Revista de Matemática: Teoría y Aplicaciones **32**(1): 55-71
CIMPA - UCR ISSN: 1409-2433(Print), 2215-3373(Online)
DOI: https://doi.org/10.15517/rmta.v32i1.59733

THE SIGNIFICANCE OF THE PERRON-FROBENIUS EIGENVALUE FOR THE CHARACTERIZATION OF A MEASURE OF CHANNEL OCCUPANCY IN DATA NETWORKS

IMPORTANCIA DEL AUTOVALOR DE PERRON-FROBENIUS PARA LA CARACTERIZACIÓN DE UNA MEDIDA DE OCUPACIÓN DEL CANAL EN REDES DE DATOS

Carina Fernández[1] José Bavio[2] Beatriz Marrón[3]

Received: 12/Jun/2024; Accepted: 28/Oct/2024

---

[1] Universidad Nacional del Sur, Departamento de Matemática, Bahía Blanca, Argentina. E-Mail: carina.fernandez@uns.edu.ar
[2] Same address as/Misma dirección que [1]. E-Mail: jmbavio@yahoo.com.ar
[3] Same address as/Misma dirección que [1]. E-Mail: beatriz.marron@uns.edu.ar

Abstract

Telecommunications is currently one of the most important technologies, utilizing various infrastructure elements to enable communication over the Internet. When an internet connection is established, traffic packets are routed from one device to another, taking different paths through routers that process each packet. These networks present multiple challenges in processing, distributing and connecting traffic. To address these challenges it is necessary, on the one hand, to have models that describe the behavior of traffic and its evolution over time, such as the Generalized Markov Fluid Model. On the other hand, it is essential to reserve part of the available resource at each node for ongoing connections. To carry out this reservation process, the Effective Bandwidth is employed. In this paper, we describe the distribution of the buffer in equilibrium for traffic sources modeled by a Generalized Markov Fluid Model, using a system of differential equations. As the main result of this paper, we prove that it is possible for this model to characterize the effective bandwidth when the most probable duration of the buffer busy period prior to overflow becomes larger and larger. Finally, we verify this result numerically from simulated traffic traces.

Keywords: effective bandwidth; Markov fluid; eigenvalues; Perron-Frobenius method; differential equations.

Resumen

En la actualidad, una de las tecnologías más importantes son las telecomunicaciones, las cual despliega diversos elementos de infraestructura para permitir la comunicación a través de internet. Cuando se establece una conexión a internet, los paquetes de tráfico se dirigen de un dispositivo a otro, recorriendo diversas rutas a través de routers que procesan cada paquete. Estas redes presentan múltiples desafíos a la hora de procesar, distribuir y conectar el tráfico. Para abordar estos retos es necesario, por un lado, disponer de modelos que describan el comportamiento del tráfico y su evolución en el tiempo, como el Modelo de Flujo Markoviano Generalizado. Por otro lado, reservar en cada nodo parte del recurso disponible para las conexiones en proceso y para llevar a cabo este proceso de reserva se utiliza el Ancho de Banda Efectivo. En este trabajo describimos, a partir de un sistema de ecuaciones diferenciales, la distribución del buffer en equilibrio en fuentes de tráfico modeladas por un Modelo de Flujo Markoviano Generalizado. Como resultado principal de este trabajo, demostramos para este modelo, que es posible caracterizar el ancho de banda efectivo cuando la duración más probable del periodo ocupado del buffer antes del desbordamiento se hace cada vez mayor. Finalmente, verificamos numéricamente este resultado a partir de trazas de tráfico simuladas.

Palabras clave: ancho de banda efectivo; flujo markoviano; autovalores; método de Perron-Frobenius; ecuaciones diferenciales.

Mathematics Subject Classification: Primary: 60J25. Secondary: 15A18.

1. Introduction

Statistical modeling of data networks plays a key role in understanding and optimizing network performance. By analyzing the statistical properties of traffic or the behavior of specific network components, such as buffer size or link capacity, we can gain insight into the system and make more informed decisions about resource allocation and congestion control.

To improve network utilization by efficiently increasing its data transmission capacity statistical multiplexing is used, which allows sources to share the available bandwidth dynamically, taking advantage of idle periods of some sources to allow data transmission from others, thus making better use of available resources. Statistical multiplexing in data networks allows us to better understand network behavior, optimize resource allocation, and design effective congestion control mechanisms, and by accurately capturing the statistical properties of network traffic, network performance is improved.

The statistical multiplexing system consists of a buffer of capacity $\boldsymbol{B}$ that receives input from multiple sources that are statistically independent. This buffer is served by a channel of constant capacity, denoted $\boldsymbol{C}$. In these systems, it is possible to allocate a bandwidth to each source based on its characteristics, including burst and service requirements. It is important to note that the bandwidth of a source is independent of the traffic sent to the multiplexer by other sources, so the complexity of its calculation depends only on the source and not on the dimensions of the system.

A problem in multiplexing is the non-zero probability of multiple sources simultaneously dispatching at the maximum rate, leading to potential overflow. To avoid data loss and guarantee the preservation of Quality of Service (QoS) for both current and future sources, it is essential to implement admission control mechanisms when accepting new connections.

Statistical tools that allow characterizing the network requirement and that have emerged as powerful methods to determine the probability of packet loss, arise from the concept of Effective Bandwidth (EB) introduced by Kelly in 1996, [10]. The EB is a useful and realistic measure of channel occupancy, reserving for each source a capacity higher than the average transmission rate, which would be a very optimistic measure, but lower than the maximum transmission rate, which would be a pessimistic measure.

Many authors have calculated the effective bandwidth for Markovian fluids from different approaches, developing formulas involving the estimable parameters of the model. Thus, Kelly [10], Courcoubetis and Siris [4], Cao et al. [3] consider the general problem of the existence of an effective bandwidth of stationary and ergodic sources, while Kesidis, Walrand and Chang [11] adopt a large deviations approach to determine the effective bandwidth. From the latter approach, an explicit formula for the effective bandwidth has been found for the Generalized Markov Fluid Model (GMFM) [2].

In this paper, the behavior of the effective bandwidth of a GMFM modulated source is studied for limiting buffer occupancy time values before overflow. Mitra [1] describes by a set of differential equations the equilibrium buffer distribution for a source modulated by a Markov fluid. In particular, we prove that this result holds for the GMFM. On the other hand, we show that the EB for this model is equal to the maximum real eigenvalue of a matrix, derived from the source parameters, the network resources and the service requirements, with dimension equal to the number of source states.

The paper is organized as follows: In Section 2, we give some notations that we use throughout the paper, define some notions of matrix theory, and state the main results of Perron-Frobenius theory that we use to prove our main result. Section 3 introduces the Generalized Markov Fluid Model. It is shown that the equilibrium buffer distribution for a GMFM-modulated source is described by a set of differential equations. Section 4 provides background material on effective bandwidth. In particular, we focus on the study of a GMFM modulated source. Section 5 shows, as a main result, that the EB of this type of source is characterized with the Perron-Frobenius eigenvalue. Section 6 provides a numerical check of the main result by using simulated traffic traces. The paper ends with some concluding remarks in Section 7.

2. Preliminares

2.1. Notations

Let A be a matrix of size $n \times n$. We denote by a_{ij} the element at position (i, j) of the matrix A, and by $a_{ij}^{(m)}$ the element at position (i, j) of the matrix A^m. We consider that $A > 0$ if every $a_{ij} \geq 0$ and at least one $a_{ij} > 0$, for $1 \leq i, j \leq n$. We consider that $A \gg 0$, if $a_{ij} > 0$, $\forall\ 1 \leq i, j \leq n$. A^T indicates the transposed A. The symbol $\langle \cdot, \cdot \rangle$ denotes the inner product of vectors.

2.2. Definitions

Definition 1. Let A be a matrix of size $n \times n$. A is said to be a *nonnegative matrix* whenever each $a_{ij} \geq 0$, and this is denoted by writing $A \geq 0$.

Definition 2. Let A be a matrix of size $n \times n$. A is said to be a *reducible matrix* when there exists a permutation matrix P such that $P^T A P = \begin{bmatrix} X & Y \\ 0 & Z \end{bmatrix}$, where X and Z are both square. Otherwise A is said to be an *irreducible matrix*.

Definition 3. Let A be a matrix of size $n \times n$. We denote $\sigma(A)$ as the set of all eigenvalues of A. We denote $\rho(A)$, the *Spectral Radius* of A, the maximum of the absolute values of the elements of $\sigma(A)$.

Definition 4. A nonzero vector $\mathbf{x}$ is a *left eigenvector* of a matrix A with an eigenvalue λ of A if $\mathbf{x}^T A = \lambda \mathbf{x}^T$. Likewise $\mathbf{x}$ is an *right eigenvector* with an eigenvalue λ of A if $A\mathbf{x} = \lambda\mathbf{x}$. Furthermore, we will say that $\mathbf{x}$ is *normalized* if it is verified that $\langle \mathbf{x}, \mathbf{x} \rangle = 1$. For further details see [8].

Definition 5. Let $\phi_d(\lambda)$ and $\phi_i(\lambda)$ be the right and left normalized eigenvectors associated with the eigenvalue λ, respectively. The *Spectral Projection Matrix* is defined as $P(\lambda) := \phi_d(\lambda)\phi_i^T(\lambda)$.

2.3. Perron-Frobenius theory

The following result can be found in the literature in different ways, depending on the required assumptions on the matrix. We give the particular case we are interested in for an irreducible and non-negative matrix. More details on this result can be found in [8].

Theorem 1. ***(Perron-Frobenius)*** *Let A be an irreducible non-negative matrix of size $n \times n$. Then,*

1. $\rho(A) \in \sigma(A)$.
2. $\rho(A)$ *has algebraic multiplicity* 1.
3. *There is a unique eigenvector* $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{x} > 0$ *such that* $A\boldsymbol{x} = \rho(A)\boldsymbol{x}$ *and* $\sum_{i=1}^n x_i = 1$.
4. *There is a unique eigenvector* $\boldsymbol{y} \in \mathbb{R}^n$, $\boldsymbol{y} > 0$ *such that* $\boldsymbol{y}^T A = \rho(A)\boldsymbol{y}^T$ *and* $\sum_{i=1}^n x_i y_i = 1$.

The following results, consequences of the Perron-Frobenius Theorem, can be found in more detail in [9] and [8], respectively.

Proposition 1. *Let $A \geq 0$ be an irreducible matrix of size $n \times n$ and P be the spectral projection matrix with respect to $\rho(A)$. Then,*

1. $P^2 = P$.
2. $AP = PA = \rho(A)P$.

Proposition 2. *If $A > 0$ and there exists $\boldsymbol{x} \gg 0$ such that $A\boldsymbol{x} \leq \mu\boldsymbol{x}$, then $\rho(A) \leq \mu$.*

Finally, the spectral radius is related to a matrix norm by the following result, introduced by Gelfand [6].

Theorem 2. ***(Gelfand formula)*** *Let $\|\cdot\|$ be a matrix norm of a matrix A of size $n \times n$. Then $\rho(A) = \lim_{k\to\infty} \left\|A^k\right\|^{1/k}$.*

3. The model

Markov fluid models have been developed for a long time and have proven to be especially valuable for accurately modeling various real data sources, such as voice and video. It is a work arrival process in which the work arrival rate depends on the state of the Markov chain. Its main property of "memorylessness" allows to deduce that the statistics of transitions between one state and another can be summarized through a matrix Q, called the Infinitesimal Generator of the chain.

In a real data, voice or video traffic, the large number of different values observed for the rates is transferred, in this model, in the number of states assumed by the modulating chain and consequently in the dimension of the matrices involved in the model, making the modeling problem almost unmanageable. This inconvenience has given rise to the so-called Generalized Markov Fluid Model (GMFM) [2]. It is a model modulated by a continuous, homogeneous and irreducible Markov chain in time, where the transfer rates are now random variables whose ranges and probability distributions are determined by the states of the modulating chain, which can be interpreted as a type of activity performed by a user, such as chat, video call, etc. A sudden change in the transfer rate can be identified as a change of state in the chain. Formally, we can define the GMFM as follows.

Definition 6. Let Z be a continuous-time, homogeneous, irreducible Markov chain with finite state space $\mathcal{K} = \{1, 2, \ldots, k\}$, invariant distribution π and infinitesimal generator Q. Furthermore, let $f_1, \ldots, f_k$, be k probability laws with disjoint and known supports, such that when the chain Z reaches at instant s the state i, the conditional random variable $Y_s|Z_s = i$ is distributed according to the known probability law f_i, with mean μ_i for $i = 1, \ldots, k$. We will call Generalized Markov Fluid Model modulated by the chain Z to the process

$$X_t = \int_0^t Y_s ds. \tag{3.1}$$

Such a process represents the accumulated work in the interval $[0, t]$ received from the source dispatching information at rate Y_s, that remains constant throughout the entire time interval in which the Markov chain stays in a given state. The process Y is observable and, since the supports I_i of the distributions f_i are disjoint and known, so the process Z is also observable.

Since this is a continuous-time work arrival process that is modulated by a Markov chain, its main property is that the statistics of the trajectory after time t depend only on the state occupied by the process at t, and not on the previous history. This property allows us to deduce that the statistics of the transitions between one state and another can be summarized through a matrix Q, the infinitesimal generator of the chain, whose entries q_{ij} represent the average number of transitions per unit of time between state i and state j, that is, the speed with which the chain leaves state i to move to state j at instant t. The elements of the main diagonal q_{ii} are related to the average time that the process remains in state i. In particular, the quantity $q_i = -q_{ii}$ can be interpreted as the number of transitions leaving state i per unit of time, or as the total speed at which state i is left to move to any other state at instant t.

The probability of the chain with k states being in state i is π_i, where the vector $\pi = (\pi_1, \ldots, \pi_k)$ satisfies the balance equation

$$\pi Q = 0.$$

The transfer rates are random variables, whose ranges and probability distributions are determined by the states of the modulating chain, which can be interpreted as a type of activity performed by a user, such as chat, video call, etc. In each state i, the data transfer rate assumes values that depend on the corresponding activity, following a certain probability distribution f_i. Using this information we define the diagonal matrix H of size $k \times k$ such that the non-zero elements are the first moments of each distribution, i.e., $h_{ii} = \mu_i$. We introduce this matrix to facilitate the handle of these values in future operations.

3.1. Equilibrium buffer distribution

For a given buffer content X, we ask whether it is possible to model, for the GMFM, the equilibrium buffer distribution. To do so, we start by studying the behavior of the source for each state. The stationary state distribution of the multiplexing system can be defined as $\pi(x) = \{\pi_s(x) \mid s \in \mathcal{K}\}$, with

$$\pi_s(x) = P(Z = s, X \leq x), \;\; s \in \mathcal{K}, 0 \leq x < \infty.$$

As an answer to our question arises the following theorem, in which we show that, for a source modulated by the GMFM, it is possible to describe the equilibrium buffer distribution by means of a set of differential equations.

Theorem 3. *The equilibrium buffer distribution for a source modulated by a GMFM, defined as in (3.1), is described by the following system of differential equations:*

$$\frac{d}{dx}\pi(x)H = \pi(x)Q.$$

Proof. Let $\pi_s(t,x) = P(Z_t = s, X \leq x)$ be the probability that at time t the modulate chain is in state s, and accumulated work does not exceed x.

Between two different time instants, the chain can remain in the same state s or change from state s' to state s and the average number of such changes is denoted by $g_{s's}$. Analyzing all possible scenarios occurring within a small increment of time Δt, for our model defined as in (3.1), we have that

$$\begin{aligned}\pi_s(t+\Delta t, x) &= \sum_{s'\in\mathcal{K}} g_{s's}\Delta t \int_{I_{s'}} \pi_{s'}(t, x - y\Delta t)\, f_{s'}(y)dy \\ &\quad + \left(1 - \sum_{s'\in\mathcal{K}} g_{s's}\Delta t\right)\int_{I_s} \pi_s(t, x - y\Delta t)\, f_s(y)dy.\end{aligned}$$

We thus have

$$\begin{aligned}\frac{\pi_s(t+\Delta t, x)}{\Delta t} &= \sum_{s'\neq s} g_{s's}\int_{I_{s'}} \pi_{s'}(t, x - y\Delta t)\, f_{s'}(y)dy \\ &\quad - \sum_{s'\neq s} g_{s's}\int_{I_s} \pi_s(t, x - y\Delta t)\, f_s(y)dy + \int_{I_s} \frac{\pi_s(t, x - y\Delta t)\, f_s(y)dy}{\Delta t}.\end{aligned}$$

Passing to the limit $\Delta t \to 0$, the last equation can be written as follows:

$$\frac{\partial \pi_s(t,x)}{\partial t} = \sum_{s'\neq s} g_{s's}\pi_{s'}(t,x) - \sum_{s'\neq s} g_{s's}\pi_s(t,x) - \mu_s\frac{\partial \pi_s(t,x)}{\partial x}.$$

We are interested only in time-independent, equilibrium probabilities. Therefore, we set $\frac{\partial \pi_s(t,x)}{\partial t} = 0$ and obtain

$$\mu_s\frac{\partial \pi_s(t,x)}{\partial x} = \sum_{s'\neq s} g_{s's}\pi_{s'}(t,x) - \sum_{s'\neq s} g_{s's}\pi_s(t,x). \tag{3.2}$$

Finally, recalling the definition of the matrix H, we can write (3.2) in matrix notation as follows:

$$\frac{d}{dx}\pi(x)H = \pi(x)Q.$$

□

Let the stationary buffer overflow distribution be given by $P(X > x) = 1 - \langle \pi(x), 1\rangle$, where $\pi(x)$ is the solution of the linear system in Theorem 3.

Knowing how much of the available resources each connection needs, has direct application to Connection Admission Control, minimizing the probability of buffer overflow. It will then be necessary to have some measure of the occupation of the transmission channel involving the parameters used to model the traffic and which are also involved in the buffer equilibrium distribution. In the following section, this measure of occupation is introduced.

4. Effective bandwidth

Multiplexing variable-speed sources on a link requires that each source be assigned a link capacity. This amount must be greater than the average transfer rate, because that would be a very optimistic estimate, but less than the maximum transfer rate to avoid unnecessary waste of resources. In order to achieve this balance, the concept of Effective Bandwidth (EB), introduced by F. Kelly in 1996 [10], provides a valuable and realistic measure of channel occupancy.

The EB is closely related to the theory of large deviations and depends on the statistical characteristics of the traffic, as well as the link capacity C, the buffer size B, and the interaction with other traffic sources sharing the buffer. Applications of this measurement to telecommunication networks are presented in [7, 13], among others.

It is useful to have an explicit formula for EB, which depends on the model parameters and can be estimated from the traffic traces. For the GMFM, the EB can be expressed in the following way:

Theorem 4. *Let $\{X_t\}_{t\geq 0}$ be a GMFM modulated by a homogeneous irreducible Markov chain Z with invariant distribution π and infinitesimal generator Q. And let the diagonal matrix H be defined as before, then the EB for a GMFM is*

$$\alpha(s,t) = \frac{1}{st}\log\left\{\pi \exp\left[(Q + sH)\,t\right] \mathbf{1}\right\}, \tag{4.1}$$

where $\mathbf{1}$ is a column vector with all of its entries equal to 1.

Its proof is developed in [2].

As we can see, (4.1) depends on the parameters s and t which are unknown, unrelated, and depend on the charasteristics of the multiplexed traffic and the link resources, capacity, and buffer. Specifically, the time parameter t, measured in sec., corresponds to the most probable duration of the buffer busy period prior to overflow. The space parameter s, measured in Mb^{-1}, corresponds to the degree of multiplexing and depends, among others, on the size of the peak rate of the multiplexed sources relative to the link capacity. In particular, for links with capacity much larger than the peak rate of the multiplexed sources, s tends to zero and $\alpha(s,t)$ approches the mean rate of the sources, while for links with capacity not much larger than the peak rate of the sources, s is large and $\alpha(s,t)$ approches the maximum value of $\frac{\bar{X}_t}{t}$, where $\bar{X}_t := \inf\{x : P(X_t > x) = 0\}$, i.e., the least upper bound on the value that X_t takes with positive probability.

5. Main result: Effective bandwidth characterization by Perron-Frobenius eigenvalue

We are interested in being able to find an expression of the EB when the most likely duration time of the buffer busy period prior to overflow becomes larger and larger.

In this case, the following Theorem shows that the EB of a GFGM modulated source is the maximum real eigenvalue of a matrix, derived from the source parameters, the network resources, and the service requirements, with dimension equal to the number of source states.

Theorem 5. *Let $\{X_t\}_{t\geq 0}$ be a GMFM modulated by a homogeneous irreducible Markov chain Z with invariant distribution π, infinitesimal generator Q and diagonal matrix H defined as before. Then*

$$\lim_{t\to\infty} \alpha(s,t) = \frac{1}{s}\rho(Q+sH). \tag{5.1}$$

To prove Theorem 5, we need the following lemma:

Lemma 1. *Let $\{X_t\}_{t\geq 0}$ be a GMFM modulated by a homogeneous irreducible Markov chain Z with invariant distribution π, infinitesimal generator Q and diagonal matrix H defined as before. Then*

$$\lim_{t\to\infty}\{\pi \exp[(Q+sH)t]\mathbf{1}\}^{1/t} = \rho(\exp(Q+sH)).$$

Proof. Let s be arbitrary. First, let us show that

$$\rho(\exp(Q+sH)) \geq \lim_{t\to\infty}\{\pi \exp[(Q+sH)t]\mathbf{1}\}^{1/t}.$$

Let us denote $A(s) := \exp(Q+sH)$. Since the matrix $A(s)$ is nonnegative and irreducible, because of the irreducibility of Q, we can then apply Theorem 1. Let $\phi_d(s)$ and $\phi_i(s)$ be the right and left normalized eigenvectors, respectively, associated with the Perron-Frobenius eigenvalue $g_p(s) := \rho(A(s))$ of the matrix $A(s)$. Then

$$\begin{aligned} A(s)\phi_d(s) &= g_p(s)\phi_d(s), \\ \phi_i^T(s)A(s) &= g_p(s)\phi_i^T(s). \end{aligned} \tag{5.2}$$

Let $P(s)$ be the spectral projection matrix with respect to $g_p(s)$, then,

$$P(s) = \phi_d(s)\phi_i^T(s). \tag{5.3}$$

Let $g_B(s)$ be an eigenvalue of the matrix $B(s) := A(s) - g_p(s)P(s)$. Let us prove that $|g_B(s)| \leq g_p(s)$. Suppose, for some $x(s) \neq 0$, that

$$B(s)x(s) = g_B(s)x(s). \tag{5.4}$$

By Proposition 1 applied to $P(s)$, (5.2), (5.3), and (5.4), we have that

$$\begin{aligned} g_B(s)P(s)x(s) &= P(s)B(s)x(s) \\ &= P(s)[A(s) - g_p(s)P(s)]x(s) \\ &= [\phi_d(s)\phi_i^T(s)A(s) - g_p(s)P(s)]x(s) \\ &= [\phi_d(s)g_p(s)\phi_i^T(s) - g_p(s)P(s)]x(s) \\ &= [g_p(s)P(s) - g_p(s)P(s)]x(s) \\ &= 0, \end{aligned} \tag{5.5}$$

so $g_B(s)x(s) \in \operatorname{Ker} P(s)$.

If $g_B(s) = 0$ then $|g_B(s)| \leq g_p(s)$, else, by (5.5) it follows that $x(s) \in \operatorname{Ker} P(s)$, then

$$\begin{aligned} g_B(s)x(s) &= B(s)x(s) \\ &= A(s)x(s) - g_p(s)P(s)x(s) \\ &= A(s)x(s), \end{aligned}$$

so that $g_B(s)$ is an eigenvalue of $A(s)$ and by Theorem 1,

$$|g_B(s)| \leq g_p(s). \tag{5.6}$$

Using Gelfand's formula for the L^1-norm, Proposition 2, (5.4), and (5.6), it follows that, for all m,

$$\rho(B(s)) = \lim_{m\to\infty}\left(\max_{1\leq j\leq k}\sum_{i=1}^{k}\left|b_{ij}^{(m)}\right|\right)^{1/m} \leq g_p(s).$$

This means that

$$\left|b_{ij}^{(m)}\right| \leq \sum_{i=1}^{k}\left|b_{ij}^{(m)}\right| \Rightarrow \max_{i,j}\left|b_{ij}^{(m)}\right| \leq \max_{j}\sum_{i=1}^{k}\left|b_{ij}^{(m)}\right|,\ \forall 1 \leq i, j \leq k,$$

then

$$\lim_{m\to\infty}\left(\max_{i,j}\left|b_{ij}^{(m)}\right|\right)^{1/m} \leq \rho(B(s)) \leq g_p(s).$$

Therefore, there exists $k_1 \in \mathbb{N}$ such that $\left(\max_{i,j}\left|b_{ij}^{(m)}\right|\right)^{1/m} \leq g_p(s),\ \forall\ m \geq k_1$, which means that

$$\max_{i,j}\left|b_{ij}^{(m)}\right| \leq g_p(s)^m,\ \forall\ m \geq k_1. \tag{5.7}$$

By induction on m, it can be proved that $B(s)^m = A(s)^m - g_p(s)^m P(s)$. Then

$$\pi B(s)^t\mathbf{1} = \pi[A(s)^t - g_p(s)^t P(s)]\mathbf{1} = \pi A(s)^t\mathbf{1} - g_p(s)^t\pi P(s)\mathbf{1}. \tag{5.8}$$

Noting that, $\pi\phi_d(s) \leq \mathbf{1}^T\phi_d(s) = 1$, by (5.3) we have

$$g_p(s)^t \pi P(s)\mathbf{1} = g_p(s)^t \pi\phi_d(s)\phi_i^T(s)\mathbf{1} \leq g_p(s)^t \phi_i^T(s)\mathbf{1}. \tag{5.9}$$

For sufficiently large t, applying (5.7) we have that

$$\pi B(s)^t \leq \max_{i,j} \left| b_{ij}^{(t)} \right| \mathbf{1}^T \leq g_p(s)^t \mathbf{1}^T,$$

therefore,

$$\pi B(s)^t \mathbf{1} \leq g_p(s)^t k. \tag{5.10}$$

From (5.8), (5.9), and (5.10), for sufficiently large t we thus obtain

$$g_p(s)^t[k + \phi_i^T(s)\mathbf{1}] = g_p(s)^t k + g_p(s)^t \phi_i^T(s)\mathbf{1} \geq \pi A(s)^t \mathbf{1}. \tag{5.11}$$

Since $\phi_i(s) \geq 0$, we have that $0 < k + \phi_i^T(s)\mathbf{1} < \infty$, so we conclude from (5.11)

$$\lim_{t\to\infty} \left\{\pi A(s)^t \mathbf{1}\right\}^{1/t} \leq \lim_{t\to\infty} g_p(s)[k + \phi_i^T(s)\mathbf{1})]^{1/t} = g_p(s),$$

and consequently

$$\rho(\exp(Q + sH)) \geq \lim_{t\to\infty} \{\pi \exp[(Q + sH)t]\mathbf{1}\}^{1/t}. \tag{5.12}$$

Next, we will show that $\rho(\exp(Q + sH)) \leq \lim_{t\to\infty}\{\pi \exp[(Q + sH)t]\mathbf{1}\}^{1/t}$.

We assume that $A(s)\mathbf{1} \ll g_p(s)\mathbf{1}$, that is, $g_p(s) > \sum_{j=1}^k a_{ij}$, $\forall\ i = 1, \ldots, k$. Then, there exists $g(s)$ such that

$$A(s)\mathbf{1} \leq g(s)\mathbf{1} \tag{5.13}$$

$$g_p(s) > g(s). \tag{5.14}$$

Applying (5.13) and Proposition 2, we have that $g_p(s) \leq g(s)$, which contradicts (5.14). Then

$$A(s)\mathbf{1} \geq g_p(s)\mathbf{1}.$$

Since $A(s) > 0$ for $s > 0$, it can be shown by induction on m that $A(s)^m\mathbf{1} \geq g_p(s)^m\mathbf{1}$, particularly,

$$\pi A(s)^t \mathbf{1} \geq \pi g_p(s)^t \mathbf{1} = g_p(s)^t,$$

concluding that

$$\lim_{t\to\infty} \{\pi A(s)^t \mathbf{1}\}^{1/t} \geq \lim_{t\to\infty} [g_p(s)^t]^{1/t} = g_p(s) = \rho(\exp(Q + sH)). \tag{5.15}$$

The result follows from (5.12) and (5.15). □

We now prove Theorem 5.

Proof of Theorem 5. From Lemma 1 we have that

$$\begin{aligned}\log \rho(\exp(Q+sH)) &= \log \lim_{t\to\infty} \{\pi \exp[(Q+sH)t]\mathbf{1}\}^{1/t} \\ &= \lim_{t\to\infty} \frac{1}{t} \log\{\pi \exp[(Q+sH)t]\mathbf{1}\}.\end{aligned} \tag{5.16}$$

Since $Q+sH$ is non-negative and irreducible, we have that $\rho(\exp(Q+sH)) = \exp(\rho(Q+sH))$. Then we can rewrite (5.16) as:

$$\rho(Q+sH) = \log \exp(\rho(Q+sH)) = \lim_{t\to\infty} \frac{1}{t} \log\{\pi \exp[(Q+sH)t]\mathbf{1}\}.$$

Finally, by a convenient operation, we have (5.1), so the theorem is proved. □

We thus demonstrate that it is possible to characterize the EB of a GMFM modulated source by the maximum real eigenvalue of a matrix that depends on the source parameters, network resources and service requirements, when the most likely duration time of the buffer busy period prior to overflow becomes larger and larger.

6. Numerical results

In this section we present some numerical results to evaluate the EB for values of the parameter t of increasing time, in order to verify that in the limit this value coincides with the maximum real eigenvalue of the matrix $Q+sH$, as we proved in Theorem 5.

Traffic simulations, generated by Markov Chain Monte Carlo algorithms, were performed according to the model presented in Section 3. This algorithm has the ability to efficiently explore the parameter space of complex models. Moreover, given the ergodicity of the modulant chain, generating a single trace, the method guarantees convergence to the target distribution, and provides accurate and representative estimates, when the generated chain is long enough. The modulant Markov chain has 9 states, each one associated to a data transfer rate interval, as shown in Table 1.

Table 1: Theoretical dispatch rates.

State	Transfer rate (Mbps)
1	$(0, 1024]$
2	$(1024, 2048]$
3	$(2048, 3072]$
4	$(3072, 4096]$
5	$(4096, 5120]$
6	$(5120, 6144]$
7	$(6144, 7168]$
8	$(7168, 8192]$
9	$(8192, 10240]$

To design the infinitesimal generator Q of the chain, we consider that it can pass from one state to another with the same probability, so we take

$$Q = \begin{bmatrix} -8 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -8 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -8 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -8 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -8 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & -8 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & -8 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & -8 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -8 \end{bmatrix}.$$

The GMFM can be interpreted as a way to reduce the Markov fluid model when its dimension is very large, grouping similar dispatch rates as a noise around a central value, so it is appropriate to model them by a Gaussian distribution. Then within each interval, we assume that the quantity actually dispatched is drawn by a Gaussian distribution with mean equal to the midpoint and variance equal to one-sixth of its length. The diagonal matrix H contains in its main diagonal the mean values of these distributions.

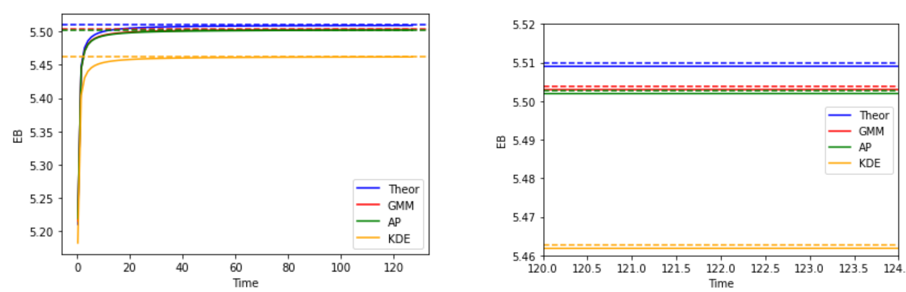
The simulated trace is a succession of pairs (v_i, t_i), where v_i is the transfer rate, t_i is the time when the chain jumps to another state, so the link is transferred at rate v_i while $t_{i-1} < t < t_i$, for $i = 1, \ldots, 14000$, the number of jumps of the chain. To estimate the infinitesimal generator $Q = (q_{ij})_{i,j=1,\ldots,k}$ we use the estimation of its elements using the Lebedev-Lukashuk maximum likelihood estimator [12], which allows us to obtain an asymptotic Gaussian estimate of the elements of Q as a function of the traffic trajectories or traces. They are $q_{ij}^{(n)}(x) = \frac{\gamma(i,j,nx)}{\tau(i,nx)}$, where $q_{ij}^{(n)}$ represents the estimated element of the (i, j) position of the Q matrix, $\gamma(i, j, h)$ is the number of transitions of the string from i to j in the interval $[0, h]$ and $\tau(i, h)$

is the time the string remained in the i state during the interval $[0, h]$. More information about this maximum likelihood estimator can be found at [2].

Simulations were performed in Python 3.7 using sklearn.neighborslibrary and codes can be provided by asking the authors.

From simulated trace data, we estimated the EB using the Kernel Density Estimator (KDE), the Gaussian Mixture Model (GMM) and the Affinity Propagation (AP) method. In [5] we compared the results obtained by calculating partition comparison indices, concluding that the most accurate methods are the Gaussian Mixture Model followed by the Kernel Density Estimator method.

From the results obtained from these estimations, we study the behavior of EB for increasingly larger t values in order to verify numerically the result obtained by the Theorem 5. Without loss of generality we consider $s = 1$. Figure 1a shows the behavior of the theoretical (Theor) and estimated EB by the different methods for sufficiently large t values. The dotted horizontal lines indicate the values of the eigenvalues of Perron-Frobenius for both the theoretical EB and the one estimated by the different methods.



(a) Convergence to Perron-Frobenius eigenvalue.

(b) Zoom Figure 1a.

Figure 1: Approximation of the theoretical and estimated Effective Bandwith to the Perron-Frobenius eigenvalue.

7. Conclusions

In this paper we have shown that it is possible to describe the equilibrium buffer distribution for a GMFM modulated source by means of a set of differential equations. Furthermore, we have shown, as a main result, that for this type of sources it is possible to characterize the EB through the maximum eigenvalue of a matrix that depends on the source parameters, the network resources and the service requirements. In this way we obtained a simple expression of the asymptotic behavior of the buffer content, which is extremely useful for network design when sources are bursty, and numerically easy to calculate. This novel result is a consequence of the application of the Perron-Fronenius theory to non-negative matrices, whose components are not deterministic but depend on the moments of the distributions involved in the model. Finally we have numerically verified this main result for simulated traffic traces.

Acknowledgments

This work was partially supported by the Research Group Project PGI 24/L112 of the Universidad Nacional del Sur.

References

[1] D. Anick, D. Mitra, M. Sondhi, *Stochastic theory of a data-handling system with multiple sources*. Bell System Tech. J. **61**(1982), no. 8, 1871–1894. DOI: 10.1002/j.1538-7305.1982.tb03089.x

[2] J. Bavio, B. S. Marrón, *Properties of the estimators for the effective bandwidth in a generalized Markov fluid model*. Open Journal of Statistics **8**(2018), 69–84. DOI: 10.4236/ojs.2018.81006

[3] J. Cao, W. S. Cleveland, D. X. Sun, *Bandwidth estimation for best-effort internet traffic*. Stat. Sci. **19**(2004), no. 3, 518–543. DOI: 10.1214/088342304000000260

[4] C. Courcoubetis, V. A. Siris, *Procedures and tools for analysis of network traffic measurements*. Perform. Eval. **48**(2002), no. 1, 5–23. DOI: 10.1016/S0166-5316(02)00028-7

[5] C. Fernández, J. Bavio, B. Marrón, *Evaluation of clustering techniques to estimate the effective bandwidth of a Markovian fluid from traffic traces*. IEEE Latin America Transactions **21**(2023), no. 5, 636–642.

[6] I. Gelfand, *Normierte ringe*. Rec. Math. Moscou **9**(1941), 3–23.

[7] R. J. Gibbens, *Traffic characterisation and effective bandwidths for broadband network traces*. Stochastic Networks: Theory and Applications. O. U. Press (Ed.). 1996, 169–180. DOI: 10.1093/oso/9780198523994.003.0009

[8] R. Horn, C. Johnson, *Nonnegative matrices*. Matrix Analysis. C. U. Press (Ed.). 2012, 487–530. DOI: 10.1017/CBO9780511810817

[9] S. Karlin, H. M. Taylor, *Review of matrix analysis*. A First Course on Stochastic Processes. A. Press (Ed.). 2nd ed. 1975, 536–551. DOI: 10.1137/1009056

[10] F. Kelly, *Notes on effective bandwidths*. Stochastic Networks: Theory and Applications. O. U. Press (Ed.). 1996, 141–168. DOI: 10.1093/oso/9780198523994.003.0008

[11] G. Kesidis, J. Walrand, C. S. Chang, *Effective bandwidths for multiclass Markov fluids and other ATM sources*. IEEE/ACM Trans. Netw. **1**(1993), no. 4, 424–428. DOI: 10.1109/90.251894

[12] E. A. Lebedev, L. ILukashuk, *Estimation of the maximum-likelihood for the infinitesimal matrix of the Markov-chain with continuous-time*. dopovidi akademii nauk ukrainskoi rsr seriya a-fiziko-matematichni ta technichni nauki **1**(1986), 12–14.

[13] G. Veciana, J. Walrand, *Effective bandwidths: Call admission, traffic policing and filtering for ATM networks*. Queueing Syst. **20**(1995), 37–59. DOI: 10.1007/BF01158431