Delay-constrained scheduling: power efficiency, filter design, and bounds

— Source link ⌕

M.A. Khojastepour, Ashutosh Sabharwal

Institutions: Rice University

Related papers:

- Communication over fading channels with delay constraints

- A calculus approach to minimum energy transmission policies with quality of service guarantees

- Energy-efficient packet transmission over a wireless link

- Optimal energy allocation for delay-constrained data transmission over a time-varying channel

- Delay-bounded packet scheduling of bursty traffic over wireless channels

Delay-constrained Scheduling: Power Efficiency, Filter Design, and Bounds

Mohammad Ali Khojastepour and Ashutosh Sabharwal*

Department of Electrical and Computer Engineering, Rice University

6100 Main St., MS-366, Houston, TX 77005

e-mail: {amir,ashu}@rice.edu

Abstract— In this paper, packet scheduling with maximum delay constraints is considered with the objective to minimize average transmit power over Gaussian channels. The main emphasis is on deriving robust schedulers which do not rely on the knowledge of the source arrival process. Towards that end, we first show that all schedulers (robust or otherwise) which guarantee a maximum queuing delay for each packet are equivalent to a time-varying linear filter. Using the connection between filtering and scheduling, we study the design of optimal power minimizing robust schedulers. Two cases, motivated by filtering connection, are studied in detail. First, a time-invariant robust scheduler is presented and its performance is completely characterized. Second, we present the optimal time-varying robust scheduler, and show that it has a very intuitive time water-filling structure. We also present upper and lower bounds on the performance of power-minimizing schedulers as a function of delay constraints. The new results form an important step towards understanding of the packet time-scale interactions between physical layer metric of power and network layer metric of delay.

I. INTRODUCTION

Source burstiness is common in many multimedia sources, which generate variable number of useful information bits per unit time, and can be used to trade average queuing delay with average transmission power [1, 2, 9, 13]. In this paper, we study the relation between average transmission power and strict delay constraints.

The motivation for this work stems from the fact that exact channel and source probability distributions are seldom known in practice. Thus, optimal schedulers designed based on an assumed knowledge of probability distributions typically incur a significant loss in the presence of distribution mismatch. Hence, seeking robust schedulers, we investigate the performance achievable with no prior assumptions about the probability distribution of the source.

In [3], we established necessary and sufficient conditions on the service rates of the wireless transmitter, needed to meet the delay deadline of every packet in the queue. Based on the derived conditions, we had demonstrated that any scheduler can be expressed as a linear time-varying low-pass filter if it meets a delay guarantee $D_{\max}$ for every packet over Gaussian channels. The result provides a proof for the intuitive explanation for power reduction due to additional queuing delay, first observed in [2].

Our main contributions are two-fold. First, the relation between linear filtering and delay bounded scheduling is used to construct robust schedulers, which do not rely on the knowledge of source statistics. Non-reliance on distributional assumptions makes our results significantly different from the common information theoretic formulations on power-efficient scheduling [1, 2]. We construct both the optimal time-varying and time-invariant robust schedulers. The optimal time-invariant robust scheduler turns out to be simply a moving average filter of length $D_{\max}$. Similarly, the time-varying scheduler has an elegant structure and water-fills the newly arrived packets over the next $D_{\max}$ time-slots.

Second, the performance of optimal scheduler (robust or otherwise) is characterized via an upper and a lower bound. The upper bound is derived using Jensen's inequality and the lower bound uses Hardy-Littlewood-Polya inequality [6]. The upper bound also characterizes the performance of the optimal time-invariant robust scheduler. Lastly, we also derive the exact performance of the optimal time-varying robust scheduler. To the best of our knowledge, these are the first bounds on average power consumption of delay-constrained scheduling, and are most useful for the case of small delays by providing trends of actual performance.

We note that the results in this paper are limited to Gaussian channels, much like the works in [2, 12, 13]. We expect that the connections between filtering and
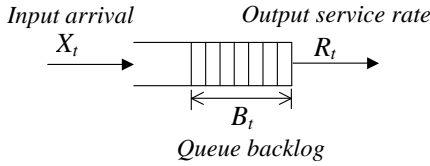
Fig. 1. Conceptual depiction of a queuing system.

scheduling will continue to hold for time-varying fading channels, albeit the nature of filters will change (possibly no longer low-pass).

The rest of the paper is organized as follows. In Section II, we formulate the problem and characterize the set of feasible schedulers. In Section III, the relation between filtering and scheduling is clarified. Some of the results in Sections II and III were presented in [3] without proofs, and hence they are included here for completeness. Motivated by the filtering property of the schedulers, the design of optimal time-invariant schedulers is presented in Section IV. In Section V, the upper and lower bounds on the performance of optimal scheduler are derived. The design of optimal robust schedulers, designed without the knowledge of source statistics, are presented in Section VI. Finally, we briefly survey the related works in Section VII and we conclude in Section VIII.

II. GUARANTEED MAXIMUM DELAY SCHEDULER

A. Problem Formulation

Consider the system in Figure 1, where the number of input packets to the queue at time t is given by a random process X_t. The arriving packets are queued in a buffer whose backlog before time t is denoted by B_t. We assume that all packets have the same size. At any time t, a *causal* scheduler transmits R_t packets out of the queue by looking at the queue size B_t, and having the knowledge of all the previous input arrivals $\{X_i\}_{i=1}^{t}$. On the other hand, a *non-causal* scheduler uses *all* arrivals $\{X_i\}_{i=1}^{\infty}$.

In this paper, we will consider the case when each packet is required to be delivered within the $D_{\max}$ time-slots of its arrival. The scheduler will be designed to adapt the transmission rate R_t such that the average transmit power $\mathbb{E}[P_t]$ is minimized while meeting the maximum queuing delay constraint for each packet. We consider a Gaussian channel between transmitter and receiver and assume that all transmissions occur in packets and the length of the packet is long enough to allow reliable communication close to the mutual information of the channel. This is motivated by the

outage versus capacity analysis in [8], and other recent works in scheduling problem [1, 2]. Solving the mutual information formula for Gaussian channels shows that the average power is exponentially related to the rate, *i.e.* $P_t \propto 2^{R_t}$. By looking at the average power required for N time slot, scheduler design based on two different optimization criteria is defined.

Criterion 1:

$$\min \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} 2^{R_t} \qquad (1)$$

Criterion 2:

$$\min \mathbb{E} \left[\lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} 2^{R_t} \right] \qquad (2)$$

The above optimization problems are not necessarily identical. It can be readily shown that a sufficient condition for the above two minimization problems to be identical is that any given time, t, the scheduler output rate, R_t, is just a function of the past M input arrivals, X_t, X_{t-1}, ..., X_{t-M+1}, for some fixed value of M which does not depend on t.

B. Set of Feasible Schedulers

We first characterize the set of all feasible schedulers by deriving the necessary and sufficient conditions for a scheduler to guarantee the maximum delay of $D_{\max}$, in terms of X_i and R_i for $i = 1, 2, \ldots, t$. The results in this section form the basis for all the subsequent results.

Lemma 1 (Feasible Schedulers): For any scheduler, maximum delay of each packet is less than or equal to $D_{\max}$ if and only if for all positive integer values of t and k, the output service rates R_i and input arrival rates X_i for $i = t, t+1, t+2, \ldots, t+k+D_{\max}$, and queue backlog B_t satisfy the following inequalities:

$$X_t + X_{t+1} + \ldots + X_{t+k} \le R_t + R_{t+1} + \ldots + R_{t+k+D_{\max}-1} \qquad (3)$$

$$R_t + R_{t+1} + \ldots + R_{t+k} \le B_t + X_t + X_{t+1} + \ldots + X_{t+k} \qquad (4)$$

Proof: First we prove the necessity of the conditions. Condition (3) gives a lower bound for the output service rates out of the queue based on the input arrival rates to the queue. It shows that for any t and k the output rate out of the queue from time t to time $t + k + D_{\max} - 1$ should be more than or equal to the input to the queue from the time t to the time $t+k$. Suppose not, thus there exist some values t and k such that the output out of the queue from time t to the time $t + k + D_{\max} - 1$ is less than the input to the queue from time t to the time $t+k$.

Therefore, even if the queue backlog at time t, is zero ($B_t = 0$) the last packet of which has arrived in the time interval of t to $t + k$ would not be serviced until time $t + k + D_{\max}$ which makes the delay more than $D_{\max}$ for this packet. On the other hand, Condition (4) gives an upper bound on the output of the queue based on the input arrivals and previous backlog in the queue. It shows that output in the time interval t to $t+k$ cannot be greater than sum of the input arrivals in the same time period plus the amount of queue backlog at the time t. This is equivalent to the condition that queue backlog cannot be negative.

Next we prove that the mentioned conditions are sufficient as well. For the same, we need to show that if input arrival rates and output service rates satisfy both of the Conditions (3) and (4), then the maximum delay for any packet would not exceed D_{max}. Consider a packet which has arrived at arbitrary time k. Assume that X_i, $i = 1, 2, \ldots, k + D_{\max} - 1$ are arbitrary input arrival rates and the output service rates R_i, $i = 1, 2, \ldots, k + D_{\max} - 1$ have been chosen such that the conditions of the lemma are satisfied. For $t = 1$, from the Condition (3) we have

$$X_1 + X_2 + \ldots + X_{t+k} \le R_1 + R_2 + \ldots + R_{t+k+D_{\max}-1} \tag{5}$$

Since there is no arrival to the queue before time $t = 1$, and the queue initially is empty, the above condition guarantees that all packets which have arrived before time $k + 1$ would be out of the queue before time $t + D_{\max}$, which means that delay for all the packets arrived at time k is at most $D_{\max}$. On the other hand we need to show that the above set of output service rates R_i, $i = 1, 2, \ldots, k + D_{\max} - 1$ do not violate the queue rule, i.e. $B_i \ge 0$, for all $i = 1, 2, \ldots, k + D_{\max} - 1$. By definition

$$B_i = X_1 + X_2 + \ldots + X_i - (R_1 + R_2 + \ldots + R_i) + B_0 \tag{6}$$

From the condition 4, for $t = 1$ and $k = i - 1$ it is immediate that $B_i \ge 0$. ∎

III. Feasible Schedulers are Low-Pass Filters

In this section, we first show that structure of any scheduler which guarantees the maximum delay of $D_{\max}$ is equivalent to a linear time-varying filter of size no more than $D_{\max}$. The characterization of the finite impulse response filter immediately leads to the fact that it is always a low-pass filter.

Theorem 1 (Filter Characterization): A scheduler which guarantees the maximum delay of $D_{\max}$ for any arrived packet is a linear time variant filter of size $D_{\max}$, denoted by:

$$R_t = \alpha_0^t X_t + \alpha_1^t X_{t-1} + \ldots + \alpha_{D_{\max}-1}^t X_{t-D_{\max}+1}, \tag{7}$$

where the filter coefficients satisfy the following

$$\sum_{i=0}^{D_{\max}-1} \alpha_i^{t+i} = 1, \forall\ t, \tag{8}$$

$$0 \le \alpha_i^t \le 1, \forall\ t, i. \tag{9}$$

Furthermore, any time variant filter of size $D_{\max}$ which satisfies the above constraints is a valid scheduler which guarantees the maximum delay of $D_{\max}$ for any packet.

Proof: First we prove the necessary condition by showing every scheduler which guarantees the maximum delay of $D_{\max}$ can be expressed as a linear time-variant filter. Consider the output of the queue at time t. Since the scheduler guarantees the maximum delay of $D_{\max}$, all of the scheduled packet at time t are packets which have arrived in the past $D_{\max}$ time slots. Assume that the first output packet has the delay of exactly i and the last output packet has the delay of exactly j, where $D_{\max} - 1 \ge i \ge j \ge 0$. Thus, the output rate R_t can be written in the form of the Equation (7) where

(i) α_i^t and α_j^t are the portions of packets which have arrived at the time $t - i$ and $t - j$ and are transmitted at time t, and

(ii) $\alpha_k^t = 1$ for any k, such that $i > k > j$ if it exists.

On the other hand, since packets which have arrived at time t should be out before the time $t + D_{\max}$ and α_i^{t+i} is the portion of it which is out at time $t + i$, the summation of the Equation (8) should hold.

Now we relax the above conditions (i) and (ii) to that of (9) and prove the sufficiency condition which means that any scheduler that has output rate R_t satisfying condition of the Theorem 1 will have delay of no more than $D_{\max}$ time unit for all of the packets. The proof follows immediately by verifying that the Inequalities (3) and (4) of Lemma 1 hold. ∎

Based on Theorem 1, whenever the scheduler is treated as a filter, we will use the term *scheduling filter* as an equivalent term for a scheduler. It is worth mentioning that a scheduler does not have a unique representation as a linear time varying filter. In most cases, there is more

than one possible way of representing a fixed scheduler in the form of a linear time varying filter.

Theorem 1 turns the design of the guaranteed maximum delay scheduler into the problem of filter design with a 'linear' structure. Therefore, we foresee the vast literature on linear filtering theory as a fundamental tool in designing power-efficient schedulers [5, 7]. Following result follows directly from basic filtering theory.

Corollary 1: Every feasible time-invariant scheduler is a low-pass filter.

Proof: Since all of the coefficients of the filter are chosen in the interval $[0, 1]$, all the zeros of the corresponding z-transform of the filter are on the left hand side of the origin which means that the filter is a low-pass filter [11]. ∎

Remark 1 : For power-efficiency, additional delay helps by smoothening the input arrival process via queuing [2]. It is clear that by increasing the number of filter taps (equivalently increasing maximum possible scheduling delay), one can design better low-pass filters, leading to transmit power reduction. The intuition behind the fact that the optimal average power scheduler would have smoother output sequence than the input sequence comes from the convexity of the objective function of the optimization problem in either of Criteria (1) or (2).

Remark 2 : It can be easily proved that the optimal scheduler with guaranteed maximum delay $D_{\max}$ depends only on the distribution of the input arrivals and the past observation of the input arrival rates X_{t-i}, $i = 0, 1, \ldots, D_{\max} - 2$, and the queue backlog B_t in order to determine the optimal value of output service rate, R_t, at time t. In other words, it does not depend on the individual values of X_{t-i} for $i \geq D_{\max} - 1$. We will use this property along with the linear filtering property of the optimal scheduler to find the optimal robust scheduler when the distribution of the input arrival rates is not known. The structure of the robust scheduler even extends to the case in which the packets have different QoS requirements (i.e. maximum delay requirements) while the supremum of all of the delays is still bounded by fixed value $D = D_{max}$.

IV. Optimal Time-invariant Scheduling

Motivated by the filtering property of the scheduler, we consider design of the optimal time-invariant scheduler for a given input arrival distribution. Theorem 2 characterizes the optimal time-invariant scheduler, and it turns out that the optimal solution is independent of the input arrival distribution. In other words, the optimal time-invariant scheduler is robust to the changes of the input arrival distribution as it will be further explored in Section VI. Also, for any input arrival distribution the performance of optimal time invariant scheduler provides an upper bound on the performance of the optimal scheduler (which is time varying in general) that would be discussed in Section V.

Theorem 2 (Time-invariant Scheduler): Let input arrival rates X_t, $t = 0, 1, \ldots$ be i.i.d. random variables with distribution $f_X(x)$. For both cases of the optimization problems posed in Criteria 1 and 2. the optimal time invariant scheduling filter (scheduler) has the form

$$R_t = \frac{X_t + X_{t-1} + \ldots + X_{t-D_{\max}+1}}{D_{\max}} \quad (10)$$

Proof : Assume that the optimal time invariant scheduler for a time index k is given by $R_k = \alpha_0 X_k + \alpha_1 X_{k-1} + \ldots + \alpha_{D_{\max}-1} X_{k-D_{\max}+1}$, where $0 \leq \alpha_i \leq 1$ and $\sum_{i=0}^{D_{\max}-1} \alpha_i = 1$.

First we prove that in the class of time invariant schedulers both of the objective functions are equivalent, and equal to the objective function $J = \mathbb{E}[2^{R_k}]$. The proof is based on the property that for all t, R_t are identically distributed and also R_t and R_{t+M} are independent if $M \geq D_{\max}$. For the first objective function, we have

$$J_1 \triangleq \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} 2^{R_t} = \lim_{L \to \infty} \frac{1}{LD_{\max}} \sum_{t=1}^{LD_{\max}} 2^{R_t} \quad (11)$$

$$J_1 = \frac{1}{D_{\max}} \sum_{i=1}^{D_{\max}} \lim_{L \to \infty} \frac{1}{L} \sum_{t=1}^{L} 2^{R_{i+tD_{\max}}} \quad (12)$$

$$J_1 = \frac{1}{D_{\max}} \sum_{i=1}^{D_{\max}} \mathbb{E}[2^{R_k}] = \mathbb{E}[2^{R_k}], \quad (13)$$

where Equation (13) is derived from the law of large numbers in probability theory. On the other hand, for the second objective function, we have

$$J_2 \triangleq \mathbb{E}\left[\lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} 2^{R_t}\right] \quad (14)$$

$$J_2 = \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} \mathbb{E}[2^{R_k}] = \mathbb{E}[2^{R_k}]. \quad (15)$$

Therefore, the optimization problem can be written as:

$$
\begin{aligned}
\min J &= \min \mathbb{E}[2^{R_k}] \\
&= \min \mathbb{E}[2^{\alpha_0 X_k + \alpha_1 X_{k-1} + \ldots + \alpha_{D_{\max}-1} X_{k-D_{\max}+1}}] \\
&= \min \mathbb{E}[2^{\alpha_0 X_k}] \mathbb{E}[2^{\alpha_1 X_{k-1}}] \ldots \\
&\qquad\qquad \ldots \mathbb{E}[2^{\alpha_{D_{\max}-1} X_{k-D_{\max}+1}}] \\
&= \min \mathbb{E}[2^{\alpha_0 X_k}] \mathbb{E}[2^{\alpha_1 X_k}] \ldots \mathbb{E}[2^{\alpha_{D_{\max}-1} X_k}]
\end{aligned}
\tag{16}
$$

It can be easily shown that $f(\alpha) = \ln \mathbb{E}[2^{\alpha X_k}]$ is a convex function of α for any distribution of X_k. The second derivative of $f(\alpha)$ is given by:

$$
\begin{aligned}
&\frac{\partial^2 f(\alpha)}{\partial \alpha^2} = \\
&\frac{\mathbb{E}[(X_k \ln 2)^2 2^{\alpha X_k}] \mathbb{E}[2^{\alpha X_k}] - (\mathbb{E}[(X_k \ln 2) 2^{\alpha X_k}])^2}{(\mathbb{E}[2^{\alpha X_k}])^2}
\end{aligned}
\tag{17}
$$

The numerator is positive by using the Cauchy-Schwartz inequality in the form $\mathbb{E}[f^2]\mathbb{E}[g^2] \geq \mathbb{E}^2[fg]$, and the denominator is obviously positive. Thus the second derivative of $f(\alpha)$ is always positive and therefore it is a convex function. By applying Jensen's inequality and considering that $\sum_{i=0}^{D_{\max}-1} \alpha_i = 1$, we have

$$
\begin{aligned}
\frac{1}{D_{\max}} \sum_{i=0}^{D_{\max}-1} f(\alpha_i) &\geq f\left(\frac{1}{D_{\max}} \sum_{i=0}^{D_{\max}-1} \alpha_i\right) \\
\sum_{i=0}^{D_{\max}-1} f(\alpha_i) &\geq D_{\max} f\left(\frac{1}{D_{\max}}\right) \\
\sum_{i=0}^{D_{\max}-1} \ln \mathbb{E}[2^{\alpha_i X_{k-i}}] &\geq D_{\max} \ln \mathbb{E}[2^{\frac{X_k}{D_{\max}}}] \\
\prod_{i=0}^{D_{\max}-1} \mathbb{E}[2^{\alpha_i X_{k-i}}] &\geq \mathbb{E}^{D_{\max}}[2^{\frac{X_k}{D_{\max}}}]
\end{aligned}
\tag{18}
$$

Thus, $J \geq \mathbb{E}^{D_{\max}}[2^{\frac{X_k}{D_{\max}}}]$ with equality if and only if $\alpha_i = 1/D_{\max}$ for all $i = 0, 1, \ldots, D_{\max} - 1$. ∎

V. Bounds on Scheduler Performance

An upper and lower bound on the optimal scheduler performance (robust or otherwise) are presented in this section. The upper bound is, in fact, the performance of the robust time-invariant scheduler presented in the previous section. Also, the bounds apply to both optimization Criteria 1 and 2.

Corollary 2 (Upper Bound): An upper bound on the optimal scheduler guaranteeing maximum delay of $D_{\max}$ is given by:

$$
P_{avg} \leq \mathbb{E}^{D_{\max}}[2^{\frac{X}{D_{\max}}}],
\tag{19}
$$

where input arrival rates are i.i.d. random variables with distribution $f_X(x)$.

Proof : Follows directly from the existence of the scheduler in Theorem 2 which requires the power average of $\mathbb{E}^{D_{\max}}[2^{\frac{X}{D_{\max}}}]$. ∎

It is easy to prove that this upper bound is a decreasing function of $D_{\max}$ for any distribution of the input process X_t, $f_X(x)$ (Jensen's inequality). It again shows that as the maximum delay increases, the number of scheduling filter (scheduler) taps $D_{\max}$ increases, and the required average transmit power is lower.

By using Jensen's inequality it is fairly easy to see that $2^{\mathbb{E}[X]}$ is a lower bound on the average power of any scheduler which satisfies conditions of the Lemma 1. In order to establish stronger lower bound, we will use Hardy-Littlewood-Polya inequality [6] which captures the effect of permutation and ordering, and is fundamentally different from the class of inequalities (such as Jensen's inequality) that are based on the property of convex functions.

Theorem 3 (Lower Bound): Let input arrival rates X_t, $t = 0, 1, \ldots$ be i.i.d. random variable with distribution $f_X(x)$. For both cases of the optimization problems posed in Criteria 1 and 2, a lower bound on the required average transmission power for any scheduler which guarantees the maximum delay of $D_{\max}$ is given by:

$$
P_{ave} \geq \mathbb{E}^{l+1}\left[2^{\frac{X}{l+D_{\max}}}\right]
\tag{20}
$$

for any value of parameter l.

Proof: Consider Criterion 1 of the objective function. We have

$$
\begin{aligned}
&\lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} 2^{R_t} \\
&\geq \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \left(2^{\frac{R_t}{l+D_{\max}}}\right)^{l+D_{\max}} \\
&\geq \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} 2^{\frac{R_t + R_{t+1} + \ldots + R_{t+l+D_{\max}-1}}{l+D_{\max}}} \\
&\geq \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} 2^{\frac{X_t + X_{t+1} + \ldots + X_{t+l}}{l+D_{\max}}} \\
&\geq \frac{1}{l+1} \sum_{i=1}^{l+1} \lim_{L\to\infty} \frac{1}{L} \sum_{t=1}^{L} 2^{\frac{X_{t(l+1)+i} + \ldots + X_{t(l+1)+i+l}}{l+D_{\max}}}
\end{aligned}
\tag{21}
$$

$$\geq \quad \frac{1}{l+1} \sum_{i=1}^{l+1} \mathbb{E} \left[2^{\frac{X_{t(l+1)+i}+X_{t(l+1)+i+1}+\dots+X_{t(l+1)+i+l}}{l+D_{\max}}} \right]$$

$$\geq \quad \frac{1}{l+1} \sum_{i=1}^{l+1} \mathbb{E}^{l+1} \left[2^{\frac{X}{l+D_{\max}}} \right] \tag{22}$$

$$\geq \quad \mathbb{E}^{l+1} \left[2^{\frac{X}{l+D_{\max}}} \right]$$

where (21) is from the application of Hardy-Littlewood-Polya inequality and (22) is derived from the law of large numbers in probability theory. Similarly, for the Criterion 2 of the objective function, we have

$$\mathbb{E} \left[\lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} 2^{R_t} \right]$$

$$\geq \quad \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} \left(2^{\frac{R_t}{l+D_{\max}}} \right)^{l+D_{\max}}$$

$$\geq \quad \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} \mathbb{E} \left[2^{\frac{R_t+R_{t+1}+\dots+R_{t+l+D_{\max}-1}}{l+D_{\max}}} \right] \tag{23}$$

$$\geq \quad \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} \mathbb{E} \left[2^{\frac{X_t+X_{t+1}+\dots+X_{t+l}}{l+D_{\max}}} \right]$$

$$\geq \quad \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}^{l+1} \left[2^{\frac{X}{l+D_{\max}}} \right]$$

$$\geq \quad \mathbb{E}^{l+1} \left[2^{\frac{X}{l+D_{\max}}} \right]$$

where (23) is from the application of Hardy-Littlewood-Polya inequality. ∎

Remark 1 : The bound of the Theorem 3 gives the trivial lower bound of $2^{\mathbb{E}[X]}$ as $l \to \infty$. A special case of this lower bound is given by letting $l = D_{\max} - 1$ as $P_{ave} \geq \mathbb{E}^{D_{\max}}[2^{\frac{X}{2D_{\max}-1}}]$ which is tight only in the case of $D_{\max} = 1$. The derived upper and lower bounds are trivially equal in the case of $D_{\max} = 1$ which corresponds to transmission of the entire packets in the queue at each time instant. On the other hand, it is possible to prove that the upper bound of the Corollary 2 converges to the lower bound $2^{\mathbb{E}[X]}$ as $D_{\max} \longrightarrow \infty$, thus, the lower and upper bound asymptotically converge together.

Theorem 4: The upper bound $\mathbb{E}^{D_{\max}}[2^{\frac{X}{D_{\max}}}]$ converges to the lower bound $2^{\mathbb{E}[X]}$ as $D_{\max} \longrightarrow \infty$.

Proof: For simplicity we assume that the random variable X is discrete, and we expand the upper bound as:

$$\mathbb{E}^{D_{\max}}[2^{\frac{X}{D_{\max}}}] =$$
$$(w_1(2^{x_1})^{\frac{1}{D_{\max}}} + w_2(2^{x_2})^{\frac{1}{D_{\max}}} + \dots + w_n(2^{x_n})^{\frac{1}{D_{\max}}})^{D_{\max}}$$

which is the weighted power mean of 2^{x_i}'s. It can be easily shown that as $D_{\max} \longrightarrow \infty$ this quantity converges to the weighted geometric mean as: [6]

$$\lim_{D_{\max} \longrightarrow \infty} \mathbb{E}^{D_{\max}}[2^{\frac{X}{D_{\max}}}] = (2^{x_1})^{w_1}(2^{x_2})^{w_2} \dots (2^{x_n})^{w_n}$$
$$= 2^{x_1 w_1 + x_2 w_2 + \dots + x_n w_n}$$
$$= 2^{\mathbb{E}[X]}$$

$$\tag{24}$$

which completes the proof. ∎

Remark 2 : The bounds on the performance of the scheduler is depicted in the Figure 2 for the Bernoulli input arrivals distribution. Usually, we need to design a scheduler which performs well for low maximum packet delay, while design of an scheduler which performs well for relatively high maximum delay constraint is not difficult. In fact, the performance of the derived robust time invariant scheduler asymptotically is optimal for large maximum delay constraint. Although the lower bound of $2^{\mathbb{E}[X]}$ is asymptotically tight, it is a loose lower bound for small values of the maximum delay (Figure 2), which are of more interest. The previous works on the power optimal packet scheduling [1, 2] are also deficient in providing a good lower bounds for small values of the average delay. On the other hand, the family of the lower bounds given by Theorem 3 as $\left\{ \mathbb{E}^{l+1} \left[2^{\frac{X}{l+D_{\max}}} \right] \right\}_{l=1}^{\infty}$ provide useful lower bounds for the required average power. For any value of $D_{\max}$, the supremum of this family of bounds is the best lower bound which is illustrated as the envelope of lower-bound curves.

VI. OPTIMAL ROBUST SCHEDULING

In practice, the exact knowledge of the input arrival distribution may not be available or it may change from time to time. Also, The scheduler might perform poorly when there is a mismatch between the actual source distribution and the assumed distribution by the scheduler. We discuss the design of a *Robust scheduler* which performs well regardless of the choice of source distribution. In Section IV, we derived the optimal time-invariant scheduler and also showed that it is independent of the distribution of the input arrival rates to the queue. Thus, it is in fact the optimal time-invariant Robust scheduler when the distribution of the input arrival rates is not known.

In this section, we present the optimal robust scheduler (which is time varying in general) for guaranteed maximum delay. The robustness of this scheduler comes from the fact that the scheduled output rate R_t at
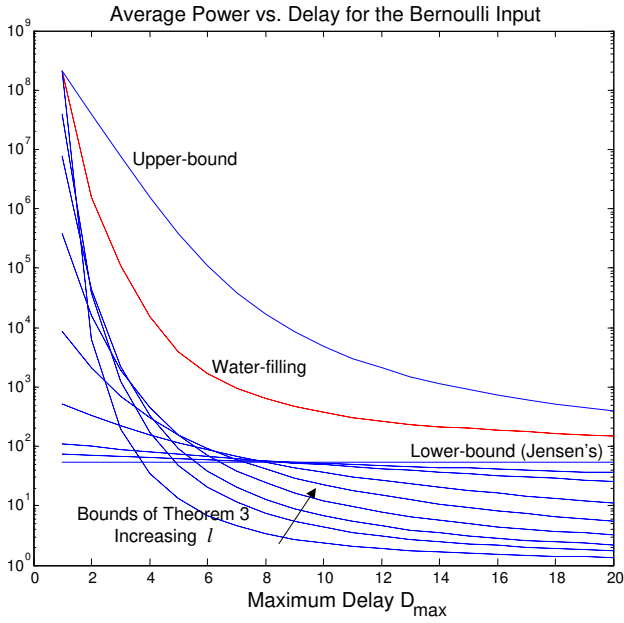
Fig. 2. Performance of the optimal time-varying robust scheduler and the bounds on performance. Note that performance of the optimal time-invariant robust scheduler matches the upper bound.

time t is chosen without the knowledge of the future input arrival statistics. The scheduler is based on the previously observed arrivals to the queue and the amount of queue backlog. On the other hand, the optimality of this scheduler guarantees that there is no scheduler which can outperform this scheduler for all possible source distributions.

We also find the exact distribution of the output service rates for the optimal time varying robust scheduler, which is then used to find the average transmission power. The average power of optimal time-varying scheduler is generally lower than that of the optimal time-invariant robust scheduler which provides a better bound than the one presented in Corollary 2.

A. Optimal Time-varying Robust Scheduler

First, we extend the problem of finding the optimal robust scheduler for guaranteed maximum delay to a more general case, where there are different maximum delays associated to different arrived packets. Then, we consider the case where the maximum delay constraint for all the packets is the same.

1) Different delay constraints for each packet: We find the optimal scheduler for the general delay-constrained scheduling problem when knowledge of input distribution is not available. Solving the general case of the problem would be of more interest in practice where different packets have different maximum delay
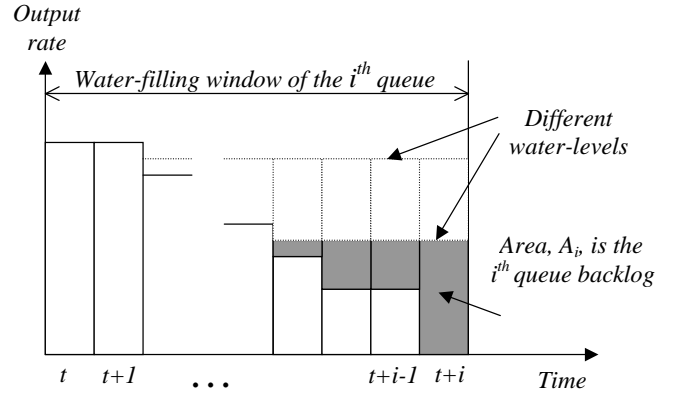

Fig. 3. Representation of the successive water-filling at time t for the ith queue.

constraints associated with different QoS requirements. We assume that the supremum of the delay constraints for all arrived packets is equal to $D_{\max}$. In this case we need to make $D_{\max}$ separate FIFO queues, one for each set of packets with delay requirement of $1, 2, \ldots, D_{\max}$, while in the simpler problem where there is only one maximum delay associated with each packet, only one FIFO queue is needed.

Assume that the scheduler does not know the distribution of future input arrivals. Thus, the only information available to the scheduler at time t would be the past observed values of input arrival rates and the delay associated with each packet. Furthermore, assume that the maximum delay bound of all the packets is no more than $D_{\max}$. We make a circle consisting of $D_{\max}$ queues and label them from 1 to $D_{\max}$ in the clockwise order. At each time t, queue i consists of the packets that have delay constraint i. So we push all the arrived packet with the same delay constraint i into this queue. After we have serviced all of the scheduled packets out of the queues in time t, we go to the time $t + 1$ and we rotate the queue labels clockwise by one position. In other words, the queue which has delay constraint i at the time t would have delay constraint $i - 1$ in the time $t + 1$. Also, the queue 1 would be empty after this queue is serviced at time t, and its label will change to $D_{\max}$ to accommodate the newly arrived packets at time $t + 1$ which have an associated delay constraint of $D_{\max}$.

Theorem 5: At any given time t, let A_i denote the backlog of the queue i for all i, $i = 1, 2, \ldots, D_{\max}$. The optimal output service rate out of the ith queue is given by $X^0_{D_{\max}-i}$, where this value is obtained through successive water-filling of the Equations (41)-(43) in Lemma 2 by letting $D = D_{\max}$.

Proof: It is clear that there is no difference between

any two different packet in the same queue in terms of our scheduling problem. Thus, at any given time t all the observed information about the input arrival rates would reduce to the size of the backlog in each of the queues, which is assumed to be A_i^t for all $i = 1, 2, \ldots, D_{\max}$. Let $\phi(x) = 2^x$ which is a convex function of x, and $A_i = A_i^t$ for all $i = 1, 2, \ldots, D_{\max}$. Using Lemma 2 from the Appendix I, successive water-filling over the next $D_{\max}$ time slot gives the optimal solution for the values of X_i^j for all $\{(i, j) : 0 \leq j \leq D_{\max} - 1,$ and $0 \leq i \leq D_{\max} - 1 - j\}$. Thus, we find the optimal service rates out of the present queue which is given by $X_{D_{\max} - i}^0$ for the $i^t h$ queue (Figure 3). Successive water-filling means that the water-filling for different queues is performed sequentially, and that the water-filling process for the ith queue treats all the other queue yet to be water-filled as an empty queue.

Going to the next time $t + 1$ we need to push new packets to the queues by considering their maximum tolerable packet delays. Thus, the new queue backlog would be $A_i^{t+1} = A_i^t - X_{D_{\max} - i}^0 + N_i^t$ for all $i = 1, 2, \ldots, D_{\max}$, where N_i is the number of arrived packets with maximum delay constraint i at time t. Therefore, in the next time $t + 1$ we need to perform successive water-filling of Lemma 2 again to find the optimal values of service rates out of the queues. Clearly, this process is optimal at each time instant t based on the known information at the time t. ∎

Note that the successive water-filling gives a nice practical way of finding the current transmission rates out of each queue. From Lemma 2, one would notice that at any time instant t, successive water-filling gives the optimal values of the output service rates not only for the present time t but also for the $D_{\max} - 1$ future time instants. But, we would like to emphasize that the entire process of successive water-filling should be redone at each time instant. The reason is that at the new time instant $t + 1$, new input packet arrivals change the backlog of the entire set of queues, thus, the known information has been changed. With this new information, it is not hard to show that the optimal values of the output service rates out of the queues at the time $t + 1$ would not necessarily match with the anticipated values at the time t for this time slot (Figure 4).

2) The same delay constraints for all packets: For the case that all packets have the same maximum delay constraint $D_{\max}$, the solution is fairly easy. Basically there are two main differences which make it simpler: First is the fact that a single FIFO queue is enough to perform scheduling. It is interesting to note that due
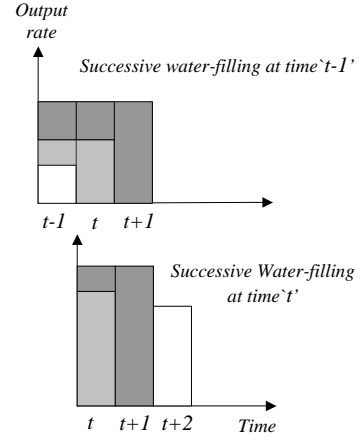

Fig. 4. Water-filling solution for the time slot $t - 1$ and subsequent time slot t.

to the scheduling, there are still different packets with different delay constraints remaining in the queue at any given time t, and the water-filling solution involves finding the optimal scheduled service rate for each of them. Still, the sum of the scheduled output service rates is enough to be considered as an optimal service rate out of the single FIFO queue to guarantee the maximum delay constraint. Second, the water-filling solution is much simpler due to the fact that at each step we just need to do water-filling for the currently arrived packets on top of the old solution. In other words, instead of making $D_{\max}$ queues, all we need is to keep track of $D_{\max}$ numbers at each time which show the water-filling solution of the previous time slot. We specifically have the following corollary based on Theorem 5.

Corollary 3: At any given time slot t, let $(S_0^t, S_1^t, \ldots, S_{D_{\max} - 1}^t)$ denote the vector of the scheduled output service rates out of the queue from the previous time slot $t - 1$, and let X_t be the new arrival which needs to be scheduled. The optimal output service rate for the current time t, R_t, and the new vector of the scheduled service rate for the time $t + 1$, $(S_0^{t+1}, S_1^{t+1}, \ldots, S_{D_{\max} - 1}^{t+1})$, are determined by the water-filling solution as:

$$R_t = (\mu - S_0^t)^+$$
$$S_i^{t+1} = (\mu - S_{i+1}^t)^+ \quad \forall \, 0 \leq i \leq D_{\max} - 2$$
$$S_{D_{\max} - 1}^{t+1} = 0$$
$$R_t + \sum_{i=0}^{D_{\max} - 1} S_i^{t+1} = X_t \tag{25}$$

Therefore, the coefficients of the optimal robust time

variant scheduling filter (scheduler) are given by:

$$\alpha_i^t = \frac{(\mu - S_{i+1}^t)^+}{X_t} \quad \forall \, 0 \le i \le D_{\max} - 1 \qquad (26)$$

The above discussion provides an extremely simple solution to the problem of robust scheduling for delay constrained inputs to a queue. There are some important observations which come out of above solution. First, to find the optimal robust scheduler in Theorem 5, and Corollary 3, although we did not use the filtering property of the scheduler, the optimal solution turns of to be exactly a linear time-variant filter of size $D_{\max}$, which is not surprising due to the Theorem 1. Second, the optimal values of filter coefficient are exactly function of the past $D_{\max} - 2$ values of the input arrivals and the queue backlog, which is expected as we discussed earlier. Third, as we mentioned earlier, the optimal scheduler intuitively should try to make the output service rate as smooth as possible, which corresponds to the low-pass property of the scheduler. Indeed, the water filling solution is the best verification of this intuition. Fourth, the water-filling solution and the Proposition 1 reveal that the optimal solution always tries to push the scheduled time of the packets as far as possible and near the maximum value of the tolerable delay.

In summary, Theorem 2 gives the best time-invariant robust scheduler, while Theorem 5 gives the best time variant robust scheduler, or simply the optimal robust scheduler. The performance of the optimal time varying robust scheduler is depicted in Figure 2. It can be observed that the performance of the optimal time varying robust scheduler is much better than the performance of the time invariant scheduler which is depicted as upper bound in Figure 2. In fact, the performance of the optimal robust scheduler follows the trend of the envelope of the lower bound curves.

B. Performance Analysis

In this section we first derive the distribution of the output service rates, $f_R(r)$, for the time-varying robust scheduler when the input arrival rates are iid with the known distribution $f_X(x)$. The derived distribution can be used to evaluate the performance of the robust time varying scheduler when the input distribution is known. In fact, it is possible to find an analytic expression for the performance of the robust time-varying scheduler which provides a better upper-bound for the performance of the optimal scheduler (which is not necessarily robust). Although this upper bound is lower than that of Corollary 2, the upper bound of Corollary 2 is still useful because of the simplicity of the expression and for the purpose of the numerical calculation.

We assume that the input arrival rate is a stationary and ergodic random process. In this case, the output service rate is an ergodic and stationary random process and it can be found by recursive solution. First, we consider the case of $D_{\max} = 2$ to show how the recursive solution works to find the distribution of the output arrival rates. Then, we drive the distribution of the output arrival rates for the general case. Although, the solution in this case is more involved, the same approach like the simple case of $D_{\max} = 2$ is used.

Suppose that $f_1(.)$ and $f_2(.)$ denote the output service rate distributions of the current and the next time interval, respectively, after water-filling process is performed. Let Y be the random variable showing the value of queue backlog and y be the actual backlog of the queue at time i before performing water-filling. Since, Y is in fact the scheduled service rate for the next time slot at time $i - 1$, its distribution is given by $f_2(y)$. Suppose that the input arrival at time i is given by the random variable X. If $X < Y$ then the scheduled rate for the next time slot would be the same as random variable X, but if $X \ge Y$, then the scheduled rate for the next time slot would be equal to $\frac{X+Y}{2}$. Therefore, at any given time, for the stationary distribution of the next time slot conditioned on the queue backlog $Y = y$ we have

$$f_2(z|Y = y) = \begin{cases} f_X(z) & \text{if } z < y \\ 2f_X(2z - y) & \text{if } z > y \end{cases} \qquad (27)$$

Thus, considering the distribution of the random variable Y, we have

$$f_2(z) = \int_0^z 2f_X(2z - y)f_2(y)dy +$$
$$\int_z^\infty f_X(z)f_2(y)dy \quad (28)$$

which can be written in the form of the homogeneous Fredholm integral equation of the second kind

$$f_2(z) = \int_0^\infty f_2(y)[f_X(z)U(y - z) +$$
$$2f_X(2z - y)(1 - U(y - z))]dy \quad (29)$$

Therefore, for any input distribution of $f_X(x)$, $f_2(.)$ can be obtained by solving the above integral equation, where the kernel is a given function of the input distribution.

Also, if $X < Y$ then the scheduled rate for the current time slot would be the same as random variable Y, but if $X \ge Y$, then the scheduled rate for the current time

slot would be equal to $\frac{X+Y}{2}$. Therefore, at any given time, for the stationary distribution of the current time slot conditioned on the queue backlog $Y = y$ we have

$$f_1(z|Y=y) = \begin{cases} 0 & \text{if } z < y \\ \int_0^z f_X(x)dx & \text{if } z = y \\ 2f_X(2z-y) & \text{if } z > y \end{cases} \quad (30)$$

Therefore, considering the distribution of the random variable Y, we have

$$f_1(z) = \int_0^z f_X(x)dx \, f_2(z) + \int_z^\infty 2f_X(2z-y)f_2(y)dy \quad (31)$$

Thus, after some manipulation the distribution of the output service rate, $f_R(r)$, which is equal to $f_1(r)$ is given by:

$$f_R(r) = \int_0^r f_X(\gamma)[f_2(r) + f_2(2r-\gamma)]d\gamma \quad (32)$$

This result is easily extendable to the case of arbitrary time delay constraint of $D_{\max}$ for $D_{\max} > 2$. Although the solution is more involved for the case of $D_{\max} > 2$, it has the same nature of first finding the solution of a homogenous Fredholm equation of the second kind and then plugging in the answer in a definite multiple integral. We use the same approach to find the distribution of the output service rate in general for an arbitrary $D_{\max}$. let $\vec{Y}$ denote the vector of the random variables $(Y_0, Y_1, \ldots, Y_{D_{\max}-1}$ which are the scheduled service rates for the current and next $D_{\max}-1$ time slots. To find the stationary distribution, we use the recursive equation which relates the values of these random variable at time i to the previous time slot $i-1$. It should be pointed out that in the case of $D_{\max} = 2$ we do not need to consider the joint distribution of the random variables Y_0, Y_1, but in the case of $D_{\max} \geq 3$ it is essential. Let $(y_0, y_1, \ldots, y_{D_{\max}-1})$ be the actual values of the scheduled service rates for the current and next $D_{\max}-1$ time slots at time $i-1$. If X denotes the input arrival at time i, then the vector of the scheduled service rate $\vec{Z}$ would be equal to $(y_1, y_2, \ldots, y_{D_{\max}-k}, z, \ldots, z)$ if and only if $k \in \{1, 2, \ldots, D_{\max}\}$ is the minimum value for which $X + y_{D_{\max}-1} + y_{D_{\max}-2} + \ldots + y_{D_{\max}-k+1} < ky_{D_{\max}-k}$. We define this minimum value $k \triangleq \theta(\vec{y}, X)$ as a function of the value of X and vector $\vec{y}$. Therefore, at any given time i, for the stationary distribution of the joint scheduled service times $\vec{Z}$ conditioned on the queue

backlog $\vec{Y} = \vec{y}$, we have

$$f_{\vec{Y}}(\vec{Z} = (y_1, y_2, \ldots, y_{D_{\max}-k}, z, \ldots, z)|\vec{Y} = \vec{y}) =$$

$$\begin{cases} f_X(z) & \text{if } z < y_{D_{\max}-1} \\ 2f_X(2z - y_{D_{\max}-1}) & \text{if } y_{D_{\max}-1} \leq z < y \\ \vdots & \\ D_{\max}f_X(D_{\max}z - \sum_{j=1}^{D_{\max}-1} y_i) & \text{if } y_1 \leq z \end{cases}$$

$$(33)$$

Using the defined function $k = \theta(\vec{y}, X)$, it can be simplified to

$$f_{\vec{Y}}(\vec{Z} = (y_1, y_2, \ldots, y_{D_{\max}-k}, z, \ldots, z)|\vec{Y} = \vec{y}) =$$

$$kf_X(kz - \sum_{j=1}^{D_{\max}-1} y_i) \quad \text{if } k = \theta(\vec{y}, X)$$

$$(34)$$

Therefore, the problem of finding the distribution of the joint output scheduled rates of the current and next $D_{\max}-1$ time slots reduces to the solving the following integral equation.

$$f_{\vec{Y}}(\vec{z}) = \int_{(\mathcal{R}^+)^{D_{\max}+1}} f_{\vec{Y}}(\vec{z}|\vec{y})f_{\vec{Y}}(\vec{y})d\vec{y}dx \quad (35)$$

$$= \int_{(\mathcal{R}^+)^{D_{\max}+1}} \theta(\vec{y}, X) \ldots$$

$$\ldots f_X\left(\theta(\vec{y}, X)z - \sum_{j=1}^{D_{\max}-1} y_i\right) f_{\vec{Y}}(\vec{y})d\vec{y}dx$$

where $\mathcal{R}^+$ is the set of nonnegative real numbers. Thus, the stationary distribution of the output service rate is given by:

$$f_R(y_0) = \int_{(\mathcal{R}^+)^{D_{\max}-1}} f_{\vec{Y}}(\vec{y})dy_1 dy_2 \ldots dy_{D_{\max}-1} \quad (36)$$

Therefore, the average required power would be

$$P_{ave} = \mathbb{E}_r[2^r] = \int_0^\infty 2^r f_R(r)dr \quad (37)$$

which gives an upper bound lower than that of Corollary 2 obtained with the assumption of a time-invariant scheduler. As we mentioned earlier, although the value of P_{ave} in Equation (37) gives a better upper bound on the P_{ave} for the optimal scheduler with the maximum delay constraint $D_{\max}$, because of its ease of computation, the simple upper-bound of Corollary 2 is still more useful in most of the cases.

VII. Related Work

It is well known that the required power for reliable communication between any two points is exponentially related to the rate of the information which is coded for transmission [1, 2]. Thus, by lowering the transmission power and transmitting over a longer period of time the energy required to transmit a packet can be significantly reduced [4, 9, 10, 12]. On the other hand, the information is usually delay-sensitive, thus, the transmission time cannot be made arbitrarily long, motivating power optimal transmission with different packet delay constraints. In [9], the off-line energy optimal solution has been found for the case that packets arrive in some time interval $[0, T)$, and they have to be delivered by time T. This is a deadline constraint for the arrived packets, thus, the delay corresponding to different packets varies with the arrival time of the packets in the queue. Therefore, the packets arrived at later time have smaller delay constraints than those that arrived earlier. The off-line optimal scheduler [9] assumes noncausal knowledge of all arrival time of the packets in this interval. Although optimality of the causal on-line scheduler is not proved, it has been shown through simulation that it is energy efficient much like its off-line counterpart. The extension of this work for the case of multiple user has been studied in [4] which is solved by proposing two algorithms, the MoveRight and the MoveRightExpress algorithms corresponding to noncausal and causal schedulers, respectively. In [10], authors combined the energy efficiency by lowering the transmission rate and increasing the transmission time with the recovery of the batteries due to the electrochemical mechanism. The authors also considered both the deadline and average delay constraint for the packets. In [1, 2], minimization of the average transmission power with the average delay constraint have been considered. This minimization problem can be turned into a convex optimization problem and the mentioned work have proposed dynamic programming formulation to find the optimal solution.

In this work, we consider strict maximum delay constraint for each arrived packet instead of the average delay [1, 2] or deadline constraint [4, 9, 12] for a group of packets. We also consider a more general case where each packet has different strict delay constraint. In [9], off-line scheduler assumes prior knowledge of the packet arrival time (noncausal scheduler), while we consider the causal scheduler that looks at the previous time and number of packet arrivals and has no prior knowledge of the time or number of packets which will arrive later. We introduce a robust scheduler that performs well regardless of the choice of the input distribution. The off-line scheduler (which is noncausal) in [9] also follows a distribution free approach, but the online version of the scheduler (which is causal) was not proved to be optimal. The robust scheduler which is presented in this paper is proved to be optimal in a sense that if the knowledge of the input arrival distribution is not available, then it is the best possible scheduler which minimizes the average power based on just the past observation of the input arrivals. The nice property of the derived analytic solution for the optimal robust scheduler is that the solution follows a simple water-filling in time, which leads to the design of a scheduler with very low-complexity in comparison to the design of schedulers in [1, 2] that require a dynamic programming based solution.

In addition, we build upon the connectiona between the filtering and scheduling [3]. We use the filtering connection to establish new lower and upper bounds on the performance of the optimal scheduler and provide the scheduling schemes which achieve the upper bounds. The new lower bounds are especially useful for low values of the maximum delay constraint where no prior good bound is known. The proof technique in deriving the lower bounds in this paper might be of independent interest. Finding a lower bound by using inequalities that rely on the property of convex functions usually reduces to Jensen's lower bound which is asymptotically optimal but does not provide a good bound for low values of the delay constraint. On the other, we use Hardy-Littlewood-Polya inequality which is inherently different from the class of inequalities which rely on convex function and it captures the effect of ordering.

VIII. Conclusion

In this paper, we presented several fundamental results pertaining to scheduling with maximum delay constraints for Gaussian channels. We established a connection between low-pass filtering and scheduling, and derived both time-invariant and time-varying optimal schedulers which do not rely on the knowledge of source distribution. New bounds on the average power consumption were presented, which are most useful for the case of small delays.

References

[1] R. Berry and R. Gallager, "Communication over Fading Channels with Delay Constraints," *IEEE Transactions on Information Theory*, vol. 48, no. 5, pp. 1135-1149, May 2002.

[2] D. Rajan, A. Sabharwal, and B. Aazhang, " Delay-bounded Scheduling of Bursty Sources over Wireless Channels," *to appear in IEEE Transactions on Information Theory*, 2004.

[3] M. A. Khojastepour and A. Sabharwal, "Power Optimal Scheduling with Maximum Delay Constraints," *Forty First Allerton Conference on Communication, Control, and Computing*, Allerton, IL, October 2003.

[4] A. El Gamal, C. Nair, B. Prabhakar, E. Uysal-Biyikoglu, and S. Zahedi, "Energy-efficient Scheduling of Packet Transmissions over Wireless Networks," *IEEE Infocom Conference on Computer Communications*, Vol. 3, pp. 1773-1782, New York City, June 2002.

[5] J.C. Geromel, M.C. de Oliveira, and J. Bernussou, "Robust Filtering of Discrete-Time Linear Systems with Parameter Dependent Lyapunov Functions," *SIAM Journal on Control and Optimization*, Vol. 41, No. 3, pp. 700-711, 2002

[6] G.H. Hardy, J.E. Littlewood, G. Polya, Inequalities, Cambridge Press 1988.

[7] S.O.R. Moheimani, A.V. Savkin, and I. R. Petersen, "Robust Filtering, Prediction, Smoothing and Observability of Uncertain Systems," *IEEE Transactions on Circuits and Systems*, Vol. 45, No. 4, pp. 446-457, April 1998.

[8] L.H. Ozarow, S. Shamai, and A.D. Wyner, "Information Theoritic Considerations for Cellular Mobile Radio," *IEEE Transactions on Vehicular Technology*, Vol 43, No 2, pp. 359-378, May 1994.

[9] B. Prabhakar, E. Uysal-Biyikoglu, A. El Gamal, "Energy-efficient Transmission over a Wireless Link Via Lazy Packet Scheduling," *Proceedings of the IEEE INFOCOM*, Vol. 1, pp. 386-394, Anchorage, Alaska, April 2001.

[10] P. Nuggehalli, V. Srinivasan, R.R. Rao, "Delay Constrained Energy Efficient Transmission Strategies for Wireless Devices ," *Proceedings of the IEEE INFOCOM*, New York, NY, June 2002.

[11] J.G. Proakis and D.G. Manolakis, Digital signal processing: principles, algorithms and applications, Macmillan (1996).

[12] E. Uysal-Biyikoglu, B. Prabhakar, A. El Gamal, "Energy-efficient Transmission over a Wireless Link," *IEEE ACM Transactions on Networking*, Vol. 10, No. 4, pp. 487-499, August 2002.

[13] W.S. Yoon and T. Klein, "Delay-Optimal Power Control for Wireless Data User with Average Power Constraints", *IEEE International Symposium on Information Theory 2002*, Lausanne, Switzerland, June 30 July 5, 2002.

APPENDIX I

Lemma 2: Consider the constrained optimization problem

$$J = \min \sum_{i=0}^{D-1} \phi \left(\sum_{j=0}^{D-1-i} X_j^i \right) \qquad (38)$$

subject to the constraints

$$\forall\, 0 \le i, j \le D-1 : X_j^i \ge 0 \qquad (39)$$

$$\forall\, 0 \le j \le D-1 : \sum_{i=0}^{D-1-j} X_j^i = A_j \qquad (40)$$

where $\phi(.)$ is a convex function and A_j's are some nonnegative real values. The successive water-filling defined in Equations (41)-(43) is the optimal solution.

$$X_{D-1}^0 = A_{D-1} \qquad (41)$$

$$X_{D-2}^0 = (\mu_{D-2} - X_{D-1}^0)^+$$
$$X_{D-2}^1 = \mu_{D-2}$$
$$X_{D-2}^0 + X_{D-2}^1 = A_{D-2} \qquad (42)$$

and successively water-fill for all $i = 3, 4, \ldots, D-1$ according to,

$$X_{D-i}^j = (\mu_{D-i} - \sum_{k=j+1}^{i-1} X_{D-k}^j)^+ \quad \forall\, 0 \le j \le (i-1)$$

$$X_{D-i}^{i-1} = \mu_{D-i}$$

$$\sum_{j=1}^{i-1} X_{D-i}^j = A_{D-i}. \qquad (43)$$

Sketch of the proof: Since it is a convex optimization problem with it is easy to find the necessary and sufficient condition for the optimal solution based on Lagrange multiplier technique. After finding the gradient and applying KKT conditions the solution turns of to be,

$$X_{D-i}^j = (\mu_{D-i} - \sum_{k=j+1}^{i-1} X_{D-k}^j - \sum_{k=0}^{i-1} X_k^j)^+$$
$$\forall\ 0 \le j \le (i-1)$$

$$X_{D-i}^{i-1} = (\mu_{D-i} - \sum_{k=0}^{i-1} X_k^j)^+$$

$$\sum_{j=1}^{i-1} X_{D-i}^j = A_{D-i}. \qquad (44)$$

Since determining X_k^j depends on for both prior $(i < k)$ and subsequent $(i > k)$ values of X_i^j the solution of the above simultaneous set of equations is not easy to see. The successive water-filling Equations (41)-(43) remove the dependency in determining the value of X_k^j to the prior values of X_i^j for the $(i < k)$. Therefore, first we use the successive water-filling Equations (41)-(43) to find the values of X_i^j for all $\{(i,j) : 0 \le j \le D-1,$ and $0 \le i \le D-1-j\}$, and μ_i for all $0 \le i \le D-1$. Then, we prove that there exist some other value of μ_i' such that the together with the same values of X_i^j satisfy the KKT conditions of the Equation (44). ∎

It should be pointed out that the optimal solution to the above optimization problem is not necessarily unique and once we get one solution (for example successive water-filling) it is easy to generate other solutions. Also, it is easy to prove the following interesting property for the water-filling solution:

Proposition 1: The argument of the function ϕ in the objective function J, $\sum_{j=0}^{D-1-i} X_j^i$, is a non-increasing function of i for $i = 0, 1, \ldots, D-1$.