issues of resource management (e.g., bandwidth allocation) and call admission control in wireless networks with the PQ-GPS mechanism and QoS constraints.

**A. Bandwidth Allocation**

For the sake of illustration, we consider a PQ-GPS system with service capacity \( C = 450 \). The buffer size assigned to traffic flows, MMPP1, \( fBm_2 \), and \( fBm_3 \) is 5, 350, and 250, respectively. The parameter settings of three traffic flows are listed in Table I (see Scenario E for the details). Given that the QoS constraints in terms of loss probability of MMPP1, \( fBm_2 \), and \( fBm_3 \) are \( L_1(5) < 10^{-8} \), \( L_2(350) < 10^{-6} \), and \( L_3(250) < 10^{-5} \) we aim to use the analytical model to investigate the proper setting of the weights, \( \mu_2 \) and \( \mu_3 \), of \( fBm_2 \) and \( fBm_3 \).

For MMPP1, the analytical model reveals that its loss probability is \( L_1(5) = 1.92 \times 10^{-11} \). That is to say, its requirement on loss probability, \( L_1(5) < 10^{-8} \), can be well guaranteed with the given buffer of size 5. Furthermore, we use the analytical model to calculate the loss probability of \( fBm_2 \) and \( fBm_3 \) under different combinations of their weights. For \( fBm_2 \), we can obtain

- \( L_2(350) \geq 2.93 \times 10^{-6} \) when \( \mu_2 \leq 0.32 \);
- \( L_2(350) \leq 5.00 \times 10^{-7} \) when \( \mu_2 \geq 0.33 \).

Similarly, for \( fBm_3 \), we have
• \( L_3(250) \geq 1.02 \times 10^{-5} \) when \( \mu_3 \leq 0.60 \);
• \( L_3(250) \leq 7.57 \times 10^{-6} \) when \( \mu_3 \geq 0.61 \).

Therefore, given that \( \mu_2 + \mu_3 = 1.0 \), the intervals \( 0.33 \leq \mu_2 \leq 0.39 \) and \( 0.61 \leq \mu_3 \leq 0.67 \) construct an admissible region of \( \mu_2 \) and \( \mu_3 \). All combinations of different values of \( \mu_2 \) and \( \mu_3 \) in this region can guarantee the above requirements of loss probability of \( fBm_2 \) and \( fBm_3 \).

B. Call Admission Control

The call admission control concerns whether to admit an incoming call into the scheduling system on the condition that the QoS requirements of all existing traffic flows are still satisfied. According to the nature (i.e., SRD or LRD) that the QoS requirements of all existing traffic flows are, the call admission control concerns whether to admit an incoming call can be readily calculated as \( \Omega = \Omega \oplus \Omega \) and \( \Lambda = \Lambda \oplus \Lambda \), where \( \oplus \) denotes the Kronecker sum [9]. In the other case, the traffic of the incoming wireless call falls into the low priority class of fBm. Let \( \bar{\lambda}_i \), \( \bar{a}_i \), \( \bar{H}_i \) represent its mean arrival rate, variance coefficient, and Hurst parameter, respectively. As the superposition of fBm traffic is still fBm [25], the mean arrival rate, \( \lambda_i^t \), variance coefficient, \( a_i^t \), Hurst parameter, \( H_i^t \), and variance function, \( v_i(t) \), of the aggregate fBm traffic flow, denoted as \( fBm_i^t \), can be given by

\[
\begin{align*}
\lambda_i^t &= \lambda_i + \lambda_i, \\
a_i^t &= (a_i + a_i)/(\lambda_i + \lambda_i), \\
H_i^t &= \text{max}(H_i, H_i), \\
v_i(t) &= \bar{a}_i\lambda_i t^{2H_i} + \bar{a}_i\lambda_i t^{2H_i}.
\end{align*}
\]

In what follows, we will illustrate the application of the developed model to call admission control in Scenario E with the parameter settings of three traffic flows listed in Table I. The buffer sizes allocated to the three traffic flows and their QoS requirements in terms of loss probability are the same as those in Section VII-A. Let the weights assigned to \( fBm_2 \) and \( fBm_3 \) be \( \mu_2 = 0.35 \) and \( \mu_3 = 0.65 \), respectively. Under such a setting, the loss probability of \( M M P P_3 \), \( fBm_2 \), and \( fBm_3 \) obtained from the model is \( L_1(5) = 1.92 \times 10^{-11} \), \( L_2(350) = 1.04 \times 10^{-8} \), and \( L_3(250) = 2.21 \times 10^{-6} \), respectively.

Case 2: In this case, given that the incoming wireless call belong to the low priority class of \( fBm_2 \) and its mean arrival rate, variance coefficient, and Hurst parameter are \( \lambda_2 = 4, a_2 = 0.6, \) and \( H_2 = 0.8 \), respectively. If admitting this call into the PQ-GPS system, the parameters for characterizing the aggregate fBm traffic flow, \( fBm_2^t \), become \( \lambda_2^t = 39, a_2^t = 0.6, \) and \( H_2^t = 0.8 \). Since \( M M P P_2 \) is served with the strict high priority, it is not affected by the variation of low priority traffic flows. Therefore, its loss probability remains unchanged. For \( fBm_2^t \) and \( fBm_3 \), using the developed model we can obtain their loss probability as \( L_2(350) = 7.70 \times 10^{-6} \) and \( L_3(250) = 4.10 \times 10^{-6} \), respectively. In other words, if the incoming wireless call is admitted, the QoS requirement of \( fBm_2^t \) cannot be satisfied under the present setting of \( \mu_2 \) and \( \mu_3 \). With the aim of admitting the incoming call while keeping the QoS constraints guaranteed, we use the model to further explore other feasible combinations of the weights, \( \mu_2 \) and \( \mu_3 \). As a result, we find that if the weights of \( fBm_2^t \) and \( fBm_3 \) are changed to \( \mu_2 = 0.37 \) and \( \mu_3 = 0.63 \), the loss probability become \( L_2(350) = 3.07 \times 10^{-7} \) and \( L_3(250) = 8.07 \times 10^{-6} \), respectively. That is to say, using the developed analytical model to tune the weights corresponding to the bandwidth allocation, the incoming call can be admitted to increase the resource utilization while the loss probability requirements remain to be satisfied.

VIII. CONCLUSIONS

Traffic scheduling plays an important role in provisioning of differentiated QoS in wireless communication networks. This paper has proposed an original analytical model for the promising integrated PQ-GPS scheduling scheme in the presence of heterogeneous LRD fBm and SRD MMPP traffic. We have derived the expression for calculating the total queue length distribution of the integrated scheduling system and then developed an efficient flow decomposition approach that can equivalently divide the PQ-GPS system into a group of SSSQ systems. We have further derived both queue length distribution and loss probability of individual traffic flows. The validity and accuracy of the model make it a practical and cost-effective tool for investigating the performance behavior of the PQ-GPS system under heterogeneous traffic. To illustrate its applications, the model has been adopted to study the bandwidth allocation problem with specific QoS requirements of individual traffic flows. The results have shown that the model can examine whether or not an incoming wireless call can be admitted into the system. If not, it can be further utilized to investigate whether the new call can be admitted by properly adjusting the bandwidth allocated to existing traffic flows.

REFERENCES


