Packet Multiplexing of Stream Sessions: Stochastic Analysis

December 2, 2002

1 Motivation for Stochastic Models

Provisioning for worst case wastes system resources

Consider n statistically independent sources $A_1(t), \ldots, A_n(t)$

Each source has envelope $\min(Rt, \sigma + \rho t)$

Envelope for superposition $(A(t) = \sum_{i=1}^{n} A_i(t))$ is $\min(nRt, n\sigma + n\rho t)$

The bandwidth required to guarantee the worst case delay of $d_{\max}$ is nC if capacity C is required for guaranteeing $d_{\max}$ for just one source.

In reality, $A(t)$ is significantly smaller than $E(t)$. Thus providing nC is redundant.

- Let $B_i(t)$ be the number of bits delivered by the ith source at time t
- $B(t) = \sum_{i=1}^{n} B_i(t)$
- $A(t) = \int_0^t B(u)du$

Let $B_i(t) = R$ for T_{on} and $B_i(t) = 0$ for T_{off}

Let $T_{\text{on}} = T_{\text{off}}$. Average bit rate is $EB_i(t) = R/2$

$EB(t) = nR/2$,

$$EA(t) = E\int_0^t B(t)dt = \int_0^t EB(t)dt = nR/2dt$$

Thus $EA(t)$ can be significantly less than its envelope depending on the values of R, ρ.

$A_i(t)$ has a standard deviation upper bounded by $\sigma_{\max}$

For large n,

$$P\left(A(t) > nR/2t + x\sqrt{n}\sigma_{\max}\right) < 1 - \phi(x)$$

where $\phi(x)$ is the cumulative distribution function of the standard normal distribution

For large n with a high probability $A(t)$ is significantly less than he deterministic envelope.

2 Example Stochastic Models

- Discrete time Batch Markovian

  System represented by a markov chain with states $1, \ldots N$

  k packets are generated every unit time in state k

- Markov modulated Poisson

  System represented by a markov chain with states $1, \ldots N$

  packets generated as a Poisson process with rate λ_k when system is in state k

3 Additional Notations

- $\mathcal{A}(t)$ is the number of packet arrivals in the time interval $[0, t]$
- $\mathcal{D}(t)$ is the number of packet departures in the time interval $[0, t]$
- W_k is the total time spent in the system by the kth packet

- $N(t)$ is the number of packets in the system at time t; $N(t) = \mathcal{A}(t) - \mathcal{D}(t)$

4 Littles Theorem and Brumelles Theorem

Theorem 1 *(***Littles Law***)*

1. *If, for a sample path denoted by ω, $\frac{\mathcal{A}(t)}{t} \rightarrow \lambda(\omega)$ and $\frac{1}{n}\sum_{j=1}^{n} W_j \rightarrow W(\omega)$, then $\lim_{t\rightarrow\infty} \frac{1}{t}\int_0^t N(u)du = N(\omega)$ exists, and $N(\omega) = \lambda(\omega)W(\omega)$*

2. *If the limits, for the various sample paths ω are equal to the constants λ and W with probability 1 then $\int_0^t N(u)du = N(\omega) = \lambda W$ with probability 1 $(N = \lambda W)$*

Littles law can be used to show that the expected waiting time of a packet in a system is the same for a large class of scheduling policies.

Assumptions:

1. The packet lengths $(L_k, k \geq 1)$ are independently and identically distributed

2. The scheduler is work conserving and the service order does not depend on packet size

3. Only one packet is served at a time

FCFS, LCFS (non-preemptive) are in this class

Shortest service first strategy is not in this class

Assume that the time averages exist with probability 1 (follows if the system is stable)

Choose any two scheduling strategies in this class (L_i, W_i)

Arrival epochs are same for both

Size of the kth served packet can be assumed to be the same for both without loss of generality

The random variables which represent the number of packets in the system have the same statistics in both ($L_1 = L_2$)

The result follows since $L_1 = \lambda W_1$ and $L_2 = \lambda W_2$.

Generalization of Littles Law: **Brummelles Law**

Associate a function $f_k(t)$ with each packet k at time t

Example: $f_k(t)$ can be the number of untransmitted bits associated with a packet k if the packet is in the system at time t, otherwise $f_k(t) = 0$

- Let G_j be the cumulative value of the function for the jth packet, $G_j = \int_0^\infty f_j(u)du$

- Let $H(t)$ be the total value of this function for all packets at time t, $H(t) = \sum_{j=1}^{\infty} f_j(t)$

$$
\begin{aligned}
G(\omega) &= \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} G_j(\omega) \\
H(\omega) &= \lim_{t\to\infty} \frac{1}{t} \int_0^t H(\omega, u)du
\end{aligned}
$$

Theorem 2 *(***Brummelles Law***)*

1. *If, for a sample path denoted by* ω, $\frac{\mathcal{A}(t)}{t} \to \lambda(\omega)$, $f_k(t) = 0$ *if the packet* k *is not in the system, and* $\frac{1}{n}\sum_{j=1}^{n} G_j \to G(\omega)$, *then* $\lim_{t\to\infty} \frac{1}{t}\int_0^t H(\omega, u)du = H(\omega)$ *exists, and* $H(\omega) = \lambda(\omega)G(\omega)$

2. *If the limits, for the various sample paths* ω *are equal to the constants* λ *and* G *with probability* 1 *then* $\int_0^t H(\omega, u)du = H(\omega) = \lambda W$ *with probability* 1 *(*$N = \lambda W$*)*

Example application: can be used to compute the expected total number of bits in the system

size of kth packet is L_k, i.i.d

$f_k(t)$ is the residual number of bits of the kth packet,

$f_k(t) = L_k$ before kth packet starts service, $f_k(t)$ linearly decreases at rate C during service,

$f_k(t) = 0$ after the kth packet completes service

$H(t)$ is the total number of bits in the system at time t

Let U_k be the wait time before the kth packet is served

From Brumelles theorem,

$$EH = \int_0^t H(\omega, u)du = \lambda \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n} (L_k U_k + \frac{1}{2}\frac{L_k^2}{C}$$

For FCFS U_k is the total number of bits in the system when the kth packet arrives $(H(a_k))$ divided by C

By PASTA $EU_k = EH/C$

It follows that

$$EH = \lambda EL \frac{EH}{C} + \frac{1}{2}\frac{L^2}{C}$$

It follows that

$$EH = \frac{\rho}{1-\rho}\frac{1}{2C}\frac{EL^2}{EL}, \text{ where } \rho = \lambda\frac{EL}{C}$$

5 Statistical Multiplexing Gain: Large Deviation Bound

Terminology

- m sources
- $B_k^{(i)}$ is the number of bits input from the ith source in the kth slot

Assumption: No Buffer, Bits which can not be served in a slot are lost

Motivation:

- Distribution of stationary occupancy for buffered multiplexer can be approximated by that for bufferless multiplexers

- Some systems are buffer less or have very low buffer, since buffers can be expensive, e.g., optical buffers

Performance Measure:

- Long run fraction of slots where data is lost?

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} I_{\sum_{k=1}^{m} B_k^{(i)} > C}$$

- Long run fraction of data lost

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n} \left(\sum_{i=1}^{m} B_k^{(i)} - C\right)^+}{\sum_{k=1}^{n} \left(\sum_{i=1}^{m} B_k^{(i)}\right)^+}$$

Assumption: $\{B_k^{(i)}, k \geq 0\}$ is stationary, ergodic for each i

$\{B_k^{(1)}, \ldots B_k^{(m)}\}$ are mutually independent for each k

It follows that $\{\vec{B}_k := \left(B_k^{(1)}, \ldots, B_k^{(m)}\right) k \geq 1\}$ is stationary, ergodic.

¿From Birkhoffs ergodic theorem, with probability 1,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} I_{\sum_{k=1}^{m} B_k^{(i)} > C} = EI_{\sum_{k=1}^{m} B_k^{(i)} > C} = P\left(\sum_{k=1}^{m} B_k^{(i)} > C\right)$$

Similarly with probability 1,

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n} \left(\sum_{i=1}^{m} B_k^{(i)} - C\right)^+}{\sum_{k=1}^{n} \left(\sum_{i=1}^{m} B_k^{(i)}\right)^+} = \frac{E\left(\sum_{i=1}^{m} B_1^{(i)} - C\right)^+}{E\left(\sum_{i=1}^{m} B_1(i)\right)^+}$$

Long run performance measures depend only on the marginal distribution of the data arrival process.

For markovian data arrival process, need to consider only the stationary distribution of the arrival process

Let $B_k^{(i)}$ be a two state markov process for each i.

The states are (a)talkspurt (mean duration: σ, emits one fixed length packet each slot) and (b) silence (mean duration: τ)

Thus $(\sum_{i=1}^m B_1(i))$ is Binomial$(m, \frac{\sigma}{\sigma+\tau})$ distributed

$$P\left(\sum_{i=1}^{m} B_1^{(i)} > C\right) = \sum_{k=C+1}^{m} (m\text{choose}k) \left(\frac{\sigma}{\sigma+k}\right)^k \left(1 - \frac{\sigma}{\sigma+k}\right)^{m-k}$$

Loss probability can be controlled if $P\left(\sum_{k=1}^m B_1^{(i)} > C\right)$ is small

C must be chosen such that $P\left(\sum_{k=1}^m B_1^{(i)} > C\right) < \epsilon$

Consider the two-state (on-off) sources with the bit rate during talk spurt being R

If $C < R$ then $P\left(B_1^{(i)} > C\right) = \sigma/(\sigma = \tau) > \epsilon$ for small ϵ

Thus for $m = 1$ for negligible loss probability $C = R$.

What happens for larger number of sources?

5.1 Analysis using Chernoff's bound and Cramer's Theorem

Consider i.i.d. random variables $\{X_1, X_2, \ldots, \}$.

Let $a > EX_1$

Interested in bounding or approximating

$$P(\sum_{i=1}^{n} X_i \geq na) = P(\sum_{i=1}^{n} (X_i - a) \geq 0)$$

For any random variable Y and for $\theta \geq 0, I_{Y \geq 0} \leq e^{\theta Y}$

Taking expectations on both sides

$$P(Y \geq 0) \leq Ee^{\theta Y}$$

$$\begin{aligned} P(\sum_{i=1}^{n} (X_i - a) \geq 0) &\leq Ee^{\theta\left((\sum_{i=1}^n (X_i - a)\right)} \\ &= \left(e^{-a\theta} Ee^{X_1 \theta}\right)^n \\ &= e^{-n(a\theta - \ln M(\theta))} \end{aligned}$$

where $M(\theta) = Ee^{X_1\theta}$

$$\begin{aligned} P(\sum_{i=1}^{n}(X_i - a) \geq 0) &\leq e^{\inf_{\theta\geq 0}(-n(a\theta - \ln M(\theta)))} \\ &= e^{-n\sup_{\theta\geq 0}(a\theta - \ln M(\theta))} \end{aligned}$$

Lemma 1 *(***Jensens inequality** *For any convex function, $f(x)$, $Ef(x) \geq f(Ex)$*

$e^{\theta X_1}$ is convex, thus $Ee^{\theta X_1} \geq e^{\theta EX_1}$

$$\begin{aligned} a\theta - \ln M(\theta) &= a\theta - \ln Ee^{\theta X_1} \\ &\leq a\theta - \ln e^{\theta EX_1} \\ &= a\theta - \theta EX_1 \end{aligned}$$

Since $a > EX_1$ for $\theta \leq 0$ $a\theta - \ln M(\theta) \leq 0$ and

$a\theta - \ln M(\theta) \leq 0$ for $\theta = 0$

It follows that for $\theta \leq 0$ $a\theta - \ln M(\theta)$ attains its maximum value at $\theta = 0$

$$sup_\theta\,(a\theta - \ln M(\theta)) = sup_{\theta\geq 0}\,(a\theta - \ln M(\theta))$$

$$l(a) = sup_{\theta\geq 0}\,(a\theta - \ln M(\theta))$$

$$P(\sum_{i=1}^{n} X_i \geq na) \leq e^{-nl(a)}$$

$$\text{Equivalently}\frac{1}{n}\ln P(\sum_{i=1}^{n} X_i \geq na) \leq -l(a)$$

Properties of the exponent $l(a)$

- $l(a) \geq 0$
- $l(a)$ is convex
- $l(EX_1) = 0$

Example: The Two State Markov Source

$$B^{(i)} = \begin{cases} R & \text{w.p.} p \\ 0 & \text{w.p.} 1-p \end{cases}$$

We wish to determine c such that

$$P\left(\sum_{i=1}^{m} B^{(i)} \geq nc\right) \leq e^{-\delta}$$

(Example upper bounds: $e^{-9} \approx 10^{-4}$ and $e^{-12} \approx 10^{-6}$)

$$M(\theta) = \ln\left(1 - p + pe^{R\theta}\right)$$

$$l(a) = \sup_{\theta \geq 0}\left(\theta a - \ln\left(1 - p + pe^{R\theta}\right)\right)$$

We need $l(c) = \delta/n$

For $\delta = 12$, $n = 100$ $R = 1$ $p = 0.4$ $c = 0.643$

Thus c is significantly less than the peak rate of R on account of statistical multiplexing

Admission Control for requisite QoS: $l(c) = \delta/n$

$C = nc$

$\frac{\delta}{C/c} = l(c)$

Admit no more than $\lfloor C/c \rfloor$ sessions

6 Analysis with Arbitrary Buffering

Characterize Stationary queue length

$$\begin{aligned} X_k &= \max\left(X_{k-1} + B_k - C, 0\right) \\ X_n &= \max_{0 \leq k \leq n}\left(A_n - A_k - (n-k)C\right) \end{aligned}$$

Assume that the arrival process is for the entire real line

Consider the continuous time process $X(t)$

Start the system with an empty buffer at $-t$

Denote the amount of traffic in the system at k as $U_{-t}(k)$, consider $U_{-t}(0)$

$$\begin{aligned} U_{-t}(0) &= \sup_{-t \le s \le 0} (A(0) - A(s) - C(0-s)) \\ &= \sup_{0 \le s \le t} (A(0) - A(-s) - Cs) \end{aligned}$$

$X(t)$ and $U_{-t}(0)$ have the same distribution.

We need the distribution of $X(t)$ as $t \to \infty$

As t increases $U_{-t}(0)$ converges possibly to ∞ (sup is over an increasing set). Denote the limit as U_∞

$$U_\infty = sup_{s \ge 0} (A(0) - A(-s) - Cs)$$

$$X(t) \to_{\text{dist}} \sup_{s \ge 0} (A(0) - A(-s) - Cs)$$

(Limit is finite with probability 1 if $\lim_{t \to \infty} \frac{A(t)}{t} = A < C$)

Similarly for discrete process

$$X_n \to_{\text{dist}} \max_{0 \le k \le n} (A_0 - A_{-k} - kC)$$

As $n \to \infty$

$$X_n \to_{\text{dist}} \sup_{k \ge 0} (A_0 - A_{-k} - kC)$$

Let $S_{-k} = A_0 - A_{-k} - kC$

$$P(X_n > x) = P(\max_{0 \le k \le n} S_{-k} > x)$$

$$
\begin{aligned}
&= P(\cup_{0\le k\le n}\{S_{-k} > x\}) \\
&\le \sum_{k=0}^{n} P(S_{-k} > x) \\
&\le \sum_{k=0}^{n} Ee^{\theta(S_{-k}-x)} \\
&\le e^{-\theta x}\sum_{k=0}^{n} Ee^{\theta(S_{-k})} \qquad (1)
\end{aligned}
$$

$$\text{Let } M(\theta) = \lim_{n\to\infty}\frac{1}{n}\ln Ee^{\theta(A_n-A_0)}$$

For large n,

$$Ee^{\theta(A_n-A_0)} \approx e^{nM(\theta)}$$

Since A_k has stationary increments, $Ee^{\theta(A_n-A_0)} = Ee^{\theta(A_k-A_0)}$

$$\text{For large } k, Ee^{\theta(S_{-k})} \approx e^{k(M(\theta)-\theta C)}$$

Let θ be such that $M(\theta)/\theta < C$

Terms in (1) are geometrically decreasing and thus the infinite summation is upper bounded, the bound is K

$$P\left(X_n > x\right) \le Ke^{-\theta x}$$

Since $P\left(X_n > x\right)$ converges to $P\left(X > x\right)$,

$$P\left(X > x\right) \le Ke^{-\theta x}$$

$$\frac{1}{x}\ln P\left(X > x\right) \le \ln K/x - \theta$$

For large x, if $M(\theta)/\theta < C$

$$\frac{1}{x}\ln P\left(X > x\right) \le -\theta$$

Thus the tail probability decay is exponential for buffer occupancy if $C > M(\theta)/\theta$

Thus $M(\theta)/\theta$ is the equivalent bandwidth

Two state Markov source

$M(\theta) = \ln(1 - p + pe^{R\theta})$

Properties of $M(\theta)$

- Effective bandwidth is between the average and the peak rate

- Effective bandwidth of n sources is the sum of that for each source

The additive properties shows that allocating resources as per effective bandwidth may be conservative

Guerin, Ahmadi, Narshineh approach: For n sources, allocate capacity which equals the minimum of that required under bufferless assumption and the sum of the effective bandwidth of each