



Balancing Elastic Traffic Sources

Thomas Bonald, Minh-Anh Tran

► **To cite this version:**

Thomas Bonald, Minh-Anh Tran. Balancing Elastic Traffic Sources. IEEE Communications Letters, Institute of Electrical and Electronics Engineers, 2007, <10.1109/LCOMM.2007.070393>. <hal-01269830>

HAL Id: hal-01269830

<https://hal.inria.fr/hal-01269830>

Submitted on 5 Feb 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Balancing Elastic Traffic Sources

Thomas Bonald and Minh-Anh Tran

Abstract—We consider heterogeneous elastic traffic sources that dynamically share a common link. We prove that balancing these traffic sources decreases the mean throughput and increases the blocking probability in the presence of admission control. The result generalizes that of Dartois derived for telephone traffic.

Index Terms—Flow-level model, finite source, traffic balancing.

I. INTRODUCTION

Consider a set of sources that dynamically share a common link. This link may be part of a cable, DSL or optical backhaul network for instance, or the access link to a Web server. Each source generates a sequence of data flows of random size, with an idle period of random duration between the end of a flow and the beginning of the following flow. Traffic is elastic in the sense that the duration of data flows increases in congestion periods where there is a high number of active sources. The duration of idle periods, on the other hand, is assumed to be independent of congestion.

This model proves very useful for dimensioning purposes [1], [2], [3]. It is the analogue of the Engset model introduced in 1915 for telephone traffic [4]. The Engset formula, that relates the call blocking probability to the traffic intensity and the number of telephone lines, was later extended to account for heterogeneous traffic sources [5]. It turned out that the corresponding results were too complex for any practical purpose. This motivated the work of Dartois who proved in 1970 that traffic balancing increases the blocking probability [6]. Thus telephone networks can be safely dimensioned assuming homogeneous traffic sources.

In this paper, we prove the analogue of Dartois' result for elastic traffic. Specifically, we prove that traffic balancing decreases the mean throughput and increases the blocking probability in the presence of admission control, when a maximum number of active sources is imposed. Thus data networks can also be safely dimensioned assuming homogeneous traffic sources, despite the strong heterogeneity observed in practice.

The model is described in the next section. The impact of traffic balancing is presented in Section III. Section IV is devoted to admission control. Some extensions are presented in Section V. All proofs are deferred to the Appendix.

II. MODEL

Consider N sources sharing a common link of capacity C bit/s. Source i generates flows of mean size σ_i , with an idle period of mean duration $1/\nu_i$ between the end of a flow and the beginning of the following flow. Successive flow sizes and

idle durations of a given source have arbitrary distributions and may be correlated. We denote by $\mu_i = C/\sigma_i$ the virtual flow completion rate of source i when active and by $\rho_i = \nu_i/\mu_i$ the corresponding virtual load. We assume capacity sharing is perfectly fair in the sense that each active source gets throughput C/n when n sources are active.

A. Stationary distribution

Let x be the N -dimensional vector whose i th component x_i gives the activity state of source i : $x_i = 1$ if source i is active and $x_i = 0$ if source i is idle. We denote by $n = \sum_{i=1}^N x_i$ the number of active sources. By the insensitivity property, the stationary distribution of x is the same as if flow sizes and idle durations were independent and had exponential distributions [7], [8]. The system state x is then a reversible Markov process on the state space $\mathcal{X} = \{0, 1\}^N$, with stationary distribution:

$$\alpha(x) = \alpha(0)n! \prod_{i=1}^N \rho_i^{x_i}, \quad x \in \mathcal{X}. \quad (1)$$

This follows from the local balance equations:

$$\alpha(x)\nu_i = \alpha(x + e_i) \frac{\mu_i}{n+1}, \quad \forall x \in \mathcal{X} : x_i = 0, \quad (2)$$

where e_i denotes the unit vector whose i -th component is equal to 1. Denoting by $\mathcal{X}(n)$ the set of states where n sources are active, we deduce the stationary distribution of the number of active sources:

$$\pi(n) = \pi(0)n! \sum_{x \in \mathcal{X}(n)} \prod_{i=1}^N \rho_i^{x_i}, \quad n = 0, 1, \dots, N. \quad (3)$$

B. Mean throughput

We are interested in the mean throughput, defined as the ratio of the mean flow size to the mean flow duration. Let τ_i be the mean flow duration of source i . The mean throughput of source i is given by:

$$\gamma_i = \frac{\sigma_i}{\tau_i}.$$

Let p_i be the probability that source i is active. By Little's law, the flow arrival rate λ_i of source i satisfies:

$$\tau_i = \frac{p_i}{\lambda_i}, \quad \frac{1}{\nu_i} = \frac{1-p_i}{\lambda_i}.$$

We deduce:

$$\gamma_i = C \frac{\rho_i(1-p_i)}{p_i}. \quad (4)$$

Now consider the mean throughput averaged over all sources:

$$\gamma = \frac{\sigma}{\tau},$$

Thomas Bonald is with Orange Labs, France,
e-mail: thomas.bonald@orange-ftgroup.com.
Minh-Anh Tran is with Ecole Normale Supérieure, France,
e-mail: minh-anh.tran@polytechnique.org.

where σ and τ denote the mean flow size and the mean flow duration, respectively:

$$\sigma = \frac{1}{\lambda} \sum_{i=1}^N \lambda_i \sigma_i \quad \text{and} \quad \tau = \frac{1}{\lambda} \sum_{i=1}^N \lambda_i \tau_i, \quad \text{with} \quad \lambda = \sum_{i=1}^N \lambda_i.$$

Let p be the probability that at least one source is active. Using the local balance equations (2), we get:

$$\begin{aligned} \sum_{i=1}^N \lambda_i \sigma_i &= \sum_{i=1}^N (1 - p_i) \nu_i \sigma_i, \\ &= \sum_{i=1}^N \sum_{x: x_i=0} \alpha(x) \nu_i \sigma_i, \\ &= \sum_{i=1}^N \sum_{x: x_i=0} \alpha(x + e_i) \frac{\mu_i}{n+1} \sigma_i, \\ &= C \sum_{i=1}^N \sum_{x: x_i=1} \frac{\alpha(x)}{n}, \\ &= C \sum_{x \neq 0} \alpha(x) \sum_{i: x_i=1} \frac{1}{n} = Cp. \end{aligned}$$

Since

$$\sum_{i=1}^N \lambda_i \tau_i = \sum_{i=1}^N p_i = \mathbb{E}[n],$$

we deduce:

$$\gamma = C \frac{p}{\mathbb{E}[n]}.$$

III. TRAFFIC BALANCING

In this section, we evaluate the impact of balancing the traffic sources. Specifically, we assume all sources generate flows of mean size σ separated by idle periods of mean duration $1/\nu$. We denote by $\mu = C/\sigma$ the virtual flow completion rate of each source and by $\rho = \nu/\mu$ the corresponding virtual load. We assume the overall load is conserved, that is:

$$\rho = \frac{1}{N} \sum_{i=1}^N \rho_i. \quad (5)$$

A. Stationary distribution

In view of (3), the stationary distribution of the number of active sources is given by:

$$\begin{aligned} \pi'(n) &= \pi'(0) n! \rho^n \sum_{x \in \mathcal{X}(n)} 1, \\ &= \pi'(0) \frac{N!}{(N-n)!} \rho^n. \end{aligned} \quad (6)$$

We have the following key result:

Theorem 1: The number of active sources is higher in distribution when the traffic sources are balanced:

$$\sum_{n \geq m} \pi'(n) \geq \sum_{n \geq m} \pi(n), \quad \text{for all } m = 0, 1, \dots, N.$$

B. Mean throughput

The higher number of active sources results in a lower mean throughput:

Theorem 2: The mean throughput is lower when the traffic sources are balanced.

Note that this result is true for the mean throughput averaged over all sources, not for the mean throughput of each source. In the case of $N = 2$ sources for instance, it follows from (4) that:

$$\gamma_1 = C \frac{1 + \rho_2}{1 + 2\rho_2}, \quad \gamma_2 = C \frac{1 + \rho_1}{1 + 2\rho_1}.$$

Thus a source benefits from traffic balancing if and only if the other source contributes most to traffic load.

IV. ADMISSION CONTROL

We now limit the number of active sources to a fixed value $M \leq N$, which guarantees a minimum throughput of C/M to all data transfers. If a source attempts to generate a new flow while M sources are active, this flow is blocked and the source reenters the idle state. We are interested in both the blocking probability and the mean throughput.

A. Stationary distribution

By the insensitivity and the reversibility properties, the stationary distribution α of the system state in the presence of admission control is the restriction of the stationary distribution (1) to the set of states \mathcal{X} such that $n \leq M$. In particular, the stationary distributions π and π' of the number of active sources without and with traffic balancing, respectively, are the restrictions of the stationary distributions (3) and (6) to $n \leq M$. We have the analogue of Theorem 1:

Theorem 3: The number of active sources is higher in distribution when the traffic sources are balanced:

$$\sum_{n \geq m} \pi'(n) \geq \sum_{n \geq m} \pi(n), \quad \text{for all } m = 0, 1, \dots, M.$$

B. Blocking probability

Let $\mathcal{X}_i \subset \mathcal{X}$ be the set of states where source i is idle and $\mathcal{X}_i(n) \subset \mathcal{X}_i$ be the set of states where source i is idle and n sources are active. The blocking probability of source- i flows is equal to the probability that M sources are active while source i is idle, that is:

$$B_i = \frac{\sum_{x \in \mathcal{X}_i(M)} \alpha(x)}{\sum_{x \in \mathcal{X}_i} \alpha(x)}.$$

The average blocking probability is given by:

$$B = \frac{\sum_{i=1}^N \nu_i \sum_{x \in \mathcal{X}_i(M)} \alpha(x)}{\sum_{i=1}^N \nu_i \sum_{x \in \mathcal{X}_i} \alpha(x)}. \quad (7)$$

It is not true in general that traffic balancing increases the blocking probability. For $N = 2$ sources and at most $M = 1$ active source for instance, we get:

$$B_1 = \frac{\rho_2}{1 + \rho_2} \quad \text{and} \quad B_2 = \frac{\rho_1}{1 + \rho_1}.$$

Thus a source benefits from traffic balancing if and only if the other source contributes most to traffic load. For the average blocking probability, we obtain:

$$B = \frac{\nu_1 \rho_2 + \nu_2 \rho_1}{\nu_1(1 + \rho_2) + \nu_2(1 + \rho_1)}.$$

which is not maximum for $\rho_1 = \rho_2$ in general.

A sufficient condition for this property to hold is that all sources have the same mean idle durations of mean $1/\nu$ in the original system with unbalanced traffic sources. It then follows from (7) that

$$B = \frac{(N - M)\pi(M)}{\sum_{n=0}^M (N - n)\pi(n)} = \frac{(N - M)\pi(M)}{N - \mathbb{E}[n]}$$

and the result is a consequence of Theorem 3.

C. Mean throughput

Unlike the blocking probability, the mean throughput is always reduced by traffic balancing independently of the original mean idle durations:

Theorem 4: The mean throughput is lower when the traffic sources are balanced.

V. EXTENSIONS

The results extend to the case where each source doesn't have full access to the link capacity but is limited by some fixed bit rate representing the speed of its access line for instance. Assuming all sources have the same rate limit, the corresponding model is a processor-sharing queue with state-dependent service rate [8]. In the particular case where the admission policy does not allow any elastic sharing, the system reduces to the Engset model and the result on the blocking probability corresponds to that of Dartois [6].

APPENDIX

The proofs are based on the following two lemmas:

Lemma 1: Let a_1, \dots, a_N , b_1, \dots, b_N and t_1, \dots, t_N be positive numbers such that $a_1/b_1 \geq a_2/b_2 \geq \dots \geq a_N/b_N$ and $t_1 \leq t_2 \leq \dots \leq t_N$. Then:

$$\frac{\sum_{i=1}^N t_i a_i}{\sum_{i=1}^N t_i b_i} \leq \frac{\sum_{i=1}^N a_i}{\sum_{i=1}^N b_i}.$$

Proof. The inequality is equivalent to:

$$\begin{aligned} \sum_{i=1}^N t_i a_i \sum_{j=1}^N b_j &\leq \sum_{i=1}^N a_i \sum_{j=1}^N t_j b_j \\ \iff \sum_{i \neq j} (t_i - t_j) a_i b_j &\leq 0 \\ \iff \sum_{i > j} (t_i - t_j) (a_i b_j - a_j b_i) &\leq 0. \end{aligned}$$

□

Lemma 2: Let a_1, \dots, a_N be positive numbers such that: $\frac{1}{N} \sum_{i=1}^N a_i = 1$. For all $n = 1, \dots, N$, let:

$$S(n) = \sum_{x \in \mathcal{X}(n)} \prod_{i=1}^N a_i^{x_i}.$$

Then:

$$\frac{S(n)}{\binom{N}{n}} \geq \frac{S(n+1)}{\binom{N}{n+1}} \text{ for all } n = 1, \dots, N-1.$$

Proof. For all $n = 1, \dots, N-1$, we have:

$$S(1)S(n) = \sum_{i=1}^N a_i S(n) = (n+1)S(n+1) + U(n),$$

with

$$U(n) = \sum_{i=1}^N a_i^2 \sum_{x \in \mathcal{X}(n-1), x_i=0} \prod_{j=1}^N a_j^{x_j}.$$

Now

$$\begin{aligned} (N-n)U(n) &= \sum_{i \neq j} (a_i^2 + a_j^2) \sum_{x \in \mathcal{X}(n-1), x_i=x_j=0} \prod_{k=1}^N a_k^{x_k}, \\ &\geq \sum_{i \neq j} 2a_i a_j \sum_{x \in \mathcal{X}(n-1), x_i=x_j=0} \prod_{k=1}^N a_k^{x_k}, \\ &= 2 \binom{n+1}{2} S(n+1). \end{aligned}$$

We deduce:

$$\begin{aligned} S(1)S(n) &\geq (n+1)S(n+1) + \frac{2}{N-n} \binom{n+1}{2} S(n+1), \\ &= N \frac{n+1}{N-n} S(n+1). \end{aligned}$$

The result then follows from the fact that $S(1) = N$. □

Using (5) and Lemma 2, we get:

$$\frac{\sum_{x \in \mathcal{X}(n)} \prod_{i=1}^N \left(\frac{\rho_i}{\rho}\right)^{x_i}}{\binom{N}{n}} \geq \frac{\sum_{x \in \mathcal{X}(n+1)} \prod_{i=1}^N \left(\frac{\rho_i}{\rho}\right)^{x_i}}{\binom{N}{n+1}}.$$

We deduce $\pi(n)/\pi'(n) \geq \pi(n+1)/\pi'(n+1)$. The proof of Theorems 1 and 3 then follows from the application of Lemma 1 to the positive numbers $t_i = 0$ if $i < m$ and $t_i = 1$ otherwise, for all $i = 0, \dots, N$ and $i = 0, \dots, M$, respectively; that of Theorems 2 and 4 follows from the application of Lemma 1 to the positive numbers $t_i = i$ for all $i = 1, \dots, N$ and $i = 1, \dots, M$, respectively.

REFERENCES

- [1] D.P. Heyman, T.V. Lakshman, A.L. Neidhardt, A new method for analysing feedback-based protocols with applications to engineering Web traffic over the Internet, in: *Proc. of ACM SIGMETRICS*, 1997.
- [2] A.W. Berger, Y. Kogan, Dimensioning bandwidth for elastic traffic in high-speed data networks, *IEEE/ACM Trans. Net.* 8(5) (2000) 643–654.
- [3] T. Bonald, P. Olivier, J. Roberts, Dimensioning high-speed IP access networks, in: *Proc. of ITC 18*, 2003.
- [4] T.O. Engset, On the calculation of switches in an automatic telephone system, in: *Tore Olaus Engset: The man behind the formula*, Eds: A. Myskja, O. Espvik, 1998.
- [5] J.W. Cohen, The Generalized Engset Formula, *Phillips Telecommunications Review* 18 (1957) 158–170.
- [6] J.P. Dartois, Lost call cleared systems with unbalanced traffic sources, in: *Proc. of ITC*, 1970.
- [7] R.F. Serfozo, *Introduction to Stochastic Networks*, Springer, 1999.
- [8] S. Ben Fredj, T. Bonald, A. Proutière, G. Régnié and J.W. Roberts, Statistical bandwidth sharing: A study of congestion at flow level, in: *Proc of ACM SIGCOMM*, 2001.