Optimal Shaping for Transmission over Wireless Fading Channels*

Farshid Agharebparast and Victor C. M. Leung

Department of Electrical and Computer Engineering, The University of British Columbia, Canada
E-mail: {farshid, vleung}@ece.ubc.ca

ABSTRACT In this paper, we derive the parameters of an optimal leaky-bucket shaper for the transmission of constrained traffic over a wireless fading channel. The analysis is based on an efficient stochastic min-plus system theory for the derivations of second-order statistics, as a wireless fading channel is a stochastic rate-latency server. The shaper is optimal in the sense that it minimizes the end-to-end delay and buffer requirements under the constraints dictated by the lower layer characteristics, in particular the wireless channel. Our model is used to plot the analytic delay and buffer sizes in comparison to simulation results for various values of shaper’s parameters, from which the optimal or sub-optimal solution is calculated.

I. INTRODUCTION

The traffic entering a network may be shaped for a number of reasons. It is shaped at the source or a network edge to conform to a predefined service level agreement to make it eligible for receiving the negotiated quality of service (QoS). Shaping is also used to ensure a fair and stable utilization of the network. Finally, it is implemented as a rate-control mechanism to smooth a bursty traffic stream.

In a wireless network, the bottleneck is usually the time-varying wireless channel. Therefore, the available effective rate should be taken into consideration. In this paper, we show how to derive the parameters of a leaky bucket shaper in order to minimize end-to-end delay at the packet level and the buffer requirements. The minimization is done under constraints dictated by the physical layer characteristics such as signal-to-noise ratio (SNR) statistics and channel rate, or by the traffic requirements like a permitted traffic burstiness.

The analytical model is based on a novel stochastic min-plus system theory that was thoroughly discussed in an earlier work [1]. This theory explains the input/output relationship of a stochastic min-plus system. It is used to derive the second-order statistics of the output, when input and/or service are stochastically characterized. As it will be explained later, a wireless fading channel is modeled by a stochastic rate-latency server whose second-order statistics are easily calculable from the knowledge of the physical layer characteristics. Therefore, we use this method to find the statistics of the traffic (the output) that passes through a wireless channel; i.e., the mean and autocorrelation of the traffic at the output of the channel. Then by using Network Calculus theorems [2], the optimal values for the parameters of the shaper are calculated.

The outline of the paper is as follows. In the next section, an overview on the theory of optimal shaping is presented. The system model, assumptions and descriptions of the underlying concepts and building blocks are addressed in Section III. Section IV presents the theory of shaping for transmission over a wireless channel and the methodology for calculating the parameters of the optimal shaper. Section V discusses the numerical and simulation results.

II. OPTIMAL LEAKY BUCKET SHAPER

In this paper, we use the widely accepted min-plus system theory or Network Calculus [2] representations for the traffic and service. A system is a representation of a network or a network element (link, node, or sub-network) which provide a service defined by \( h(t) \). The incoming traffic (input) is represented by \( x(t) \), and \( y(t) \) represents the outgoing traffic (output received by the end element), both in units of bytes. The traffic and service are characterized by the arrival and service curves, respectively, and are all represented by non-decreasing functions. The output of a system with a service curve \( h(t) \) is expressed by \( y(t) \geq L^{h(t)}(x(t)) \). The min-plus convolution is defined by: \( x(t) \circ h(t) = \inf_{0 \leq s \leq t} (x(t-s)+h(s)) \).

A leaky bucket server is a well-known shaper with two parameters: the average rate \( a \) and the burst tolerance \( b \). Any traffic passing through such a shaper will have an arrival curve of \( a(t) = at+b \). The rate-latency server \( \beta_{\mu,T} = \mu(t-T)^+ \) is a widely-used model. It describes a server that guarantees a minimum service rate \( \mu \) and maximum service latency \( T \). As explained later, a wireless fading channel can be modeled by a stochastic rate-latency server.

An optimal leaky bucket shaper is the one that minimizes the maximum delay and buffer requirement. The minimization is subject to the constraint dictated by the QoS policy (affecting the selection region for the shaper’s parameters) and the lower layer characteristics such as SNR (affecting the service curve).

There are two analytical methods for delay and buffer calculations, noted here by A and B. In method A, the input and output functions are used. The data transmitted at time \( t \) encounters a delay of \( d(t) = \inf \{ \Delta : \Delta \geq 0 \text{ and } x(t) \leq y(t+\Delta) \} \), and the virtual backlog at time \( t \) inside the system is \( B(t) = x(t) - y(t) \). Then the maximum delay and required buffer are, respectively:

\[
d_{\max} = k(x,y) = \sup_{t \geq 0} (d(t)) \quad \text{and} \quad B_{\max} = v(x,y) = \sup_{t \geq 0} (B(t)).
\]

The maximum delay \( d_{m} \) is the maximum horizontal distance, and the maximum required buffer \( B_{m} \) is the maximum vertical.
distance between the transmitted (input) and the received (output) traffic curves.

Method B gives a looser bound, but is more convenient. It is based on finding the maximum delay and backlog from the arrival and service curves instead. When the arrival and service curves tightly represent the input and service, method B gives a good estimate for the delay and buffer size. Similarly, it is shown that for a system with a service curve of \( h(t) \) and input traffic with an arrival curve of \( \xi(t) \), the delay bound (maximum buffer size) is graphically the maximum horizontal (vertical) distance between the graphs of \( h(t) \) and \( \xi(t) \) [2].

Method A is preferred and used in section V. However, method B is used in the following to provide an insight to the rule of the parameters involved which is not obvious from method A. Let assumed a traffic shaped by a leaky bucket passes through a rate-latency server. Then it can be shown that when \( a \leq \mu \), for the desired values of \( d_{\text{max}} \) and \( b_{\text{max}} \), we need \( b = (d_{\text{max}} - T)\mu \) and \( a = (b_{\text{max}} - b) / T \). If the QoS policy indicates a shaping of \( a_p t + b_p \), then we should choose \( \min(a, a_p) \) and \( \min(b, b_p) \). When \( a > \mu \), the arrival and service curves diverge. Therefore, the maximum delay depends on the length of the transmission session and increases accordingly.

III. SYSTEM MODEL AND BUILDING BLOCKS

The packet-level transmission over a wireless channel is modeled as a system with shaped input \( x(t) \) as the traffic from the packet layer entering the lower layers, a stochastic server with a service curve \( h(t) \), and output \( y(t) \) which is received at the packet level at the other end of the channel.

Figure 1 shows the system model. The traffic is shaped with a leaky bucket and is sent to a buffer for transmission over the channel via the lower layers. In most wireless implementations, such as in UMTS, the radio link layer permits the network layer to see the wireless link as a radio bearer providing a logical transmission link. For loss free transmissions, we consider that a stop-and-wait Automatic Repeat-Request (ARQ) mechanism is deployed. The buffer requirement is dependent on the available lower layers’ buffer sizes for a reliable transmission specially when an ARQ mechanism for loss free transmissions is deployed. We assume that the packets are transmitted in frames so that the transmitter will be acknowledged by the next available slot, whether a retransmission is required. We consider a Rayleigh fading channel with Alamouti diversity. Such a channel will have a SNR statistics defined by a chi-square function. The physical layer provides an estimate of the average SNR \( (E[\gamma]) \) to the shaper module. In the following, we show that the channel can be modeled by a stochastic rate-latency server.

**Wireless Channel Service Curve**: A function that practically characterizes the instantaneous service rate of a wireless channel observed at the packet layer is

\[
r_c(t) = (1 - \text{PER}(t))\mu,
\]

when the packet error rate (PER) at any given time \( \text{PER}(t) \)

and the channel transmission capacity \( (\mu) \) measured in packet units at \( \Delta \) time interval \( (\Delta \to 0) \) are known [4]. What (3) describes is the instantaneous rate that is deemed available by the network layer for successful transmissions.

Since the channel is time-varying, the PER is a random process whose statistics need to be characterized. This can be done by sequential calculations of statistics of SNR, then (BER), and finally PER.

The instantaneous SNR per bit, \( \gamma \), for an un-correlated time-varying Rayleigh fading channel with Alamouti transmit diversity has a central chi-square distribution with two degrees of freedom. The Alamouti transmit diversity scheme is one of the proposed methods for IEEE 802.16 (WiMAX) to achieve 2-way diversity without adding an extra antenna (described in the IEEE 802.16abc-01/53 specification). The probability density function (pdf) of \( \gamma \) is:

\[
\gamma(\gamma) = \frac{\gamma(\gamma) - \gamma(2N_0/E)\exp(-\frac{\gamma(2N_0/E)})}{E/N_0 = E(\gamma)}.
\]

Without loss of generality, we assume a \((\pi/4)\)-QPSK modulation, independent distribution of bit errors in each \( n \)-bit packet and no channel coding. Then, BER and PER are defined by

**BER(t) = Q(\sqrt{2SNR(t)}) and PER(t) = 1 - (1 - BER(t))^n**.

Since the service curve of the channel is the cumulative service, which is the integral (or summation for sampled or discrete representation) of (3), then:

\[
h(t) = \int_0^t r_c(t)dt = \int_0^t (1 - \text{PER}(t))\mu dt = \left[ t - \int_0^t \text{PER}(t)dt \right] \mu. \quad (4)
\]

It is interesting to note that (4) is in the format of a rate-latency server, where its latency \( (T) \) is a stochastic process.

**Stochastic Output Characterization**: In an earlier paper [1], we proved two fundamental theorems that define the second-order statistics of the output of a min-plus stochastic system, when the second-order statistic of service curve are given. The second-order statistics of \( h(t) \) are its mean \( E(h(t)) = \eta_h \) and autocorrelation \( R_{hh}(t_1, t_2) = E[h(t_1)h(t_2)], \ t_2 \geq t_1 \). The variance is then \( \text{Var}(h(t)) = R_{hh}(t, t) - \eta_h^2 \).

**Theorem I**: In a min-plus system with lower service curve \( h(t) \):

\[
\eta_{\gamma} \geq \eta_h \otimes x. \quad (5)
\]
find the parameters of the optimal leaky bucket shaper that and $B$ are presented in this section. The methods are used to rated in the service curve), as well as other factors such as constraints are the characteristics of the wireless channel (incorpo-
data over a wireless fading channel. The main challenge is the optimal values for the shaper.

Two methodologies based on the aforementioned methods $A$ and $B$ are presented in this section. The methods are used to find the parameters of the optimal leaky bucket shaper that minimizes the overall delay and required buffer for transmitting data over a wireless fading channel. The main challenge is the stochastic behaviour of the channel. The optimization constraints are the characteristics of the wireless channel (incorporated in the service curve), as well as other factors such as allowing a certain desirable burst tolerance.

Method $A$: In order to use method $A$, we first use (5) and (6) to find the second-order statistics (mean and autocorrelation) of the output; i.e., its mean $\eta_i(t)$ and variance $Var(y)(t)$. Then we calculate the mean and worst case output curves. The worst case output is defined using $\eta_i(t) – \sqrt{Var(y)(t)}$. As an example, for a normal distribution, it is within one standard deviation around the mean with a probability of 68%, and within two standard deviations around the mean with a probability of 95%. So we can use (1) and (2) to find the values of the expected and worst delays and buffer requirements for different values of the parameters. These results are then used to select the parameters of the optimal (or sub-optimal) shaper. Figure 2-A graphically shows this method.

Method $B$: As (4) suggests, a wireless channel is a stochastic rate-latency server whose $T$ is a random process, which its second-order statistics can be calculated. Having the mean latency $T_m$ and worst case latency $T_w$, the mean and worst case delays are calculated (Figure 2-B). We then use (4) to find the optimal values for the parameters $a$ and $b$, using the value of either mean or worst latency, depending on the required level of service quality.

Method $A$ is the preferred methodology for a number of reasons. First, the bound derived by method $B$ are loose and overestimated, specially when the actual traffic and service are not close to their curves. In addition, method $B$ is only applicable for a leaky bucket and rate-latency server, and only when $a \leq \mu$. In contrast, method $A$ is general and may be used for any type of shaper and service curve.

**V. Numerical Results**

We consider a practical scenario in which a server is transmitting continuous-media data to a wireless user via a base station. The base station can be either the originating source or a last-mile connecting point. For an end-to-end scenario with concatenations of $n$ rate-latency nodes, we will have an overall rate latency service curve of $\mu_a(t – T_a)$ where $\mu_a = min(\mu_1, ..., \mu_n)$ and $T_a = \sum_n T_1$.

We assume exponentially distributed interarrival times and packet sizes. The parameters used in the numerical example and their values are: mean interarrival times (MIntArr = 0.7 ms), mean packet size (MPkSz = 3 KB), shaper’s rate ($a$ = MPkSz/MIntArr), burst tolerance ($b$ = [2 12]*MPkSz), average SNR ($E[\gamma] = 10$ dB) and channel rate ($\mu = [5 1.2]*a$).

We present the mean and worst case plots for the delay and buffer sizes for a wide range of values: $\mu/a \in [0.5, 0.7, 0.8, 0.9, 1, 1.1, 1.2]$ and $b \in [2, 3, 5, 7, 9, 12]*MPkSz$. The $\mu/a$ ratio is used here to emphasize on the importance of their relative values. However, it goes without saying that the results depend on the values of $\mu$ and $a$, not the ratio. We consider a wide range of values for this ratio, even for values smaller than one in which the delay and buffer sizes are session-length dependent. This condition is normally avoided, unless when the transmission session is short.

The analytical results using method $A$ are presented and compared with the simulation results. The numerical analysis results in three dimensional plots. Due to the difficulty in legibly presenting such graphs, as well as lack of space, the contour of these graphs are presented here.

Figure 3 and Figure 4 show the plots of the mean and worst delays, respectively, from analysis and simulation. While Figure 5 and Figure 6 show the plots of the mean and worst buffer sizes from analysis and simulation. The values on the contours represent the delay (or the buffer size) for the corresponding values of $b$ and $\mu/a$.

From the graphs, those values for $\mu/a$ and $b$ are chosen that either are optimal or satisfy a given delay and/or buffer requirement as closely as possible (sub-optimal). An optimal set of values results in the minimum worst delay and buffer size. For example, if we are constrained by a worst case delay of 15s and...
buffer size of 60MB, and the available channel rate is \( \mu = 4.2 \) MB/s, then from the graphs \( \mu/a = 1 \) which results in \( b = 25 \).

Figure 3 to Figure 6 are plotted with the assumption that the average SNR is 10 dB. The statistics of the latency (\( T \)) of the rate-latency server is directly related to SNR. Therefore, as seen in Figure 2, the value of the delay is accordingly affected by SNR. Figure 7 shows the values of the mean and worst delays versus average SNR changing from -20 to 40 dB, with a 95% confidence interval (30 iterations of 50 instances each). The figure shows that for any value larger than 10 dB, the delay almost stays constant. Therefore, it is the best choice for this scenario.

**VI. CONCLUSIONS**

In this paper, we have presented a methodology for the derivation of optimal parameters of a leaky bucket shaper for transmission over wireless fading channels. The method is based on a novel stochastic min-plus system theory approach for second-order statistics. The shaper results in minimum (or feasible minimum) delay and buffer requirements, considering the imposed constraints. We have presented numerical and simulation results to support the proposed methodology.

**REFERENCES**


