Optimal Routing Policy

R. Kühn and S. M. Mostafavi

***Abstract*— We investigate the problem of resource allocation in heterogeneous networks of computational resources. We provide an explicit analytical solution for a situation where the computational environment can be described by M/M/1 queueing theory. We illustrate the quality of our solution by comparing results with those obtained via a simple ad hoc resource allocation in large heterogeneous networks consisting of $N = 10^4$ nodes with computational resources distributed either uniformly in a given interval, or exponentially in $\mathbb{R}^+$.**

***Index Terms*— Optimal resource allocation, heterogeneous networks.**

I. Introduction

ROUTING, multiplexing and switching are principal techniques used to efficiently transfer and process information in digital networks. Intelligent routing is a major ingredient for enhancing functionality, performance, and flexibility of networks of computational resources. This is why a large amount of research had been conducted to devise efficient routing algorithms [1], [2], [3], for obtaining better quality of service (QoS) through routing [4], in particular on scheduling techniques in routers to achieve desired QoS objectives [5]. Some of this work falls into the category of shortest path problems which consist of finding the path along which the delay is minimized. A large number of algorithms has been proposed as a solution to this problem under a variety of conditions and constraints [6].

In this work we will model the delay for a fixed network where there exist several resources with server characteristics which follow the M/M/1 queueing model [7]. Data packets arrive according to a Poisson process and must be routed to parallel queues with exponential service time distributions. We investigate the problem of task allocation in this heterogeneous network of computational resources which is optimal in the sense of minimising the expected average delay for time of end-to-end processing of the data.

II. Problem Definition

In this letter we investigate the problem of optimal resource allocation in a heterogeneous network of (computational) resources.

The resource allocation model we are looking at is characterized as follows:

Definition 2.1: A heterogeneous network consisting of n resources is characterized by the set $\boldsymbol{\mu} = \{\mu_1, \mu_2, \dots, \mu_n\}$ of values for the individual resources. We shall denote by $\mu = \sum_{i=1}^n \mu_i$ the sum of resources available in the system.

A resource could be a node in a network of processors, measured by its processing capability (e.g. in floating point operations per second), or the data transmission capacity of a given communication channel (e.g. in bits/sec).

Definition 2.2: We denote by $\lambda = \lambda(t)$ the time dependent computational demand or total traffic on the network; the set $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$ represents the distribution of the load across the set of resources with $\lambda_i = \lambda_i(t)$ denoting the share of the load allocated to resource i. Clearly, $\lambda = \sum_{i=1}^n \lambda_i$.

Assuming an M/M/1 queueing model for the computational environment, we will have an expected delay D_i for completion of (computational) tasks at resource i of the form

$$D_i = \frac{1}{\mu_i - \lambda_i} \qquad \text{(II.1)}$$

The optimization problem we are going to solve here, is to allocate computational tasks across the heterogeneous network of resources in such a way that the *average delay per computational task*,

$$D = \frac{1}{\lambda}\sum_{i=1}^n \lambda_i D_i = \frac{1}{\lambda}\sum_{i=1}^n \frac{\lambda_i}{\mu_i - \lambda_i}\ , \qquad \text{(II.2)}$$

is *minimized*. In this expression one readily identifies $p_i = \lambda_i/\lambda$ as the fraction of the total load allocated to resource i. The minimizing optimal resource allocation shall be denoted by $\boldsymbol{\lambda}^o$, and will be computed in Section III.

III. Optimal Resource Allocation

The aim is to find the minimum of the average delay $D = D(\boldsymbol{\lambda}|\boldsymbol{\mu})$ as given by (II.2), subject to the constraint that $\sum_{i=1}^n \lambda_i = \lambda$, within the space of permissible $\boldsymbol{\lambda}$, i.e. $\mu_i > \lambda_i \ge 0$ for all i.

As the space of permissible solutions is defined in terms of a set of inequalities, the theoretical framework used to solve the present optimization problem is Karush-Kuhn-Tucker theory [8].

To this end one introduces the Lagrangian

$$\begin{aligned}\mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\phi}) &= D - \phi_0\left(\sum_{i=1}^n \lambda_i - \lambda\right) - \sum_{i=1}^n \phi_i\lambda_i \\ &= \frac{1}{\lambda}\sum_{i=1}^n \frac{\lambda_i}{\mu_i - \lambda_i} - \phi_0\left(\sum_{i=1}^n \lambda_i - \lambda\right) - \sum_{i=1}^n \phi_i\lambda_i\end{aligned} \qquad \text{(III.1)}$$

with $\boldsymbol{\phi} = (\phi_0, \phi_1, \dots, \phi_n)$ a set of Lagrangian multipliers; here ϕ_0 is introduced to deal with with the equality constraint $\sum_{i=1}^n \lambda_i = \lambda$, while the ϕ_i, $1 \le i \le n$ are introduced to handle the inequality constraints $\lambda_i \ge 0$. The constrained

Manuscript received October 25, 2007. The associate editor coordinating the review of this letter and approving it for publication was J. Holliday.

R. Kühn is with the Department of Mathematics, King's College, London, UK (e-mail: reimer.kuehn@kcl.ac.uk).

S. M. Mostafavi is with the Centre for Communication Systems Research, University of Surrey, Guildford GU2 7XH, UK.

Digital Object Identifier 10.1109/LCOMM.2008.071777.

optimum $\boldsymbol{\lambda}^o$ is then found by solving the so-called Karush-Kuhn-Tucker equations

$$\frac{\partial \mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\phi})}{\partial \lambda_i} = \frac{1}{\lambda}\left[\frac{1}{(\mu_i - \lambda_i)} + \frac{\lambda_i}{(\mu_i - \lambda_i)^2}\right] - \phi_0 - \phi_i = 0 \qquad \text{(III.2)}$$

for $i = 1, \ldots, n$, together with

$$\frac{\partial \mathcal{L}(\boldsymbol{\lambda}, \phi_0)}{\partial \phi_0} = \lambda - \sum_{i=1}^{n} \lambda_i = 0 \ , \qquad \text{(III.3)}$$

and

$$\phi_i \lambda_i = 0, \qquad \lambda_i \geq 0 \ , \qquad i = 1, \ldots, n \ . \qquad \text{(III.4)}$$

Eq. (III.3) expresses just the the global constraint of complete load distribution; Eqs. (III.4) address the inequality constraints $\lambda_i \geq 0$, $i = 1, \ldots, n$; one may distinguish between so-called *active* constraints for which $\lambda_i = 0$ for the optimal solution, allowing for the corresponding Lagrange multiplier ϕ_i to be non-zero, $\phi_i \neq 0$, and the so-called *in-active* constraints for which $\lambda_i > 0$, and hence $\phi_i = 0$ by (III.4).

Eqs (III.2), have an explicit solution for λ_i, for any given ϕ_0 and ϕ_i, which can be written in the form

$$\mu_i - \lambda_i = \sqrt{\frac{\mu_i}{\lambda(\phi_0 + \phi_i)}} \ , \qquad 1 \leq i \leq n \ . \qquad \text{(III.5)}$$

It is useful to give explicit versions for inactive and active constraints respectively, viz.

$$\mu_i - \lambda_i \ = \ \sqrt{\frac{\mu_i}{\lambda \phi_0}} \ , \qquad \lambda_i > 0 \ , \qquad \text{(III.6)}$$

$$\mu_i \ = \ \sqrt{\frac{\mu_i}{\lambda(\phi_0 + \phi_i)}} \ , \qquad \lambda_i = 0 \ . \qquad \text{(III.7)}$$

The version for the active constraints (III.7) fixes the unknown Lagrangian multipliers ϕ_i in terms of the corresponding computational resource μ_i, once ϕ_0 is known.

In order to determine the unknown ϕ_0, we sum (III.5) over all i, and use the conditions described by (III.6) and (III.7) for the inactive and active constraints, respectively, giving $\mu - \lambda = \sum_{i=1}^{n}(\mu_i - \lambda_i)$, thus

$$\mu - \lambda \ = \ \sum_{i=1}^{n}\left[\sqrt{\frac{\mu_i}{\lambda \phi_0}}\Theta\left(\mu_i - \sqrt{\frac{\mu_i}{\lambda \phi_0}}\right) + \mu_i \Theta\left(\sqrt{\frac{\mu_i}{\lambda \phi_0}} - \mu_i\right)\right] \qquad \text{(III.8)}$$

By solving this equation for ϕ_0, one obtains the unknown value of the Lagrangian parameter ϕ_0 as a function of the total traffic λ, for given $\boldsymbol{\mu}$ defining the collection of resources.

Inserting the solution $\phi_0 = \phi_0(\lambda|\boldsymbol{\mu})$ of (III.8) into (III.6) and (III.7) one obtains the following solution to the optimal load distribution problem

$$\lambda_i^o = \max\left(\mu_i - \sqrt{\frac{\mu_i}{\lambda \, \phi_0}} \ , \ 0\right) \qquad \text{(III.9)}$$

The actual value of the minimal average delay per computational task $D_{\rm opt} = D(\lambda|\boldsymbol{\mu})$ is then simply obtained by inserting $\lambda_i = \lambda_i^o$ from (III.9) into (II.2), giving

$$D_{\rm opt} \ = \ \frac{1}{\lambda}\sum_{i=1}^{n}\frac{\lambda_i^o}{\mu_i - \lambda_i^o}$$
$$= \ \frac{1}{\lambda}\sum_{i=1}^{n}\left(\sqrt{\lambda \phi_0 \mu_i} - 1\right)\Theta\left(\sqrt{\lambda \phi_0 \mu_i} - 1\right) \qquad \text{(III.10)}$$

Here we have used the fact that for a valid solution $\mu_i - \sqrt{\frac{\mu_i}{\lambda \phi_0}} > 0$, if and only if $\sqrt{\lambda \phi_0 \mu_i} - 1 > 0$.

Note the following features: (i) The optimal values λ_i^o for the task allocation (routing) problem depend *only* on the properties of the set of resources, i.e. on $\boldsymbol{\mu}$, and of course on the total traffic λ. (ii) Whenever the capacity of the entire network of resources is not exceeded, i.e. whenever $\mu \geq \lambda$, the optimal solution leads to a complete and feasible allocation of tasks to the the system of resources. (iii) Once the solution $\phi_0 = \phi_0(\lambda|\boldsymbol{\mu})$ of (III.8) for a given system $\boldsymbol{\mu}$ of resources is known as a function of the traffic λ — a problem that can be numerically solved upfront and its solution stored, e.g., in a simple look-up table or in the form of some interpolating function — the solution (III.9) is *explicit*.

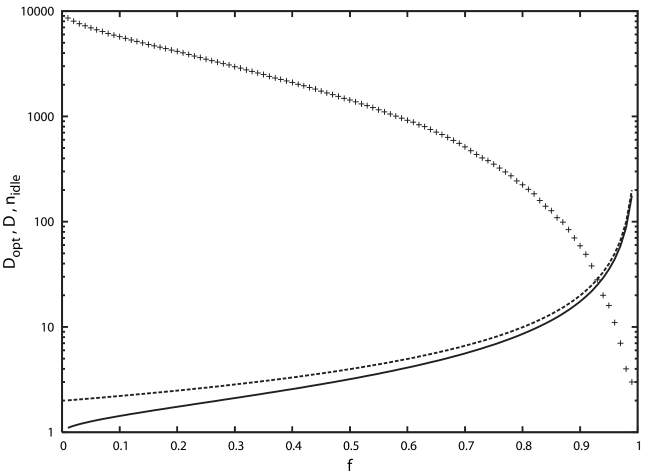

Fig. 1. Heterogeneous network of $N = 10^4$ resources with μ_i distributed uniformly in [0,1]. Shown are the number of idle resources (crosses), the average delay D per package in the ad hoc solution (dashed line) and the average delay per package $D_{\rm opt}$ in the optimal solution (full line), as a functions of the ratio $f = \lambda/\mu$ of total traffic and total computational resource.

We illustrate the quality of the solution in Fig. 1 for a random network of $N = 10^4$ resources, with μ_i uniformly distributed in $[0, 1]$, by comparing the cost per package for the optimal solution as given by (III.9), with that given by a simple alternative ad-hoc solution, where $\lambda_i = \frac{\lambda}{\mu}\mu_i$, i.e. tasks are allocated in such a way that each resource utilizes the same fraction of its computational power. The speed-up is most pronounced in the low load limit, where the optimal solution reduces the delay by a factor close to 2.

Fig. 1 also exhibits the number of idle processes as a function of of the ratio $f = \lambda/\mu$ of total traffic and total computational resource. It shows that at low computational load, speed-up is obtained by distributing the load only over the most powerful resources in the system, and that weaker resources are recruited progressively only with increasing load. Fig. 2 shows the behaviour of the Lagrange parameter

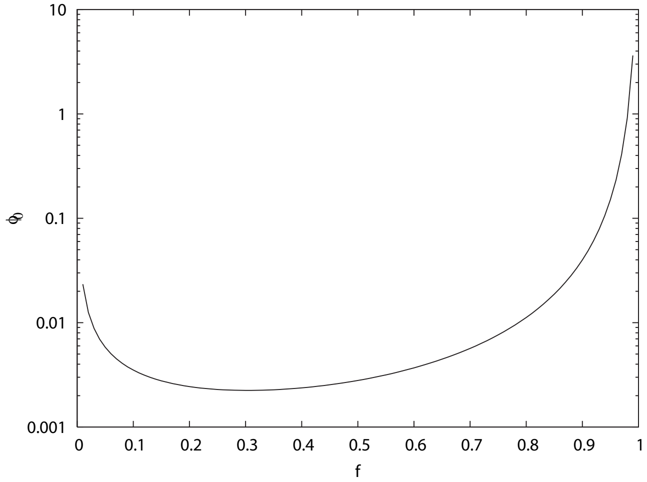

Fig. 2. Lagrange parameter ϕ_0 for the system of resources studied in Fig. 1, as a function of the ratio $f = \lambda/\mu$ of total traffic and total computational resource.

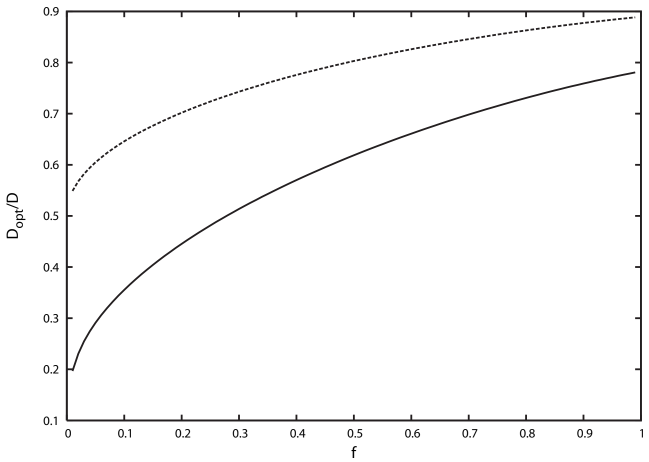

Fig. 3. Ratio $D_{\rm opt}/D$ of average delays in the optimal and the heuristic solution as a function of f for resources uniformly distributed in $[0, 1]$ (upper curve), and for exponentially distributed computational resources with probability density function $p(\mu) = \exp(-\mu)$ for the values μ_i of the resources (lower curve).

$\phi_0 = \phi_0(\lambda|\boldsymbol{\mu})$, here expressed as a function of f.

It is worth noting that the performance of the system, as given by the value of $D_{\rm opt}(\lambda|\boldsymbol{\mu})$ as a function of $f = \lambda/\mu$, as well as the value of the Lagrange parameter ϕ_0 will in the large N limit depend *only* on the *statistical* properties of the μ_i. The same is true for the actual speed-up that the the optimal solution realizes in comparison with the heuristic solution. We illustrate this in Fig. 3 by plotting the ratio of delays $D_{\rm opt}(\lambda|\boldsymbol{\mu})/D$, both for the system with uniformly distributed computational resources, and for a system of the same size with exponentially distributed resources. The latter system exhibits greater heterogeneity, and the gain obtainable by optimizing according to our algorithm is therefore larger, reaching a factor greater than 5 in the low load limit.

IV. Conclusion

We have provided an explicit solution to an optimal routing problem for a heterogeneous network of (computational) resources.

We have illustrated our solution by comparing the performance of the optimal solution with that of a simple ad-hoc solution for a large heterogeneous network consisting of $N = 10^4$ resources, with μ_i uniformly distributed in $[0, 1]$, or exponentially distributed in $\mathbb{R}^+$. The gain obtained by optimizing resource allocation is most pronounced at low loads, where the space of feasible solutions is largest (roughly a factor 2 in the case of uniformly distributed resources, better than a factor 5 for exponentially distributed resources). The optimal solution allocates tasks mainly to the most powerful resources in the system, recruiting weaker ones only as the load increases.

Based on the above explicit solution, a number of meaningful other optimization problems might sensibly be addressed. Here we mention just two to give a flavour of the possibilities.

Assuming that the total load λ on the system over over its life time varies according given probability density function $p(\lambda)$, one might wish to *optimize the set* $\boldsymbol{\mu} = \{\mu_1, \mu_2, \ldots, \mu_n\}$ *of computational resources itself* — on average over the distribution of loads — subject to further constraints such as $\mu = \sum_{i=1}^{n} \mu_i = \lambda_{\max} + \Delta$ where $\lambda_{\max}$ denotes a hypothetical maximum load and Δ a desired safety margin.

Additional constraints might come into play, like budgetary constraints: denoting by b_i the cost of acquiring a unit of computational power of the type of resource i, a fixed budget constraint of the form $\sum_{i=1}^{n} \mu_i b_i = B$ may exist. Budget constraints for acquisition and runtime costs may be separately considered, and so on.

All of these 'second order' optimization problems would benefit from the fact that the resource allocation or routing problem addressed above has an *explicit* solution. Note that the fact that ϕ_0 is known only *implicitly* as solution of (III.8), while complicating the solution of these second order problems, does not pose problems of principle, though their solution will generally require numerical tools. As the second order problems mentioned are concerned with one-off investment and design decisions, this aspect remains unproblematic.

We have not discussed the problem of (optimal) load shedding when the capacity of the network of resources is exceeded, i.e., when $\lambda(t) > \mu$.

References

[1] A. Orda, R. Rem, and M. Sidi, "Minimum delay routing in stochastic network," *IEEE/ACM Trans. Networking*, vol. 1, no. 2, pp. 187–198, 1993.

[2] S. Chen and K. Nahrstedt, "An overview of quality-of-service routing for the next generation high-speed networks: problems and solutions," *IEEE Network Mag.*, vol. 12, no. 6, pp. 64–79. 1998.

[3] Y. D. Lin, N. B. Hsu, and R.-H. Hwang, "QoS routing granularity in MPLS networks," *IEEE Commun. Mag.*, vol. 46, no. 2, pp. 58–65, 2002.

[4] S. Nelakuditi and Z.-L. Zhang, "A localized adaptive proportioning approach to QoS routing granularity," *IEEE Commun. Mag.*, vol. 46, no. 2, pp. 66–71, 2002.

[5] F. Hao, E. W. Zegura, and M. H. Ammar, "QoS routing for anycast communications: motivation and an architecture for Diffserv networks," *IEEE Commun. Mag.*, vol. 46, no. 2, pp. 48–56, 2002.

[6] N. Deo and C. Y., Pang, "Shortest path algorithms: taxonomy and annotation," *Networks*, vol. 14, pp. 275–323, 1984.

[7] L. Kleinrock, *Queuing Systems*, vol. 1. New York: Wiley Interscience, 1975.

[8] J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization. Theory and Examples*. New York: CMS Books in Mathematics, Springer-Verlag, 2000.