Effective Bandwidths for Multiclass Markov Fluids and Other ATM Sources*

G. Kesidis[1,2] J. Walrand[1] C.-S. Chang[3]

1. EECS Dept., University of California, Berkeley, CA94720.
2. Elec. & Comp. Eng. Dept, University of Waterloo, Waterloo, ON, Canada, N2L 3G1.
3. IBM Research Div., T.J. Watson Research Center, Yorktown Heights, NY 10598.

*Supported by: NSERC of Canada, Pacific Bell, Micro Grant of the State of California, ARO Grant Number DAAL03-89-K-0128

Abstract

We show the existence of effective bandwidths for multiclass Markov fluids and other types of sources that are used to model ATM traffic. More precisely, we show that when such sources share a buffer with deterministic service rate, a constraint on the tail of the buffer occupancy distribution is a linear constraint on the number of sources. That is, for a small loss probability one can assume that each source transmits at a fixed rate called its effective bandwidth. When traffic parameters are known, effective bandwidths can be calculated and may be used to obtain a circuit-switched style call acceptance and routing algorithm for ATM networks. The important feature of the effective bandwidth of a source is that it is a characteristic of that source and the acceptable loss probability only. Thus, the effective bandwidth of a source does not depend on the number of sources sharing the buffer nor on the model parameters of other types of sources sharing the buffer.

1 Introduction

Effective bandwidths have been discovered for certain traffic models and certain performance criteria (see [16], [12],[11], [4],[17],[3]). For example, consider a buffer of infinite size with service rate c cells/s. Assume that the buffers sources and buffer occupancy are in steady state. Let X be the number of cells in the buffer found by a typical arriving cell. Suppose that

$$\mathrm{P}\{X > B\} \;\leq\; \mathrm{e}^{-B\delta} \tag{1}$$

must be satisfied (the performance criterion). Suppose further that there are N_j independent on-off Markov fluids [1] of type j ($j = 1, 2, ..., K$) sharing the buffer. There exist functions α_j that depend only on the parameters of a type j source and δ, such that the constraint (1) holds for $B\delta >> 1$ if and only if

$$\sum_{j=1}^{K} N_j \alpha_j \;\leq\; c.$$

We call α_j the *effective bandwidth* of an on-off Markov fluid of type j (see [11] and [12] for proofs of this result and numerical examples that explore the accuracy of the effective bandwidth approach).

In general, effective bandwidths depend on both the traffic/buffer models and the performance criterion. Kelly [16] finds effective bandwidths for GI/G/1 queues under (1) and for M/G/1 queues with the performance criterion taken to be the buffer utilization (fraction of time $X \neq 0$) or mean workload ($EX < B$). Courcoubetis and Walrand [4] find effective bandwidths for stationary Gaussian sources under (1). Recently, Elwalid and Mitra [7] have also obtained effective bandwidth results for the case of continuous-time Markovian sources under (1) (c.f., Sections 3.3 and 3.4 and the Conclusions). The open question answered in this note is the existence of effective bandwidths for more general source models under (1).

We start by heuristically deriving an expression for $\mathrm{P}\{X > B\}$ for general source models. Consider an infinite buffer with service rate c shared by N_i sources of type i, $i = 1, ..., K$. All the sources are assumed independent. For all M_i greater than the average rate of cells produced by a source of type i, assume that the probability that a source of type i produces $M_i T$ cells over a period of time of length T is approximately $\exp(-TH_i(M_i))$ where H_i is convex and non-negative (this is assumption is motivated by the theory of large deviations and is discussed below). This approximation is sharpest for $T >> 1$.

By independence, the probability that, for $j = 1, ..., N_i$, the j^{th} source of type i produces $\mu_j T$ cells over time T is about

$$\exp\left(-T\sum_{j=1}^{N_i} H_i(\mu_j)\right).$$

Consequently, the probability that all sources of type i produce a total of $N_i M_i T$ cells over large time T is about

$$\sum_{\mu:\sum\mu_j = N_i M_i} \exp\left(-T\sum_{j=1}^{N_i} H_i(\mu_j)\right)$$

where $\mu = (\mu_1, ..., \mu_{N_i})$. Indeed, each choice of μ such that $\sum \mu_j = N_i M_i$ is one particular way for $N_i M_i T$ cells to get produced. This sum of exponentials can be approximated by the largest term (originally an argument of Laplace):

$$\begin{aligned} \sum_{\mu:\sum\mu_j = N_i M_i} \exp\left(-T\sum_{j=1}^{N_i} H_i(\mu_j)\right) &\approx \exp\left(-\inf_{\mu:\sum\mu_j = N_i M_i} T\sum_{j=1}^{N_i} H_i(\mu_j)\right) \\ &= \exp\left(-TN_i H_i(M_i)\right) \end{aligned}$$

where the last equality is due to the convexity of H_i. Therefore, by independence, the probability that, for $i = 1, ..., K$, the sources of type i produce $N_i M_i T$ over time T is about

$$\exp\left(-T\sum_{i=1}^{K} N_i H_i(M_i)\right).$$

Thus, the probability that, starting from an empty buffer, the sources of type i produce cells at rate $N_i M_i$ until the buffer occupancy exceeds B is

$$\exp\left(-B\frac{\sum N_i H_i(M_i)}{\sum N_i M_i \ - c}\right).$$

Indeed $T = B/(\sum N_i M_i - c)$ is the time the buffer occupancy takes to reach B when the aggregate cell arrival rate is $\sum N_i M_i$. By the argument of Laplace, the probability that the buffer occupancy, starting from empty, reaches B before it returns to empty is about

$$\exp\left(-B \inf_{\sum N_i M_i \ > c} \frac{\sum N_i H_i(M_i)}{\sum N_i M_i \ - c}\right) \ \approx \ \mathrm{P}\{X > B\}. \qquad (2)$$

Given the effective bandwidths of a buffer's sources, one can determine its *spare capacity* to accept more calls at any time. For instance, say we want to determine if a call of type j can be accommodated (i.e., constraint (1) is preserved) in a buffer that is currently being used by N_i calls of type i, $i = 1, ..., K$. If $\alpha_j(\delta) < c - \sum_{i=1}^K N_i \alpha_i(\delta)$ then the call can be accommodated, else it cannot. See, for example, [9], [12],[13], and [18] for further discussion on how effective bandwidths can be used for network resource management.

This note is organized as follows. In Section 2, we show the existence of effective bandwidths in the multiclass case when the sources satisfy certain conditions. In Section 3, we give expressions for the effective bandwidths of Markov-modulated fluids (or just "Markov fluids"), Markov-modulated Poisson processes, and discrete-time Markov sources. Finally, conclusions are drawn in Section 4.

2 General Effective Bandwidths

We now show the existence of effective bandwidths. First, some assumptions on the sources are made, then effective bandwidths are defined by considering the single source case, and finally the multiclass case is considered.

Consider an infinite buffer with deterministic service rate c cells/s, shared by N_i independent sources of type i, $i = 1, ..., K$. Denote by $[\cdot, \cdot]$ the scalar product. Let $\Gamma_i \in (0, \infty]$ (respectively $\gamma_i \in [0, \infty)$) denote the maximum (respectively minimum) possible cell arrival rate of a type i source. Let $\overline{\gamma}_i \in (0, \infty)$ be the average arrival rate of a type i source. We assume that

$$N \ \in \ \mathbf{C} := \{N \in \mathbf{Z}_+^K \ : \ [N, \Gamma] \ > \ c \ \text{ and } \ [N, \overline{\gamma}] \ < \ c\}$$

where $\mathbf{Z}_+ = \{0, 1, 2, ...\}$, $\overline{\gamma} = (\overline{\gamma}_1, ..., \overline{\gamma}_K)$ and $\Gamma := (\Gamma_1, ..., \Gamma_K)$. Let $M = (M_1, ..., M_K)$.

Motivated by equation (2), we take the measure of congestion in the buffer to be

$$\exp(-BI(N,c)+o(B)) \quad \text{where} \quad I(N,c) \ := \ \inf_{M\in\mathbf{A}(N,c)} \frac{\sum_{i=1}^{K} N_i H_i(M_i)}{[N,M]-c} \tag{3}$$

and $\mathbf{A}(N,c) := \{M \in \mathbf{R}_+^K \ : \ \gamma_i < M_i < \Gamma_i \ \forall i \ \text{ and } \ [N,M] > c\}$ (c.f., equation (5) for the definition of the H_i).. Thus, when $B\delta >> 1$, the constraint (4) is

$$I(N,c) \ \geq \ \delta. \tag{4}$$

Assume that the sources are stationary and ergodic. Consider a single source of type i. Let the number of arrivals of this type i source in the time interval $[0,t]$ be $A_i(t)$. Assume that A_i satisfies the conditions of the Gärtner-Ellis theorem [10],[6],[2]. That is, assume that the asymptotic log moment generating function of A_i,

$$h_i(\delta) \ := \ \lim_{t\to\infty} \frac{1}{t} \log \mathrm{E} \exp\left(A_i(t)\delta\right)$$

exists and is finite for all real δ, and that h_i is differentiable. We can directly verify that h_i is convex, positive and increasing for $\delta > 0$.

By the Gärtner-Ellis theorem, H_i is the Legendre transform of h_i:

$$H_i(M_i) \ := \ \sup_{\delta\in\mathbf{R}} \{\delta M_i - h_i(\delta)\}. \tag{5}$$

We can directly verify that H_i is non-negative, convex and differentiable, $H_i(\overline{\gamma}_i) = 0$, and $H_i(M) = \infty$ for all $M > \Gamma_i$ or $M < \gamma_i$. We also assume that H_i is *strictly* convex on the interval (γ_i, Γ_i).

Consider the case of a single source of type i. For $\delta > 0$, define $\alpha_i(\delta)$ to be the value of a such that

$$I_i(a) \ := \ \inf_{M_i\in\mathbf{A}_i(a)} \frac{H_i(M_i)}{M_i - a} = \delta$$

where $\mathbf{A}_i(a) := \{M_i \ : \ a < M_i < \Gamma_i\}$. Thus, $\alpha_i(\delta) = I_i^{-1}(\delta)$ can be interpreted as the rate at which to serve a single source of type i so that the constraint (4) is satisfied. We call α_i the *effective bandwidth* of the type i traffic. The following theorem gives us a more manageable form for α_i.

Lemma 1: Under the above conditions, for all $\delta > 0$,

$$\alpha_i(\delta) \;=\; \frac{h_i(\delta)}{\delta}.$$

Proof: Since H_i and h_i are convex conjugates, $h_i(\delta) = \sup_M\{M\delta - H_i(M)\}$. It then follows from the differentiability of H_i and h_i and the strict convexity of H_i that

$$h_i(\delta) \;=\; \delta H_i'^{-1}(\delta) - H_i(H_i'^{-1}(\delta)). \tag{6}$$

Define the function $g_i(M) := M - H_i(M)/H_i'(M)$. From the strict convexity of H_i it follows that g_i is strictly increasing on (γ_i, Γ_i). Thus we can define g_i^{-1} as the inverse of g_i; i.e., for $a \in (\overline{\gamma}_i, \Gamma_i)$, $g_i^{-1}(a)$ is the solution of the equation $a = M - H_i(M)/H_i'(M)$. Since $H' > 0$ on $(\overline{\gamma}_i, \Gamma_i)$, $g_i^{-1}(a) > a$ so that $g_i^{-1}(a) \in \mathbf{A}_i(a)$. Thus, $I_i(a) = H_i'(g_i^{-1}(a))$ and, in conjunction with equation (6), we have that $I_i^{-1}(\delta) = h_i(\delta)/\delta$ as desired. ♠

With this lemma, the following "effective bandwidth" theorem for multi-class sources is immediate by independence.

Theorem 1: Assume that the arrival processes A_i all satisfy the conditions of the Gärtner-Ellis theorem and that the H_i are all *strictly* convex. For any $\delta > 0$ and $N \in \mathbf{C}$,

$$I(N, c) \geq \delta \quad \Leftrightarrow \quad \sum N_i \alpha_i(\delta) \leq c.$$

Proof: Let h be the log moment generating function for the *aggregate* arrival process. Clearly

$$h(\delta) \;=\; \sum N_i h_i(\delta).$$

Let the inverse of $I(N, \cdot)$ be I_N^{-1}. Thus, by the argument in the lemma above,

$$I_N^{-1}(\delta) \;=\; \frac{h(\delta)}{\delta} \;=\; \frac{\sum N_i h_i(\delta)}{\delta} \;=\; \sum N_i \alpha_i(\delta)$$

as desired. ♠

This theorem shows that under weak conditions on the arrival processes, effective bandwidths exist for the measure of congestion (3). The large deviations approach used is a unified framework to handle buffer sources modeled in different ways as we shall see in the next section.

3 Models of ATM Buffer Sources

We now consider several models of buffer sources used to characterize bursty ATM traffic. In each case, an expression for the effective bandwidth is found.

3.1 Constant Rate and Memoryless Sources

For sources with a constant arrival rate of R cells/s, $A(t) = Rt$ for $t > 0$. Thus, $h(\delta) = R\delta$, $H(R) = 0$ and $H(M) = \infty$ for all $M \neq R$. Therefore, the hypothesis of Theorem 1 is satisfied and the effective bandwidth of this source is $\alpha(\delta) = R$. Note that, in the notation of Section 2, $\gamma = \Gamma = R$ for a constant rate source.

For memoryless (Poisson) sources with intensity R cells/s, $h(t) = R(e^{\delta} - 1)$. Thus, $H(M) = M\log(M/R) - M + R$. So, the hypothesis of Theorem 1 is satisfied and the effective bandwidth of this source is $\alpha(\delta) = R(e^{\delta} - 1)/\delta$. Note that $\gamma = 0$ and $\Gamma = \infty$ for a Poisson source.

3.2 Discrete-Time Markov Sources

We call a buffer source a "discrete-time" Markov source if there is a discrete-time Markov chain Z_n and a real constant R such that the number of arrivals to the buffer in interval of (continuous) time $(nR^{-1}, (n+1)R^{-1})$ is a function of Z_n. We take the state space of Z to be $1, 2, ..., m$ and we let Q be its irreducible and aperiodic transition probability matrix. Let Λ_i be the number of cells that arrive in the interval $(nR^{-1}, (n+1)R^{-1})$ when $Z_n = i$. We assume $0 \leq \Lambda_i \leq \Lambda_{i+1} < \infty$ for all $i = 1, ..., m-1$. Therefore, in the notation of Section 2, $\gamma = R\Lambda_1$, $\Gamma = R\Lambda_m$, and $\overline{\gamma} := R\sum_i \pi_i \Lambda_i$ where π is the invariant of Q: $\pi Q = \pi$.

By an argument using the backward equation and Perron-Frobenius theory [3],

$$h(\delta) \quad = \quad R\log\left[\rho(\mathrm{e}^{\delta\Lambda}Q)\right] \tag{7}$$

where $\Lambda = \mathrm{diag}(\Lambda_1, ..., \Lambda_m)$, and $\rho(F)$ is the spectral radius of the matrix F.

h is differentiable (and analytic) as a consequence of perturbation theory of matrices (see [2], p. 190-191) and, therefore, satisfies the conditions of the Gärtner-Ellis theorem. In Section 2, we established that h is convex. A simple consequence of Lemma 3.4 in [15] is that either h is affine or strictly convex.

$h(0) = 0$ implies that the affine case is the constant rate source of Section 3.1. If h is strictly convex, by direct calculation starting from equation (5), we get that $H' = h'^{-1}$. Thus H' is strictly increasing which implies that H is strictly convex as well. So, the hypothesis of Theorem 1 is satisfied, and the effective bandwidth of this source is $\alpha(\delta) = h(\delta)/\delta$. This source is a special case of an example 2.3 in [3] wherein the rates Λ_i are random.

3.2.1 Two-State Discrete-Time Markov Source Example

If the Markov chain considered is of the two state ($m = 2$) type, then by direct calculation,

$$h(\delta) \quad = \quad R \log \left[\frac{1}{2} \left(a(\delta) + \sqrt{a^2(\delta) + 4b(\delta)} \right) \right]$$

where

$$a(\delta) \;=\; Q_{1,1} e^{\delta \Lambda_1} + Q_{2,2} e^{\delta \Lambda_2} \quad \text{and} \quad b(\delta) \;=\; e^{\delta(\Lambda_1 + \Lambda_2)}(1 - Q_{1,1} - Q_{2,2}).$$

3.3 Markov Fluids

A source is called a Markov fluid if its time-derivative is a function of a continuous-time Markov chain on a finite state space. As for the discrete-time Markov sources above, we let $1, ..., m$ be the state space and Q be the irreducible transition *rate* matrix of the Markov fluid's time-derivative. Let Λ_i be the arrival rate of cells when the time-derivative of the Markov fluid is in state i. We make the same assumption on the parameters Λ_i that we made in the discrete-time Markov source case.

By an argument similar to that for discrete-time Markov sources (see the Appendix below),

$$h(\delta) \quad = \quad \mu(Q + \delta\Lambda)$$

where Λ is defined above and $\mu(F)$ is the largest real eigenvalue of the matrix F. The same argument used for discrete-time Markov sources verifies that the hypothesis of Theorem 1 is satisfied.

3.3.1 Two-State Markov Fluids Example

If the Markov fluid considered is of the two state ($m = 2$) type, then by direct calculation,

$$h(\delta) \;=\; \frac{1}{2}\left(-a(\delta) + \sqrt{a^2(\delta) - 4b(\delta)}\right)$$

where

$$a(\delta) \;=\; Q_{1,2} + Q_{2,1} - \delta(\Lambda_2 - \Lambda_1) \quad \text{and} \quad b(\delta) \;=\; \delta^2\Lambda_2\Lambda_1 - \delta(Q_{1,2}\Lambda_2 + Q_{2,1}\Lambda_1).$$

This is the effective bandwidth result in [11],[12].

3.4 Markov-Modulated Poisson Process

A source to a buffer is called a Markov-modulated Poisson process (MMPP) if the cell arrivals are Poisson with intensity λ, where λ is a function of a continuous-time Markov chain. We assume that the space $\Lambda_1, ..., \Lambda_m$ of intensities satisfies the conditions of the previous examples and that the transition rate matrix Q is irreducible.

By an argument similar to that for discrete-time Markov sources (again, see the Appendix below),

$$h(\delta) \;=\; \mu(Q + (e^\delta - 1)\Lambda)$$

and the hypothesis of Theorem 1 is satisfied.

4 Conclusions

Effective bandwidth results for the continuous-time Markovian sources of Sections 3.3 and 3.4 were also obtained in [7] using spectral decomposition methods [20],[8]. They found the same effective bandwidth formulas and establish equation (2) for buffers with multiclass Markov fluid sources and buffers with multiclass MMPP sources. The effective bandwidth results in Section 2 (using the large deviations approach) are more general than those of [7] and our measure of congestion (equation (3)) allows us to handle a buffer with sources modeled in different ways (e.g., a buffer with two sources:

one modeled as a Markov fluid and the other as a MMPP). Recently, in [5], equation (2) was established for the stationary Lindley buffer process (discrete time) and they find an effective bandwidth result for a buffer using a simple "randomized priority" processor sharing rule [19].

In summary, we have shown the existence of effective bandwidths for a large class of sources commonly used to model ATM traffic. Given the effective bandwidths of a buffer's sources (i.e., the functions α_i for the buffer of Section 2), one can determine its spare capacity to accept more calls, $c - \sum_{i=1}^{K} N_i \alpha_i(\delta)$, which can be an integral part of network resource management [12],[13],[18].

References

[1] D. Anick, D. Mitra, and M. M. Sondhi. Stochastic theory of a data-handling system with multiple sources. *Bell Sys. Tech. J.*, Vol. 61 No.8:1871–1894, 1982.

[2] J.A. Bucklew. *Large Deviation Techniques in Decision, Simulation and Estimation.* John Wiley and Sons, Inc., New York, NY, 1990.

[3] C.-S. Chang. Stability, queue length and delay, part II: Stochastic queueing networks. *Proc. IEEE CDC, Tucson, AZ*, .:1005–1010, 1992.

[4] C. Courcoubetis and J. Walrand. Note on effective bandwidth of ATM traffic. *preprint*, .:., .

[5] G. de Veciana and J. Walrand. Effective bandwidths: Call admission, traffic policing & filtering for ATM networks. *preprint*, .:., 1993.

[6] R. Ellis. Large deviations for a general class of random vectors. *Ann. Probab.*, Vol. 12:1–12, 1984.

[7] A.I. Elwalid and D. Mitra. Effective bandwidth of general Markovian traffic sources and admission control of high speed networks. *Proc. IEEE INFOCOM*, Vol. 1:3a.2.1–3a.2.10, 1993.

[8] A.I. Elwalid, D. Mitra, and T.E. Stern. Statistical multiplexing of Markov modulated sources: theory and computational algorithms. *Proc. ITC'13*, .:495–500, Copenhagen, Denmark, June 1991.

[9] G. Gallassi, G. Rigolio, and L. Fratta. ATM: bandwidth assignment and bandwidth enforcement policies. *Proc. Globecom*, Vol. 3:1788–1793, 1989.

[10] J. Gartner. On large deviations from invariant measure. *Theory Probab. Appl.*, Vol. 22:24–39, 1977.

[11] R.J. Gibbens and P.J. Hunt. Effective bandwidths for multi-type UAS channel. *Queueing Systems*, Vol. 9:17–28, 1991.

[12] R. Guérin, H. Ahmadi, and M. Naghshineh. Equivalent capacity and its application to bandwidth allocation in high-speed networks. *IEEE JSAC*, Vol. 9, No. 7:968–981, 1991.

[13] R. Guérin and L. Gun. A unified approach to bandwidth allocation and access control in fast packet-switched networks. *Proc. IEEE INFOCOM*, Vol. 1:1–12, 1992.

[14] R.A. Horn and C.R. Johnson. *Topics in Matrix Analysis.* Cambridge, ., 1986.

[15] I. Iscoe, P. Ney, and E. Nummelin. Large deviations of uniformly recurrent Markov additive processes. *Adv. Appl. Prob.*, Vol. 6:373–412, 1985.

[16] F.P. Kelly. Effective bandwidths of multi-class queues. *Queueing Systems*, Vol. 9:5–16, 1991.

[17] G. Kesidis. Cell loss estimation in high-speed digital networks. *Ph.D. Dissertation, EECS Dept, U.C. Berkeley*, .:., 1992.

[18] G. Kesidis and J. Walrand. Traffic policing and enforcement of effective bandwidth constraints in ATM networks. *submitted to IEEE-ACM Trans. Networking*, .:., 1992.

[19] J.M. Pitts and J.A. Schormans. Analysis of ATM switch model with time priorities. *Electronic Letters*, Vol. 26, No. 15:1192,1193, July 1990.

[20] T.E. Stern and A.I. Elwalid. Analysis of seperable Markov-modulated models for information handling systems. *Adv. Appl. Prob.*, .:., March 1991.

Appendix: Backward Equation Approach to Evaluate the Effective Bandwidth for Markovian Sources

For the Markov fluid source of Section 3.3, let $A(s,t)$ be the number of arrivals in the interval (s,t), x be the irreducible modulating Markov chain with rate matrix Q and invariant distribution π, and $\psi_j(\delta,t) = E[\exp(\delta A(0,t))|x(0) = j]$.

The claim is that

$$\begin{aligned} h(\delta) &:= \lim_{t\to\infty} t^{-1} \log E \exp(\delta A(0,t)) \\ &= \mu(Q + \delta\Lambda). \end{aligned}$$

To show this, we begin with a standard backward equation argument: for positive $\epsilon << 1$,

$$\begin{aligned} \psi_j(\delta,t) &= E\left(E[e^{\delta A(0,t)}|x(\epsilon)]\ |x(0) = j\right) &\quad (8)\\ &= \sum_i \psi_i(\delta, t-\epsilon) e^{\epsilon Q}(j,i) e^{\epsilon\delta\Lambda_j} + o(\epsilon). &\quad (9) \end{aligned}$$

Since $\exp(\epsilon Q)(j,i) = (I + \epsilon Q)(j,i) + o(\epsilon)$ and $\exp(\epsilon\delta\Lambda_j) = 1 + \epsilon\delta\Lambda_j + o(\epsilon)$, we get, after a little rearrangement,

$$\frac{\psi_j(\delta,t) - \psi_j(\delta,t-\epsilon)}{\epsilon} = \psi_j(\delta,t-\epsilon)(Q(j,j) + \delta\Lambda_j) + \sum_{i\neq j} \psi_i(\delta,t-\epsilon) Q(j,i) + O(\epsilon).$$

Letting $\epsilon \to 0$, we get

$$\frac{\partial}{\partial t}\psi_j(\delta,t) = \psi_j(\delta,t)(Q(j,j) + \delta\Lambda_j) + \sum_{i\neq j} \psi_i(\delta,t) Q(j,i).$$

In matrix form this equation is

$$\frac{\partial}{\partial t}\Psi(\delta,t) = (Q + \delta\Lambda)\Psi(\delta,t)$$

where $\Psi^T(\delta,t) = (\psi_1(\delta,t), ..., \psi_m(\delta,t))$. Thus,

$$\Psi(\delta,t) = \exp((Q + \delta\Lambda)t)\mathbf{1}$$

where $\mathbf{1} = \Psi(\delta, 0)$ is a column of 1's.

Therefore,

$$h(\delta) \quad = \quad \lim_{t\to\infty} \frac{1}{t} \log \left(\pi^T \exp((Q + \delta\Lambda)t)\mathbf{1}\right).$$

First note that $\exp(Q+\delta\Lambda)$ is a nonnegative matrix (see [14], Exercise 6.5.4e and Theorems 6.2.9(g) and 6.2.38). Choose a large enough such that $aI+Q+\delta\Lambda \geq 0$. This is possible since $Q_{i,j} \geq 0$ for all $i \neq j$. Thus, $\exp(Q + \delta\Lambda) = \exp(aI + Q + \delta\Lambda)\exp(-aI) \geq e^{-a}\exp(aI + Q + \delta\Lambda) \geq 0$. Because of the irreducibility assumption, we can use the same Perron-Frobenius argument in [3] on the matrix $\exp(Q + \delta\Lambda)$ to obtain $h(\delta) = \log(\rho(\exp(Q + \delta\Lambda)))$. The result then follows from $\rho(\exp(F)) = \exp(\mu(F))$, where $\mu(F)$ is the largest eigenvalue of F.

For the case of the MMPP source of Section 3.4, we use the fact that if ξ is a Poisson random variable with mean $\epsilon\Lambda_j$, then $E\exp(\delta\xi) = \exp(\epsilon\Lambda_j(e^{\delta}-1))$. So, the above argument will give us the formula for h in Section 3.4 by simply substituting the expression
"$\exp(\epsilon\Lambda_j(e^{\delta} - 1))$" for "$\exp(\epsilon\delta\Lambda_j)$" in equation (9) above.