Performance analysis of a fluid GPS multiplexer

Ernst Nordström

October 22, 2009

1 Introduction

This paper describes a method for calculating key performance measures of class-based GPS systems offered many fluid Markov sources. The task is decoupled into a set of single queue problems, which are analysed using a lower bound model outlined by Presti, Zhang amd Towsley [10]. The upper bound model presented in the same reference is omitted for simplicity reasons.

2 Traffic assumption

We consider a class-based GPS system that is offered traffic from K QoS classes. The capacity of the link is denoted C [Mbps]. The weight of queue i is denoted ϕ_i and the buffer capacity is denoted B_i [Mbit]. The i-th class, $i \in I = \{1, \ldots, K\}$, is characterized by the following:

- Number of calls: N_i,
- Peak bit rate requirement: p_i [Mbps],
- OFF to ON state transition rate: α_i [s^{-1}],
- ON to OFF state transition rate: β_i [s^{-1}],

In order to have a stable system we assume the total offered mean rate is is less than the link capacity:

$$\sum_{i \in I} N_i m_i < C, \tag{1}$$

where $m_i = p_i \frac{\alpha_i}{\alpha_i + \beta_i}$.

3 Model

Each buffer is modeled as a fluid reservoir with a hole in the bottom and arriving information is modeled as a fluid running into the reservoir. We model the input from class i as a Markov modulated fluid source characterized by the pair $(\mathbf{M}^{(i)}, \lambda^{(i)})$, with $\mathbf{M}^{(i)}$ being the irreducible generator of the Markov chain on state space $S^{(i)} = \{1, \cdots, N_i\}$ and $\lambda^{(i)}$ the rate vector. Let the stochastic variables Σ and Q_i denote the stationary state of the system fluid process, and i-th queue length, respectively.

The sample path description of the buffer dynamics is:

$$\frac{dQ_i}{dt} = \begin{cases} r_i(t) - s_i(\phi, t) & \text{if} \quad Q_i(t) > 0 \\ (r_i(t) - s_i(\phi, t))^+ & \text{if} \quad Q_i(t) = 0 \\ (r_i(t) - s_i(\phi, t))^- & \text{if} \quad Q_i(t) = B_i \end{cases} \tag{2}$$

where generally $(x)^+ = \max(x, 0)$ and $(x)^- = \min(x, 0)$ denotes the positive and negative part of x, respectively, $r_i(t)$ and $s_i(\phi, t)$ denote the arrival rate and the service rate, respectively, at time t.

The *arrival rate* at time t is $r_i(t) = k_i(t)p_i$. The *service rate* at time t is:

$$s_i(\phi, t) = \begin{cases} g_i + \frac{\phi_i}{\sum_{i\in J} \phi_i \mathbf{1}_{Q_i(t)>0}} \sum_{i\in J\setminus\{j\}} r_i(t), & Q_i(t) > 0, \\ g_i, & Q_i(t) = 0, \end{cases} \tag{3}$$

where $r_i(t) = (g_i - r_i(t))^+ \mathbf{1}_{Q_i(t)=0}$ denotes the *residual service rate*, and $g_i = \phi_i c$ denotes the guaranteed minimum service rate. The *departure rate* represents the service rate actually used.

When the queue is backlogged,i.e. $Q_i(t) > 0$, class i will make full use of the available service rate. When the queue is empty, i.e. $Q_i = 0$, and input rate $r_i(t)$ is less than the minimum service rate g_i, the excess (residual) service rate is shared among the other classes with backlogged queues. The sharing of residual service rate among the backlogged classes makes the system of differential equations coupled. To simplify the analysis, the system of differential equations can be *decoupled* into K independent differential equations by imposing assumptions on the queue states. The queue length $\tilde{Q}_i(t)$ in a decoupled differential equation is either a lower or upper bound of the real queue length $Q_i(t)$.

4 Analysis

4.1 Statistical multiplexing system with modulated service process

The decoupled multiplexing system with modulated service process is very similar to the producer-consumer system studied by Mitra [7]. The queue dynamics follows the following sample path equation

$$\frac{dQ}{dt} = \lambda^{(a)}_{\Sigma^{(a)}(t)} - [C - \lambda^{(s)}_{\Sigma^{(s)}(t)}] \tag{4}$$

where $(\lambda^{(a)}_{\Sigma^{(a)}(t)}, \lambda^{(s)}_{\Sigma^{(s)}(t)})$ is a pair of random variables representing the states of the arrival process and the service process at time t. By considering (4) it is clear that the system is equivalent to a statistical multiplexing system with constant service rate C, the input of which is produced by the superposition of the Markov modulated fluid sources $(\mathbf{M}^{(a)}, \lambda^{(a)})$ and $(\mathbf{M}^{(s)}, \lambda^{(s)})$.

4.2 Lower bound model

Under the assumption that the queue for class $j \in I\setminus\{i\}$ is always empty, the service process of class i becomes

$$\tilde{s}_i(t) = \begin{cases} g_i + \sum_{j\in I\setminus\{i\}} r_j(t) = C - \sum_{j\in I\setminus\{i\}} \tilde{r}_j(t), & Q_i(t) > 0, \\ g_i, & Q_i(t) = 0, \end{cases} \tag{5}$$

where $\tilde{r}_j(t) = \min(r_j(t), g_j)$. In reality, the queue of class j may sometime be busy, resulting in zero residual service rate. Hence, the service process $\tilde{s}_i(t)$ provides a lower bound of the queue length. The arrival process for class i is described by the pair $\mathbf{M}^{(i)}, \lambda^{(i)}$ and the modulating process by the pair $(\tilde{\mathbf{M}}^{(i)}, \tilde{\lambda}^{(i)})$. According to the expression for service rate, it follows that $\tilde{\mathbf{M}}^{(i)} = \mathbf{M}^{(i)}$ and $\tilde{\lambda}^{(i)} = \{\tilde{\lambda}^{(i)}_1, \cdots, \tilde{\lambda}^{(i)}_{N_i}\}$ where $\tilde{\lambda}^{(i)}_s = \min(\lambda^{(i)}_s, g_i)$, $1 \le s \le N_i$.

4.3 System equation

Let $\mathbf{F}_i(x) = \{F_i(\mathbf{k}, x)\}_{\mathbf{k}\in\mathbf{S}}$ denote the stationary buffer distribution column vector, where $F_i(\mathbf{k}, x) = \Pr(\Sigma = \mathbf{k}, Q_i \leq x), \mathbf{k} \in \mathbf{S}, 0 \leq x \leq B_i$. The (decoupled) lower bound model fullfils the following Kolmogorov differential equation [1]:

$$\mathbf{D}_i \frac{\mathrm{d}}{\mathrm{dx}} \mathbf{F}_i(x) = \mathbf{M}_i \mathbf{F}_i(x), 0 < x < B_i, \tag{6}$$

where $\mathbf{D}_i$ denotes the drift matrix and $\mathbf{M}_i$ the generator matrix.

The drift matrix is defined by $\mathbf{D}_i = \mathbf{\Lambda}_i^{(a)} \ominus \mathbf{\Lambda}_i^{(s)}$ where $\ominus$ denotes the Kronecker difference, $\mathbf{\Lambda}_i^{(a)} = \mathrm{diag}(\lambda^{(i)})$ denotes the arrival rate matrix and $\mathbf{\Lambda}_i^{(s)} = \mathrm{diag}(\tilde{\lambda}^{(i)})$ denotes the service rate matrix. The generator matrix for the superposition of the arrival and service fluid process is defined as $\mathbf{M}_i = \mathbf{M}^{(i)} \oplus \tilde{\mathbf{M}}^{(i)}$, where $\oplus$ denotes the Kronecker sum.

5 Solution

The solution to the differential equation for class i in the decoupled GPS system is given by the spectral expansion:

$$\mathbf{F}_i(x) = \sum_{\mathbf{k}\in\mathbf{S}} a_i(\mathbf{k}) \exp(z_i(\mathbf{k})x) \varphi_i(\mathbf{k}) \tag{7}$$

where $a_i(\mathbf{k})$ are scalar cofficients found from boundary conditions, and $z_i(\mathbf{k})$, $\varphi_i(\mathbf{k})$ are eigenvalues and (left) eigenvectors, respectively, found by solving the generalized eigenvalue problem $z_i \varphi_i \mathbf{D}_i = \varphi_i \mathbf{M}_i$.

The coefficients $a_i(\mathbf{k})$ can be found in at least two ways. One method finds the set of exact coefficients $a_i(\mathbf{k}, B_j)$ by solving a linear equation system formulated from the boundary condition [5]:

$$\sum_{\mathbf{k}\in\mathbf{S}} a_i(\mathbf{k}, B_i) \varphi_i(\mathbf{k}) = 0, \quad r_i(k_i) > s_i(\phi, \tilde{k}_i) \tag{8}$$

$$\sum_{\mathbf{k}\in\mathbf{S}} a_i(\mathbf{k}, B_i) \exp(z_i(\mathbf{k})B_i) \varphi_i(\mathbf{k}) = \pi_i(\mathbf{k}), \quad r_i(k_i) < s_i(\phi, \tilde{k}_i) \tag{9}$$

where k_i,$\tilde{k}_i$ denotes the state of the arrival and service process for class i, and $\pi_i(\mathbf{k})$ denotes the overall probability of sources being in state $\mathbf{k}$. This probability is found from the multi binomial distribution:

$$\pi_i(\mathbf{k}) = \prod_{j\in J} \binom{N_i}{k_i} \left(\frac{\alpha_i}{\alpha_i + \beta_i}\right)^{k_i} \left(\frac{\beta_i}{\alpha_i + \beta_i}\right)^{N_i - k_i} \tag{10}$$

For large B_i, the exponential function $\exp(z_i(\mathbf{k})B_i)$ in (9) may overflow for positive eigenvalues. As only $a_i(\mathbf{k}, B_i) \exp(z_i(\mathbf{k})B_i)$ rather than $a_i(\mathbf{k}, B_i)$ is needed to compute packet loss, this problem is overcome by solving for $a_i(\mathbf{k}, B_i) \exp(z_i(\mathbf{k})B_i)$ instead of $a_i(\mathbf{k}, B_i)$ in (8) and (9). In practice, this means removing $\exp(z_i(\mathbf{k})B_i)$ from (9) and inserting $\exp(-z_i(\mathbf{k})B_i)$ in (8), thus turning the potential overflow into a potential underflow [11].

Individual approximate coefficients $a_i(\mathbf{k}, 0)$ can be calculated directly from the diagonal elements of the drift matrix $\mathbf{D}_i$, the eigenvectors $\varphi_i(\mathbf{k})$, and the overall probabilities $\pi_i(\mathbf{k})$ as follows [2]:

$$a_i(\mathbf{k},0)=\frac{\sum_{\mathbf{n}:l_i(\phi,\mathbf{n})<0} d_i(\mathbf{n})\varphi_i(\mathbf{k})(\mathbf{n})}{\sum_{\mathbf{n}\in\mathbf{S}} d_i(\mathbf{n})\pi_i(\mathbf{n})^{-1}[\varphi_i(\mathbf{k})(\mathbf{n})]^2} \tag{11}$$

where $l_i(\phi,\mathbf{k})=r_i(k_i)-s_i(\phi,\tilde{k}_i)$ denotes the fluid loss rate.

6 Fluid performance measures

In this section we present performance measures for the lower-bound fluid GPS model.

6.1 Buffer overflow probability

6.1.1 Exact formula

The buffer overflow probability $G_i(x)=\mathrm{P}(Q_i>x)$ can be written as

$$G_i(x)=1-\sum_{\mathbf{k}\in\mathbf{S}} F_i(\mathbf{k},x) \tag{12}$$

6.1.2 Asymptotic approximation

The lower bound procedure proposed in the previous section assume a complete analysis of the decomposed bounding system. As the dimension of the system grows, this becomes increasingly expensive in terms of computational cost, especially when performed on-line. To remedy the situation, we use the refined effective bandwidth approximation (REB). Namely, we approximate the queue length of class i by the formula:

$$G_i(x)\approx L_i\exp(z_i(\mathbf{0})x) \tag{13}$$

where L_i is an appropriate prefactor and $z_i(\mathbf{0})$ is the dominant eigenvalue of the system matrix in question. As when $x=0$, the approximation yields $\mathrm{P}r(Q_i>0)\approx L_i$, thus L_i approximates the probability that the buffer is not empty. Here, we calculate L_i by simply computing the probability that the input rate $r_i(k_i)$ exceeds the service rate $s_i(\phi,\tilde{k}_i)$ The dominant eigenvalue is the unique solution to the equation:

$$g^{(a)}(z)+g^{(s)}(z)=C \tag{14}$$

where $g^{(i)}$ is the effective bandwidth of source $i\in\{a,s\}$, i.e. the maximum real eigenvalue of the matrix $(\Lambda^{(i)}-\frac{1}{z}\mathbf{M}^{(i)})$. Computing the maximum real eigenvalue of irreducible, essentially nonnegative matrices is almost identical to the task of computing the Perron-Frobenius eigenvalues of irreducible, nonnegative matrices, and various standard, simple techniques are available. In the special case of the arrival process being a superposition of ON/OFF sources with exponentially distributed activity periods, $g^{(a)}(z)$ in (14) is given by:

$$g^{(a)}(z)=\frac{N_i}{2z}\left[zp_i+\beta_i+\alpha_i-\sqrt{(zp_i+\beta_i-\alpha_i)^2+4\alpha_i\beta_i}\right] \tag{15}$$

6.2 Fluid loss probability

6.2.1 Exact formula

The overall fluid loss probability p_{loss} is defined as the fraction lost information to offered information. Loss can only take place when the buffer is at maximum and the input rate is larger than the output rate. Therefore we get:

$$p_{\text{loss},i} = \frac{1}{N_i m_i} \sum_{\mathbf{k}: l_i(\phi,\mathbf{k})>0} l_i(\phi,\mathbf{k}) u_i(\mathbf{k}) \tag{16}$$

where $u_i(\mathbf{k}) = \pi_i(\mathbf{k}) - \lim_{x\to B_i} F_i(\mathbf{k},x)$.

6.2.2 Segarra-Haro approximation

Segarra and Haro have proposed following upper bound for the fluid loss probability [9]:

$$p_{\text{loss},i} = \frac{A_{1,i}(\phi)}{1+(A_{2,i}(\phi)/A_{3,i}(\phi))\exp(-z_i(\mathbf{0})x)} \tag{17}$$

where coefficients $A_{1,i}(\phi)$,$A_{2,i}(\phi)$, and $A_{3,i}(\phi)$ for a GPS system are defined by:

$$\begin{aligned} A_{1,i}(\phi) &= \frac{1}{N_i m_i} \sum_{\mathbf{k}\in\mathbf{S}} l_i(\phi,\mathbf{k})\pi_i(\mathbf{k}) \\ A_{2,i}(\phi) &= \sum_{\mathbf{k}: l_i(\phi,\mathbf{k})<0} l_i(\phi,\mathbf{k})\pi_i(\mathbf{k}) \\ A_{3,i}(\phi) &= \sum_{\mathbf{k}: l_i(\phi,\mathbf{k})>0} l_i(\phi,\mathbf{k})\pi_i(\mathbf{k}) \end{aligned} \tag{18}$$

6.2.3 Kvols-Blaabjerg approximation

Kvols and Blaabjerg have proposed the following upper bound for the fluid loss probability [6]:

$$p_{\text{loss},i} = A_i(\phi)\exp(z_i(\mathbf{0})x) \tag{19}$$

where the coefficient $A_i(\phi)$ denotes the fluid loss probability in a bufferless GPS system:

$$A_i(\phi) = \frac{1}{N_i m_i} \sum_{\mathbf{k}: l_i(\phi,\mathbf{k})>0} l_i(\phi,\mathbf{k})\pi_i(\mathbf{k}) \tag{20}$$

6.3 Fluid mean delay

6.3.1 Exact formula

The distribution of queue length for class i is given by

$$\Pr(Q_i \le x) = \sum_{\mathbf{k}\in\mathbf{S}} F_i(\mathbf{k},x) \tag{21}$$

The mean queue length is given by

$$\overline{Q}_i = \int_0^B x \frac{\mathrm{dPr}(Q_i \leq x)}{\mathrm{d}x} \mathrm{d}x = \sum_{\mathbf{k} \in \mathbf{S}} \left\{ [xF_i(\mathbf{k}, x)]_0^{B-} - \int_0^{B-} F_i(\mathbf{k}, x)\mathrm{d}x + Bu_i(\mathbf{k}) \right\} \tag{22}$$

where $F_i(\mathbf{k}, B-)$ is defind as $\lim_{x \to B} F_i(\mathbf{k}, x)$. The mean queueing delay for class i, $\overline{W}_i$, is obtained from Little's formula:

$$\overline{W}_i = \frac{\overline{Q}_i}{N_i m_i (1 - p_{loss}(i))} \tag{23}$$

6.3.2 Asymptotic approximation

From the dominated eigenvalue $z_i(\mathbf{0})$ and its eigenvector $\varphi_i(\mathbf{0})$ the buffer distribution can be approximated by

$$\tilde{\mathbf{F}}_i(\mathbf{k}, x) = a_i(\mathbf{0}) \exp(z_i(\mathbf{0})x)\varphi_i(\mathbf{0})(\mathbf{k}) \tag{24}$$

Inserting this expression in the formula for mean queue length gives:

$$\overline{Q}_i = \int_0^B x \frac{\mathrm{dPr}(Q_i \leq x)}{\mathrm{d}x} \mathrm{d}x \approx \sum_{\mathbf{k} \in \mathbf{S}} \left\{ [x\tilde{F}_i(\mathbf{k}, x)]_0^{B-} - \int_0^{B-} \tilde{F}_i(\mathbf{k}, x)\mathrm{d}x + B\tilde{u}_i(\mathbf{k}) \right\} \tag{25}$$

which finally is inserted in Little's formula to yield the approximate mean fluid delay.

References

[1] D. Anick, D. Mitra, M. Sondhi, "Stochastic theory of a data handling system with multiple sources", *Bell System Technical Journal*, Vol. 61, No. 8, pp. 1871-1894, 1982.

[2] A. Baiocchi, N. Blefari-Melazzi, "An error-controlled approximative analysis of a stochastic fluid flow model applied to an ATM multiplexer with heterogeneous on-off sources", *IEEE/ACM Transactions on Networking*, Vol. 1, No. 6, 1993.

[3] A. Elwalid, D. Mitra, "Effective bandwidth og general Markovian traffic sources and admission control of high speed networks", *IEEE/ACM Transactions on Networking*, Vol. 1, No. 3, pp. 329-357, 1992.

[4] A. Elwalid, D. Mitra, "Analysis, approximation and admission control of a multi-service multiplexing system with priorities", In *Proc. of Infocom'95*, Vol. 2, pp. 463-472, 1995.

[5] S. Jacobsen, L. Dittman, "A fluid flow queueing model for heterogeneous on/off traffic", In *Proc. RACE BLNT Workshop*, Munchen, 1990.

[6] K. Kvols, S. Blaabjerg, "Bounds and approximations for the periodic on/off queue with application to ATM traffic control", In *Proc. IEEE Infocom*, pp. 487-494, 1992.

[7] D. Mitra,"Stochastic theory of a fluid model of producers and consumers coupled by a buffer", *Adv. Applied Probability*, Vol. 20, pp. 646-676, 1988.

[8] E. Nordström, "Performance evaluation of a heterogeneous fluid FIFO multiplexer", *working paper*, 2009.

[9] E. Pallarés-Segarra, J. Garcia-Haro, “Fluid flow approach to evaluate the information loss probability in a finite buffering swicthing node under heterogeneous ON/OFF input traffic sources”, *Performance Evaluation*, Vol. 51, No. 2-4, pp. 153-169, Feb. 2003.

[10] F. Presti, Z. Zhang, D. Towsley, “Bounds, approximations and applications for a two-queue GPS system”, In *Proc. Infocom’96*, Vol. 3, pp. 1310-1317, 1996.

[11] R.G. Tucker, “Accurate method for analysis of a packet-speech multiplexer with limited delay”, *IEEE Transactions on Communications*, Vol. COM-36, pp. 479-483, 1988.