Entropy-optimal Generalized Token Bucket Regulator

Ashutosh Deepak Gore and Abhay Karandikar
Information Networks Laboratory
Department of Electrical Engineering
Indian Institute of Technology - Bombay
{adgore,karandi}@ee.iitb.ac.in

***Abstract*— We derive the maximum entropy of a flow (information utility) which conforms to traffic constraints imposed by a generalized token bucket regulator, by taking into account the covert information present in the randomness of packet lengths. Under equality constraints of aggregate tokens and aggregate bucket depth, a generalized token bucket regulator can achieve higher information utility than a standard token bucket regulator. The optimal generalized token bucket regulator has a near-uniform bucket depth sequence and a decreasing token increment sequence.**

I. Introduction

In Internet Quality of Service (QoS) parlance, as a part of the service level agreement (SLA) between a subscriber (source) and an Internet service provider (ISP), a token bucket regulator (TBR) can be used to smoothen the bursty nature of a subscriber's traffic [1]. The SLA mandates that the ISP provide end-to-end loss and delay guarantees to a subscriber's packets, provided the traffic profile of the subscriber adheres to certain TBR constraints. The standard token bucket regulator (STBR), as defined by the Internet Engineering Task Force (IETF), enforces linear-boundedness on the flow and is characterized by the token increment rate r and the bucket depth B. We will be more general and consider a TBR in which the token increment rate and bucket depth (maximum burst size) can vary from slot to slot. Such a TBR, which we define as a generalized token bucket regulator (GTBR), can be used to regulate variable bit rate (VBR) traffic[1] from a source [2]. The continuous-time analogue of a GTBR is the time-varying leaky bucket shaper [3] in which the token rate and bucket depth parameters can change at specified time instants. In [3], the authors determine the optimal parameters (rates and bucket sizes) and apply it to the renegotiable VBR service.

Our primary contribution is developing the notion of information utility of a GTBR. Specifically, we derive the maximum information that a GTBR-conforming traffic flow can convey in a finite time interval, by taking into account the additional information present in the randomness of packet lengths. The idea of using a covert channel to convey side information[2] in data networks has been investigated earlier in the classic papers [4] [5]. In this paper, the side information is considered in the lengths of the packets only. Of all the packet length schedules that conform to a given GTBR, our objective is to stochastically characterize the flow that has the maximum entropy.

In [6], the authors have derived the information utility of an STBR and suggested a pricing viewpoint for its application. Our interest is more theoretical – we consider an STBR as a special case of a GTBR and describe a framework for their information-theoretic comparison. We investigate whether a GTBR can achieve higher flow entropy than an STBR and explain the properties of entropy-maximizing GTBRs.

Section II explains our system model. In Section III, we derive the optimal flow entropy equation and define the information utility of a GTBR. In Section IV, we formulate the optimal GTBR and derive a necessary condition. In Section V, we compute the optimal GTBR. We interpret our results in Section VI, and conclude in Section VII.

[1]For example, a pre-recorded video stream.

[2]Information present in packets other than the actual packet contents.

II. System model

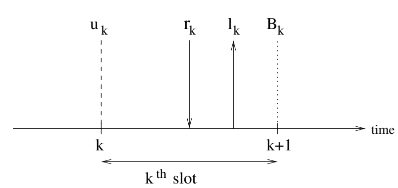

Fig. 1. Relative time instants of parameters defined in (1).

Consider a system in which time is divided into slots and a source which has to complete its data transmission within N slots. In our discrete-time model, we will evaluate the system at time instants $0, 1, \ldots, N-1, N$. The k^{th} slot is defined to be the time interval $[k, k+1)$. The traffic from the source is regulated by a GTBR. Define

$$\begin{aligned} r_k &:= \text{ token increment for the } k^{th} \text{ slot} \\ B_k &:= \text{ bucket depth for the } (k+1)^{th} \text{ slot} \\ \ell_k &:= \text{ length of packet transmitted in the } k^{th} \text{ slot} \\ u_k &:= \text{ residual tokens at start of the } k^{th} \text{ slot} \end{aligned} \quad (1)$$

r_k, B_k, ℓ_k and u_k, whose relative time instants are shown in Figure 1, are all non-negative integers. Let $\mathbf{r} := (r_0, r_1, \ldots, r_{N-1})$ denote the token increment sequence and $\mathbf{B} := (B_0, B_1, \ldots, B_{N-2})$ denote the bucket depth sequence. The system starts with zero tokens; $u_0 = 0$. A GTBR $\mathcal{R}$ with the above parameters, written as $\mathcal{R}(N, \mathbf{r}, \mathbf{B})$, constrains the packet lengths according to

$$\ell_i \leq u_i + r_i \quad \forall\, i : 0 \leq i \leq N-1 \tag{2}$$

If (2) is satisfied, then $\ell = (\ell_0, \ell_1, \ldots, \ell_{N-1})$ is a conforming packet length vector and u_i evolves as

$$\begin{aligned} u_{i+1} &= \min(u_i + r_i - \ell_i, B_i) \;\; \forall\, i : 0 \leq i \leq N-2 \\ u_N &= u_{N-1} + r_{N-1} - \ell_{N-1} \end{aligned} \tag{3}$$

If $r_i = r$ and $B_i = B$ for all i, then the GTBR $\mathcal{R}_g(N, \mathbf{r}, \mathbf{B})$ degenerates to the STBR $\mathcal{R}_s(N, r, B)$.

III. Information utility

Consider a source which has a large amount of data to send and whose traffic is regulated by a GTBR. We seek to maximize the information that the source can convey in the given time interval or the entropy present in the source traffic flow in an information-theoretic sense. The maximum entropy achievable by any flow which is constrained by the GTBR $\mathcal{R}(N, \mathbf{r}, \mathbf{B})$ is defined to be its information utility. The source can send information to the destination via two channels:

i) Overt channel: The contents of each packet. Let ℓ_i be the length of a packet in bits. The value of each bit is 0 or 1 with equal probability and is independent of the values taken by the preceding and succeeding bits. The packet thus contributes ℓ_i bits of information.
ii) Covert channel: We consider the length of a packet as an event and associate a probability with it. Thus, side information is transmitted by the randomness in the packet lengths.

At time k, the only method by which past transmissions can constrain the rest of the flow is by the residual number of tokens u_k. The key observation is that the future entropy depends only on the buffer level u_k at time k. So, u_k captures the state of the system. Entropy is a function of system state u_k and is denoted by $H_k(u_k)$.

At time N, the source signals the termination of the current flow by transmitting a special string of bits (flag). The information transmitted by this fixed sequence of bits is zero. Thus

$$H_N(u_N) = 0 \tag{4}$$

For a given state u_k of the system, if a packet of length ℓ_k bits is transmitted with probability $p_{\ell_k}(u_k)$, then:

1) The overt information transmitted is ℓ_k bits.
2) As the event occurs with probability $p_{\ell_k}(u_k)$, the covert information transmitted is $(-\log_2 p_{\ell_k}(u_k))$ bits.
3) Since ℓ_k is random, u_{k+1} is also random (from (3)). Thus, $H_{k+1}(u_{k+1})$ is also a random variable.

Adding all of the above and averaging it over all conforming packet lengths, we obtain the entropy of the current stage:

$$H_k(u_k) = \sum_{\ell_k=0}^{u_k+r_k} p_{\ell_k}(u_k)\Big(\ell_k - \log_2(p_{\ell_k}(u_k)) + H_{k+1}(\min(u_k + r_k - \ell_k, B_k))\Big) \;\; \forall\, k = 0, \ldots, N-1 \tag{5}$$

Finally, the above probabilities must satisfy

$$\sum_{\ell_k=0}^{u_k+r_k} p_{\ell_k}(u_k) = 1 \;\; \forall\, k = 0, \ldots, N-1 \tag{6}$$

Let $\mathbf{p}_k(u_k) := (p_0(u_k), p_1(u_k), \cdots, p_{u_k+r_k}(u_k))$. Our objective is to determine the sequence of probability mass functions[3] $(\mathbf{p}^*_{N-1}, \mathbf{p}^*_{N-2}, \cdots, \mathbf{p}^*_0)$ which maximizes the flow entropy $H_0(0)$ for a given GTBR $\mathcal{R}(N, \mathbf{r}, \mathbf{B})$. From (4)

$$H^*_N(u_N) = 0$$

From (5)

$$H_k(u_k) = \sum_{\ell_k=0}^{u_k+r_k} p_{\ell_k}\Big(\ell_k - \log_2(p_{\ell_k}) + H^*_{k+1}(\min(u_k + r_k - \ell_k, B_k))\Big) \;\; \forall\, k = 0, \ldots, N-1$$

Given $H^*_{k+1}(u_{k+1}) \;\forall\; u_{k+1}$, there exists an optimum probability vector $\mathbf{p}^*_k = (p^*_0, p^*_1, \ldots, p^*_{u_k+r_k})$ which maximizes the flow entropy $H_k(u_k)$.

$$\therefore H^*_k(u_k) = \sum_{\ell_k=0}^{u_k+r_k} p^*_{\ell_k}\Big(\ell_k - \log_2(p^*_{\ell_k}) + H^*_{k+1}(\min(u_k + r_k - \ell_k, B_k))\Big) \;\; \forall\, k = 0, \ldots, N-1 \tag{7}$$

Thus, the problem of computing the entire sequence of probability vectors $(\mathbf{p}^*_{N-1}, \mathbf{p}^*_{N-2}, \cdots, \mathbf{p}^*_0)$ has now been decoupled into a sequence of subproblems. The subproblem for time k is:
Given the function $H^*_{k+1}(u_{k+1}) \;\forall\; u_{k+1}$, determine the probability vector $\mathbf{p}_k = (p_0, p_1, \ldots, p_{u_k+r_k})$ so as to

$$\text{maximize} \sum_{\ell_k=0}^{u_k+r_k} p_{\ell_k}\Big(\ell_k - \log_2(p_{\ell_k}) + H^*_{k+1}(\min(u_k + r_k - \ell_k, B_k))\Big) \;\text{ subject to } \sum_{\ell_k=0}^{u_k+r_k} p_{\ell_k} = 1 \tag{8}$$

(8) can be solved using Lagrange multipliers.

$$\mathcal{L}(\mathbf{p}_k, \lambda_k) := \sum_{\ell_k=0}^{u_k+r_k} p_{\ell_k}\Big(\ell_k - \log_2(p_{\ell_k}) + H^*_{k+1}(\min(u_k + r_k - \ell_k, B_k))\Big) + \lambda_k\Big(\sum_{\ell_k=0}^{u_k+r_k} p_{\ell_k} - 1\Big) \tag{9}$$

At the optimal point $(\mathbf{p}^*_k, \lambda^*_k)$

$$\left.\frac{\partial \mathcal{L}}{\partial p_{\ell_k}}\right|_{(\mathbf{p}^*_k, \lambda^*_k)} = 0 \quad \forall\, \ell_k = 0, \ldots, u_k + r_k \tag{10}$$

$$\left.\frac{\partial \mathcal{L}}{\partial \lambda_k}\right|_{(\mathbf{p}^*_k, \lambda^*_k)} = 0 \tag{11}$$

[3] The dependence of p_{ℓ_k} and $\mathbf{p}_k$ on u_k is assumed to be understood and is not always stated explicitly. So, $\mathbf{p}_k = (p_0, p_1, \cdots, p_{u_k+r_k})$.

Solving (11)

$$\sum_{\ell_k=0}^{u_k+r_k} p^*_{\ell_k}(u_k) = 1 \tag{12}$$

Solving (10)

$$p^*_{\ell_k}(u_k) = 2^{\ell_k - \log_2 e + H^*_{k+1}(\min(u_k+r_k-\ell_k, B_k)) + \lambda^*_k(u_k)} \tag{13}$$

From (12) and (13)

$$\lambda^*_k(u_k) = \log_2\left(\frac{e}{\sum_{\ell_k=0}^{u_k+r_k} 2^{\ell_k + H^*_{k+1}(\min(u_k+r_k-\ell_k, B_k))}}\right) \tag{14}$$

From (13) and (14)

$$p^*_{\ell_k}(u_k) = \frac{2^{\ell_k + H^*_{k+1}(\min(u_k+r_k-\ell_k, B_k))}}{\sum_{\alpha_k=0}^{u_k+r_k} 2^{\alpha_k + H^*_{k+1}(\min(u_k+r_k-\alpha_k, B_k))}} \tag{15}$$

From (7) and (15), we finally obtain

$$H^*_k(u_k) = \log_2\left(\sum_{\ell_k=0}^{u_k+r_k} 2^{\ell_k + H^*_{k+1}(\min(u_k+r_k-\ell_k, B_k))}\right) \tag{16}$$

Starting with $H^*_N(u_N) = 0$, we use (16) to compute the optimal flow entropy $H^*_k(u_k)$ for all u_k and then proceed backward recursively for $k = N-1, N-2, \ldots, 0$. The information utility of the GTBR is $H^*_0(0)$.

IV. Problem Formulation

For the information-theoretic comparison of a GTBR $\mathcal{R}_g(N, \mathbf{r}, \mathbf{B})$ and an STBR $\mathcal{R}_s(N', r, B)$, we impose the following conditions:

a) $\mathcal{R}_g$ and $\mathcal{R}_s$ must operate over the same number of slots.

$$N = N'$$

b) The aggregate tokens of $\mathcal{R}_g$ and $\mathcal{R}_s$ must be equal.

$$\sum_{i=0}^{N-1} r_i = Nr \tag{17}$$

c) The aggregate bucket depth of $\mathcal{R}_g$ must not exceed that of $\mathcal{R}_s$[4].

$$\sum_{i=0}^{N-2} B_i \leq (N-1)B \tag{18}$$

d) The bucket depth of $\mathcal{R}_s$ cannot be very high compared to its token increment rate.

$$2r \leq B \leq 5r \tag{19}$$

For example, in [3], the authors use $r_{max} = 6$ Mbps and $B_{max} = 12$ Mbps for their simulations.

e) The token increment rate of $\mathcal{R}_g$ at every stage must not be higher than the bucket depth of $\mathcal{R}_s$.

$$r_i \leq B \tag{20}$$

The optimal GTBR problem is:
Given an STBR $\mathcal{R}_s(N, r, B)$, determine $\mathbf{r}$ and $\mathbf{B}$ of a GTBR $\mathcal{R}_g(N, \mathbf{r}, \mathbf{B})$ so as to maximize $H^*_0(0)$ subject to (17), (18), (19) and (20).

The following result significantly reduces the search space for the optimal GTBR.

Proposition: For an optimal GTBR, equality must hold in (18), except when N is small.

Proof: We prove by contradiction. Define $g_k(u) = 2^{H^*_k(u)}$. Since $H^*_k(u) \geq 0$, $g_k(u) \geq 1$. From (16)

$$g_k(u) = \sum_{\ell=0}^{u+r_k} 2^\ell g_{k+1}(\min(u + r_k - \ell, B_k)) \tag{21}$$

$g_{N-1}(u) = 2^{u+r_{N-1}+1} - 1$ is an increasing sequence in u. Using (21), we can show that $g_k(u)$ is an increasing sequence in u $\forall$ $k = 0, \ldots, N-1$. Let ϕ_i = maximum number of tokens possible at time i. Thus, $\phi_0 = 0$ and

$$\phi_i = \min(\phi_{i-1} + r_{i-1}, B_{i-1}) \ \forall i = 1, \ldots, N-1 \tag{22}$$

If $u_i \leq \phi_i$, then we say that state u_i is reachable at stage i, otherwise it is unreachable.

Let $\mathcal{R}(N, \mathbf{r}, \mathbf{B})$ be an optimal GTBR, for which equality does not hold in (18). Then $\sum_{i=0}^{N-2} B_i \leq (N-1)B - 1$. Consider another GTBR $\mathcal{R}'(N, \mathbf{r}', \mathbf{B}')$ with $\mathbf{r}' = \mathbf{r}$ and $\mathbf{B}' = (B_0, \ldots, B_{k-1}, B_k + 1, B_{k+1}, \ldots, B_{N-2})$ for some k. $\mathbf{B}'$ satisfies (18). $g'_i(u) = g_i(u)$ $\forall$ $i = k+1, \ldots, N$ and $\forall$ u. Since $\min(u + r_k - \ell, B_k + 1) \geq \min(u + r_k - \ell, B_k)$, $g_k(\min(u+r_k-\ell, B_k+1)) \geq g_k(\min(u+r_k-\ell, B_k)) \geq 1$. If we determine a reachable state u such that $g'_k(u) > g_k(u)$, then $g'_0(0) > g_0(0)$, since the flow entropy at stage 0 is computed stage-by-stage as a linear sum of future possible flow entropies with positive weights. Thus, the problem now reduces to determining a stage k and a reachable state u such that $g'_k(u) > g_k(u)$. One of the following must hold:

Case 1 There exists an $i \in \{1, \ldots, N-1\}$ such that $\phi_i = B_{i-1} < \phi_{i-1} + r_{i-1}$.

Case 2 There is no i such that $\phi_i = B_{i-1} < \phi_{i-1} + r_{i-1}$.

Case 1: Consider the smallest i such that $\phi_i = B_{i-1} < \phi_{i-1} + r_{i-1}$. Take $k = i-1$. From (21)

$$g_{i-1}(u) = \sum_{\ell=0}^{u+r_{i-1}} 2^\ell g_i(\min(u + r_{i-1} - \ell, B_{i-1}))$$
$$= \sum_{\ell=0}^{u+r_{i-1}-B_{i-1}-1} 2^\ell g_i(B_{i-1}) + \sum_{\ell=u+r_{i-1}-B_{i-1}}^{u+r_{i-1}} 2^\ell g_i(u + r_{i-1} - \ell) \tag{23}$$

[4]Equality is present in (17) because every additional token directly translates to the permission to transmit one more bit, leading to increase in information utility. As this may not be necessarily true for bucket depth, we permit inequality in (18).

$$g'_{i-1}(u) = \sum_{\ell=0}^{u+r_{i-1}} 2^{\ell} g_i(\min(u+r_{i-1}-\ell, B_{i-1}+1)) =$$

$$\sum_{\ell=0}^{\substack{u+r_{i-1}\\-B_{i-1}-1}} 2^{\ell} g_i(B_{i-1}+1) + \sum_{\substack{\ell=u+r_{i-1}\\-B_{i-1}}}^{u+r_{i-1}} 2^{\ell} g_i(u+r_{i-1}-\ell) \quad (24)$$

(23) and (24) hold only if

$$u + r_{i-1} - B_{i-1} - 1 \geq 0 \quad (25)$$

$u = \phi_{i-1}$ is a state which is reachable in the original system as well as in the primed system and satisfies (25). Since $g_i(u)$ is an increasing sequence in u, (23) and (24) imply $g'_{i-1}(\phi_{i-1}) > g_{i-1}(\phi_{i-1})$. Consequently, $g'_0(0) > g_0(0)$.

Case 2: If no such i exists, then $B_i \geq r_0 + \cdots + r_i \;\forall\; i = 0, \ldots, N-2$. Adding and using (20)

$$\begin{aligned}\sum_{i=0}^{N-2} B_i &\geq (Nr - r_{N-1}) + (Nr - r_{N-1} - r_{N-2}) + \cdots \\ &\geq (Nr - B) + (Nr - 2B) + \cdots \quad (26) \\ &= N(N-1)r - \alpha B \quad (27)\end{aligned}$$

From (17), (19) and (20), we cannot have $r_i = B \;\forall\; i$. So, α cannot be of the order of N^2. Thus, the lower bound on $\sum_{i=0}^{N-2} B_i$ given by (26) and (27) is a loose lower bound. From (18), (19) and (27), $\sum_{i=0}^{N-2} B_i$ grows as N^2 and is upper-bounded by $5(N-1)r$, which is impossible, except when N is small. So, we discard Case 2.

From the result of Case 1, $H_0^{*'}(0) > H_0^*(0)$. So, our assumption that $\mathcal{R}$ is an optimal GTBR is incorrect. Therefore, equality must hold in (18) for every optimal GTBR. ■

V. Optimal GTBR

(N,r,B)	$\mathbf{r}^*$	$\mathbf{B}^*$	H_s (bits)	H_g^* (bits)	incr. (%)
(4,3,6)	(6 3 3 0)	(6 6 6)	20.04	20.92	4.4
(4,3,9)	(8 3 1 0) (9 2 1 0)	(8 10 9) (9 10 8)	20.10	21.44	6.7
(4,3,12)	(12 0 0 0)	(12 12 12)	20.10	21.56	7.2
(4,4,8)	(8 4 4 0)	(8 8 8)	25.08	26.04	3.8
(4,4,10)	(9 5 2 0)	(9 12 9)	25.13	26.39	5.0
(4,4,12)	(11 4 1 0)	(11 14 11)	25.14	26.59	5.8
(4,4,16)	(16 0 0 0)	(16 16 16)	25.14	26.70	6.2
(4,5,10)	(10 5 5 0)	(10 10 10)	29.91	30.92	3.4
(4,5,12)	(11 6 3 0)	(11 14 11)	29.96	31.24	4.3
(4,6,12)	(11 7 6 0) (12 7 5 0)	(11 13 12) (12 13 11)	34.60	35.66	3.1
(5,3,6)	(6 3 3 3 0)	(6 6 6 6)	25.68	26.57	3.5
(5,3,9)	(8 3 3 1 0)	(8 10 10 8)	25.88	27.33	5.6
(5,3,12)	(11 2 2 0 0)	(11 13 13 11)	25.90	27.59	6.5
(5,3,15)	(15 0 0 0 0)	(15 15 15 15)	25.90	27.64	6.7
(6,3,6)	(6 3 3 3 3 0)	(6 6 6 6 6)	31.33	32.23	2.9

TABLE I
ENTROPY-MAXIMIZING GTBR FOR GIVEN N, r AND B.

We determined the optimal GTBR by exhaustive search over the reduced search space obtained from the proposition. Our computation results are shown in Table I. H_s and H_g^* denote the information utility of the STBR $\mathcal{R}_s(N, r, B)$ and the optimal GTBR $\mathcal{R}_g(N, \mathbf{r}^*, \mathbf{B}^*)$ respectively. Based on our computations, we infer:

1) A generalized token bucket regulator can achieve *higher* information utility than a standard token bucket regulator. The increase in information utility is significant (up to 7.2%), esp. for higher values of B.
2) The optimal bucket depth sequence $\mathbf{B}^*$ is uniform or near-uniform, i.e., the standard deviation is very small compared to the mean.
3) The optimal token increment sequence $\mathbf{r}^*$ is a decreasing sequence and is not uniform.
4) For a fixed N and r:
    a) If $B = 2r$, $\mathbf{B}^*$ is always uniform and $\mathbf{r}^*$ is uniform except for the terminal values.
    b) As B increases from $2r$ to $\min(5, N)r$, the variance of $\mathbf{r}^*$ increases rapidly with a concentration of tokens in first few stages, the variance of $\mathbf{B}^*$ increases slowly, while H_g^* initially increases and then saturates at some final value. H_g^* is an increasing and concave sequence[5] in B (Figure 2).
5) For a fixed N and B, H_g^* is an increasing, highly linear and slightly concave sequence in r (Figure 3). For the STBR, Results 4b and 5 have been observed in [6].

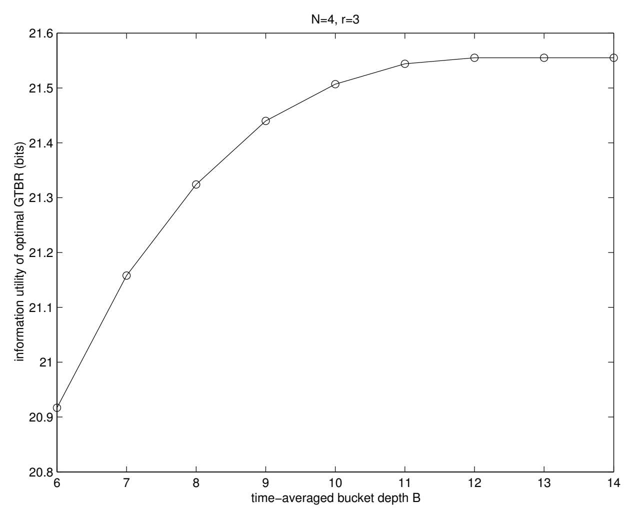

Fig. 2. H_g^* vs. B is concave.

VI. Information-theoretic interpretation

From classical information theory, if $\sum_{i=1}^{n} p_i = 1$, system entropy H increases with decreasing Kullback-Leibler distance between the given probability mass function (pmf) and the uniform pmf. H is maximized only if $p_1 = \cdots = p_n = \frac{1}{n}$. Also, maximum system entropy H^* increases with n [7]. Analogously, a GTBR can achieve higher information utility than an STBR because the pmfs of the packet lengths at each stage are more uniform and have a larger support. For a given

[5] The first-order differences form a decreasing, non-negative sequence.

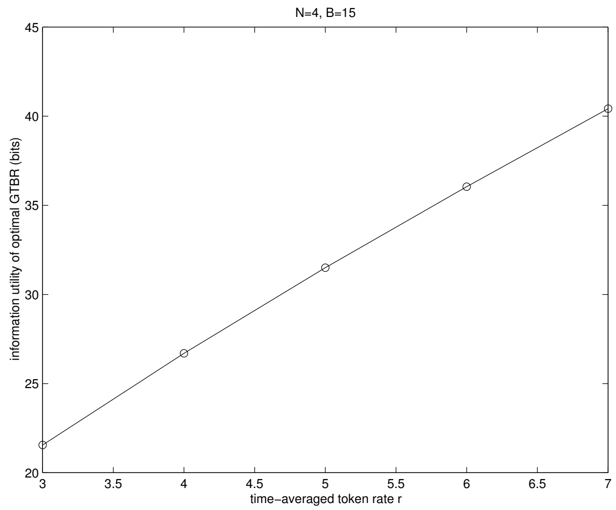

Fig. 3. H_g^* vs. r is highly linear.

$\mathbf{r}$ and $\mathbf{B}$, recall that information utility is computed recursively by (3) and (16).

We argue that $\mathbf{B}^*$ must be uniform or near-uniform for maximum information utility. If $\mathbf{B}^*$ is neither uniform nor near uniform, then $B_j = \min_i B_i$ is much smaller than B. This restricts the range of values taken by u_{j+1} and ℓ_{j+1} (from (2) and (3)). The support of packet length pmfs at stage $j+1$ is reduced, leading to lower flow entropy at stage $j+1$ and consequently lower information utility. Thus, $\mathbf{B}^*$ must be uniform or near-uniform to maximize the minimum support of the packet length pmfs *at each stage*. Our argument is corroborated by the observation that in Table I, $\min_i B_i^* = B-1$ or $\min_i B_i^* = B$ throughout.

We argue that for maximum information utility, $\mathbf{r}^*$ must be a decreasing sequence, subject to $r_i \le B_i$ for every i. If $r_i > B_i$ for any i, then a zero length packet cannot be transmitted in slot i (from (3)) and will have zero probability. This decreases the support of the packet length pmfs in slot i and leads to lower information utility. More importantly, from (7),

$$H_0^*(0) = \sum_{\ell_0=0}^{r_0} p_{\ell_0}^*(0)\Big(\ell_0 - \log_2(p_{\ell_0}^*(0)) + H_1^*(\min(r_0 - \ell_0, B_0))\Big)$$

The major contribution to information utility $H_0^*(0)$ is from the support of the packet lengths $[0, r_0]$ and the pmf of the packet lengths $\mathbf{p}_0^*(0)$, while the contribution from $H_1^*(\cdot)$ is insignificant. So, to maximize $H_0^*(0)$, r_0 should be allowed to take its maximum possible value, subject to $r_0 \le B_0$, and the pmf of the packet lengths should be close to the uniform pmf. The observation that $r_0 = B_0$ consistently in Table I corroborates this argument. Also, a high value of r_0 leads to larger supports of packet length pmfs at intermediate and later stages. Similarly, the first few elements of $\mathbf{r}^*$ tend to take large values till the aggregate tokens are exhausted. However, their contribution to $H_0^*(0)$ is not significant and equality may not hold in $r_i \le B_i$. Thus, $\mathbf{r}^*$ must be a decreasing sequence and the first few elements of $\mathbf{r}^*$ tend to take their maximum possible values, subject to $r_i \le B_i$, to achieve uniformity and larger supports of packet length pmfs *at intermediate and later stages*.

This "greedy" nature of $\mathbf{r}^*$ is evident when N and r are kept constant and B increases (Result 4b). A similar argument is applicable when N and B are kept constant and r increases (Result 5). The only difference is that a unit increase in r will necessarily increase H_g^* by at least N bits (N bits are contributed by the packet contents alone, which also explains the dominant linear variation in Figure 3), while a unit increase in B will increase H_g^* only by an amount equal to the difference in covert information. This increase in covert information is positive only if the entropy-optimal token increment and bucket depth sequences $(\mathbf{r}^*, \mathbf{B}^*)$ result in larger support and more uniformity for the packet length pmfs. Indeed, when B increases beyond the maximum number of tokens possible at any stage ($\max_i\{\phi_i\}$), clamping the residual number of tokens at every stage becomes ineffective and the system behaves as if bucket depth constraints were not imposed at all (Figure 2).

VII. Discussion

In this paper, we have considered a problem where a source whose traffic is regulated by a generalized token bucket regulator, seeks to maximize the entropy of the resulting flow. The source can achieve this by recognizing that the randomness in packet lengths acts as a covert channel in the network and sizing its packets appropriately. We have formulated the problem of computing the GTBR with maximum information utility in terms of constrained token increment and bucket depth sequences. A GTBR can achieve higher information utility than a standard IETF token bucket regulator. Finally, we have information-theoretically interpreted the observation that an entropy-maximizing GTBR always has a near-uniform bucket depth sequence and a decreasing token increment sequence.

Our results show the existence of upper bounds on the entropy of regulated flows. It would be interesting to construct source codes which come close to this bound. The development of a rate-distortion framework for a generalized token bucket regulator is currently under investigation.

References

[1] S. Keshav, *An Engineering Approach To Computer Networking*. Addison Wesley Longman, 1997.

[2] P. Shah and A. Karandikar, "Optimal packet length scheduling for regulated media streaming," *IEEE Communications Letters*, vol. 7, pp. 409–411, August 2003.

[3] S. Giordano and J.-Y. L. Boudec, "On a class of time varying shapers with application to the renegotiable variable bit rate service," *Journal on High Speed Networks*, vol. 9, pp. 101–138, June 2000.

[4] R. G. Gallager, "Basic limits on protocol information in data communication networks," *IEEE Transactions on Information Theory*, vol. 22, pp. 385–398, July 1976.

[5] V. Anantharam and S. Verdu, "Bits through queues," *IEEE Transactions on Information Theory*, vol. 42, pp. 4–18, January 1996.

[6] P. Shah and A. Karandikar, "Information utility of token bucket regulator," *Electronics Letters*, vol. 39, pp. 581–582, March 2003.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 1991.