Load balancing and the power of preventive probing

B. Van Houdt
Department of Mathematics and Computer Science
University of Antwerp - IBBT
Antwerpen, Belgium
benny.vanhoudt@ua.ac.be

ABSTRACT

Consider a randomized load balancing problem consisting of a large number n of server sites each equipped with K servers. Under the greedy policy, clients randomly probe a site to check whether there is still a server available. If not, $d-1$ other sites are probed and the task is assigned to the site with the fewest number of busy servers. If all the servers are also busy in each of these $d-1$ sites, the task is lost.

This short paper analyzes a set of policies, i.e., (L,d) policies, that will occasionally probe additional sites even when there is still a server available at the site that was probed first. Using mean field methods, we show that these policies, that preventively probe other sites, can achieve the same loss probability while requiring a lower overall probe rate.

1. PROBLEM DESCRIPTION

Consider a network consisting of many clients that generate tasks and n server sites that each consist of K servers. Denote λn as the mean overall task generation rate and assume tasks are generated according to a Poisson process. We will refer to such a system as a (λ, K) grid, where $K \geq 1$ is an integer and $\lambda > 0$.

Clients use a so-called (L,d) policy, where, for now, L and d are assumed to be integers. Under an (L,d) policy a client will probe a random site whenever it generates a task. If there are less than L servers busy, the task is assigned to the server site, meaning the task is assigned after transmitting a single probe. Otherwise, the client will randomly probe another $d-1$ sites and will assign its task to the site with the fewest number of busy servers (ties are broken arbitrarily), hence in this case d probes were used to assign the task. Notice, if $L < K$ the task may end up in the site that was probed first. Finally, if all the servers are busy at each of the d probed sites, the task is considered lost.

The main question addressed in this paper is which (L,d) policy minimizes the average number of probes required per task for a given (λ, K) grid, such that the loss probability is below a predefined threshold ϵ. Setting $L = K$ corresponds to a greedy policy as we only send multiple probes whenever a single probe does not suffice. However, we will show that this policy is not always optimal as increasing L will imply that a larger d value will be required to obtain the targeted loss probability ϵ. For a fixed L, increasing d will reduce the loss rate and as such we can determine the required d as a function of L.

We will also consider non-integer values for L and d. For general L, the task is assigned to the first site with probability 1, if there are less than $\lfloor L \rfloor$ busy servers and with probability $L - \lfloor L \rfloor$ if there are $\lfloor L \rfloor$ busy servers. Although many numerical experiments indicate that non-integer L values are never optimal, they still provide inside into the erratic behavior of the curves when only integer L values are considered. Similarly, for general d values we state that $\lfloor d-1 \rfloor$ additional probes are send with probability $\lceil d \rceil - d$, while $\lceil d-1 \rceil$ more probes are send with probability $d+1-\lceil d \rceil$; hence, on average $d-1$ additional probes are used whenever the first probe did not result in a task assignment.

We will study the behavior of this stochastic system when the number of server sites n becomes large. When n approaches infinity the theory of density dependent Markov chains [3] shows that the system becomes deterministic and its behavior over time can be described by means of a system of ordinary differential equations (ODEs).

There have been many studies on load balancing and the power of $d > 1$ choices (see [1, 4, 5] and the references therein). These works typically consider server sites where each site has a single server and an infinite waiting room to store tasks and the main objective is to minimize the response time. We consider K servers per site and no waiting room and wish to minimize the required probe rate to guarantee a predefined loss probability.

2. THE MEAN FIELD MODEL

In this section we consider the system with integer L and d values, the generalization to arbitrary L and d values requires some additional care but is not hard. The service times of the tasks are assumed to be i.i.d. and exponentially distributed with a mean $\mu = 1$. Consider the system with n sites and define $X_i^{(n)}(t)$ as the proportion of the sites having i or more busy servers at time t. Then, $X^{(n)}(t) = (X_1^{(n)}(t), \ldots, X_K^{(n)}(t))$ for $t \geq 0$ is clearly a Markov chain and this chain is stable for any arrival rate $\lambda > 0$. It is not hard to show that this set of Markov chains forms a density dependent Markov chain as defined by Kurtz in [3].

Next, consider the deterministic system described by the following set of ODEs. To easy the notation, let $w_0(t) = 1$ and $w_{K+1}(t) = 0$ for all t. For $i = 1, \ldots, L$, define

$$\frac{d}{dt}w_i(t) = \lambda w_L(t)(w_{i-1}^{d-1}(t) - w_i^{d-1}(t)) + \lambda(w_{i-1}(t) - w_i(t)) - i(w_i(t) - w_{i+1}(t)), \quad (1)$$

while for $i = L+1, \ldots, K$ let

$$\frac{d}{dt}w_i(t) = \lambda(w_{i-1}^{d}(t) - w_i^{d}(t)) - i(w_i(t) - w_{i+1}(t)). \quad (2)$$

If we denote this system of ODEs as $\frac{d}{dt}w(t) = F(w(t))$,

with $w(t) = (w_1(t), \ldots, w_K(t))$ and F a function from $\mathbb{R}^K$ to $\mathbb{R}^K$, then it is not hard to see that F is Lipschitz on $E = \{(x_1, \ldots, x_K) | 0 \leq x_i \leq 1, i = 1, \ldots, K\}$. Therefore, the following theorem is immediate by Kurtz [3]:

Theorem 1. *Suppose that* $\lim_{n\to\infty} X^{(n)}(0) = w(0)$ *a.s. and consider the path* $\{w(u), u \leq t\}$*, then*

$$\lim_{n\to\infty} \sup_{u\leq t} |X^{(n)}(u) - w(u)| = 0 \quad a.s.$$

Kurtz's theorem states that up to time t the limiting process is indeed the deterministic process given by the above set of ODEs. We are however interested in the limit of the stationary distributions $\pi^{(n)}$ of the Markov chains $\{X^{(n)}(t), t \geq 0\}$. This limit will coincide with a fixed point of the system of ODEs if all the trajectories can be shown to converge to this fixed point (see [2, Corollary 5]). Using the L_1-norm as a Lyapunov function, the following theorem can be proven similar to [4, Theorem 3]:

Theorem 2. *The set of ODEs given by Equations (1-2) has a unique fixed point* π *in* E *and all the trajectories starting in* E *converge to* π *exponentially fast. More specifically,* $\sum_{i=1}^{K} |w_i(t) - \pi_i| \leq Ke^{-t}$.

This means the trajectories converge exponentially fast with parameter $\delta = 1$ and the fixed point can be determined numerically in a fraction of a second by a short simulation of the system of ODEs.

3. NUMERICAL RESULTS

Figure 1 shows the impact of d and L on the loss probability of a task for $K = 10$ and $\lambda = 6$, similar observations were made for other K and λ values. As expected, increasing d or decreasing L reduces the loss probability. More importantly, we see that the curves are not smooth whenever d or L becomes an integer. Thus, even if we send $\lceil d - 1 \rceil$ additional probes with a probability close to one, the loss may be substantially larger than sending $\lceil d - 1 \rceil$ additional probes with probability one. Similarly, assigning a task when the number of busy servers equals $\lfloor L \rfloor$ with a small probability may also result in a strong increase in the loss rate, compared to assigning the task with probability zero.

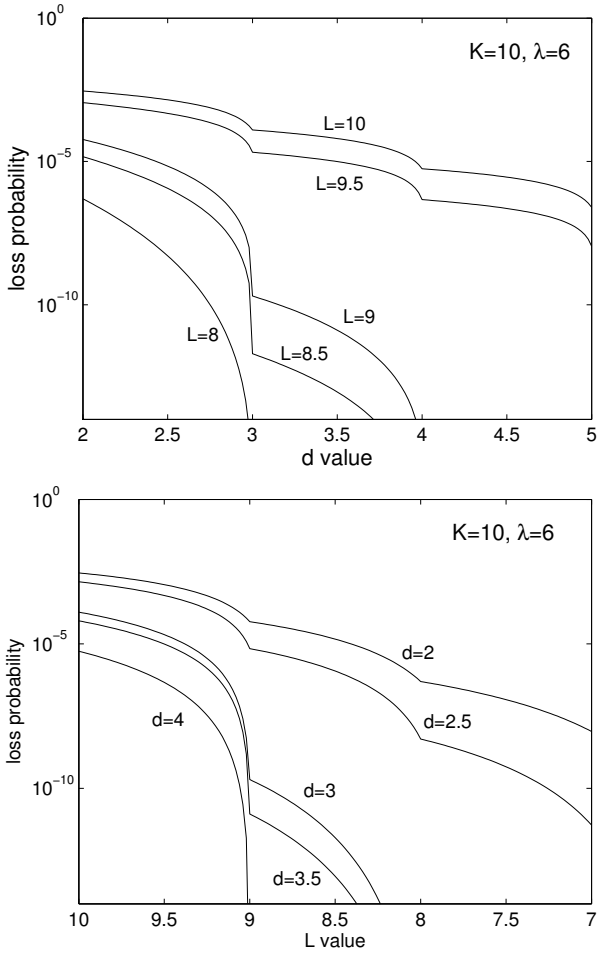

Figure 1: Impact of d and L on the loss probability for $K = 10$ and $\lambda = 6$

Figure 2(top) shows the required additional probe rate as a function of L to attain a loss probability of $\epsilon = 10^{-9}$ for $K = 10$ and $\rho = \lambda/K = 0.6$ and 0.7. The curves are not smooth whenever L is an integer, but also at some non-integer L values. For these latter values the required d value, which clearly increases with L, becomes an integer, as illustrated in Figure 2(bottom) depicting the required d value. This figure also indicates that the greedy policy (i.e., setting $L = K$) is not always optimal, as setting $L = K - 1$ is optimal for $K = 10$ and $\rho = 0.6$ or 0.7. The fact that the optimal L value is an integer does not appear to be a coincidence, other experiments also suggest that integer L values are optimal.

Next, let us get some insights on the integer value of L that results in the lowest probe rate. Before we generate any numerical results, it should be clear that setting L below λ is probably not very useful, as one should not try to avoid the assignment of a task to a server with a lower than average load. As such, for small K values, the greedy policy $L = K$ should be optimal. As K increases while the load $\rho = \lambda/K$ remains fixed, the $L = K - 1$ policy might also become eligible, followed by the $L = K - 2$ policy and so on. On the other hand, for very large K values, the greedy policy should also be optimal, as the loss probability in an Erlang loss system reduces to zero if the number of servers K grows to infinity (while the load remains fixed). Hence, no matter how small the targeted loss probability ϵ is, if K is sufficiently large there is no need to send additional probes. For instance, if $\rho = 0.5$ and the targeted loss probability is $\epsilon = 10^{-9}$, setting $d = 1$ suffices for $K \geq 91$ servers. In conclusion, we expect to see a finite region of K values where the greedy policy $L = K$ might be outperformed by a policy with $L < K$.

Figure 3(top) depicts the additional probe rate for the $L = K, K - 1$ and $K - 2$ policies when the load $\rho = 0.8$ and K ranges from 5 to 100 (further decreasing L did not result in a smaller probe rate). The curves in this figure are clearly not smooth and this behavior can once more be understood by looking at the corresponding required d value in Figure 3(bottom), where the curves tend to change direction whenever the required d value becomes an integer value. This also makes the regions in which a particular L value is optimal somewhat irregular.

For $\rho = 0.8$ and $K \leq 100$, we see that $L = K$ is optimal for $K \leq 7$, $L = K - 1$ is optimal for $8 \leq K \leq 33$ and for

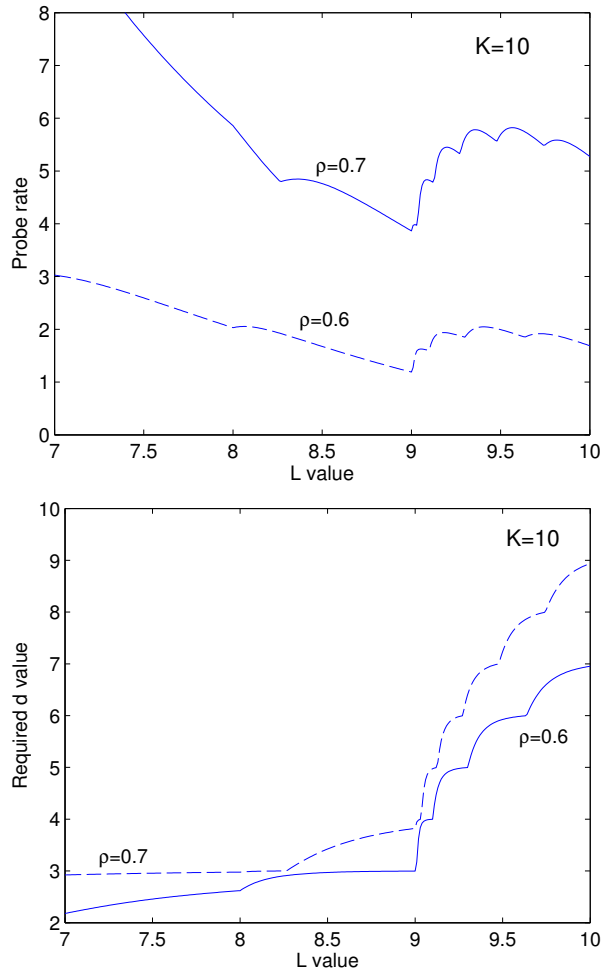

Figure 2: Impact of L on the required additional probe rate to achieve a loss probability of 10^{-9} and the required corresponding d value for $K = 10$ and $\lambda = 6$ and 7

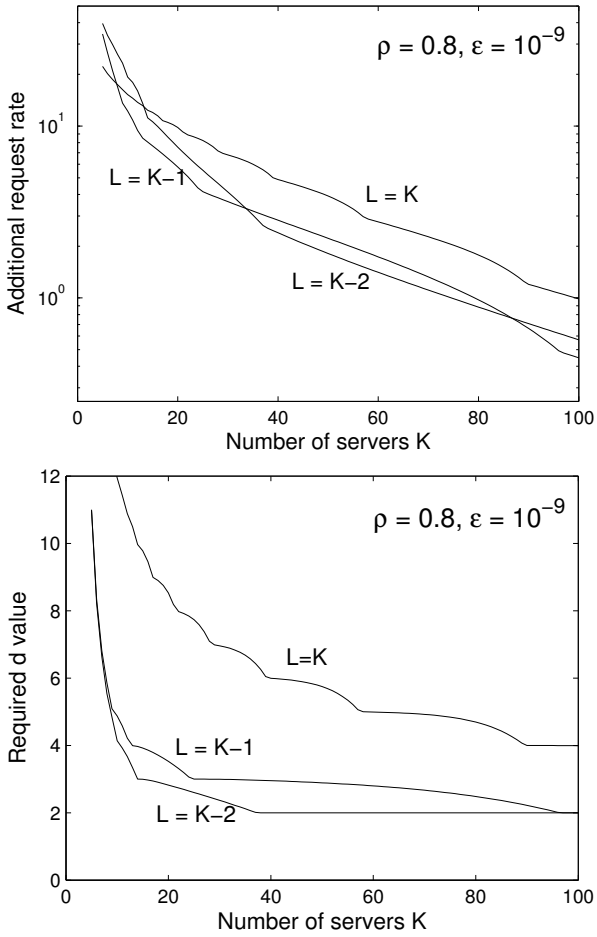

Figure 3: Optimal (integer) L value and its corresponding d value as a function of K with $\rho = 0.8$

$K \geq 87$, while having $L = K-2$ is optimal for $34 \leq K \leq 86$. When we reduce the load, the size of the region where the greedy $L = K$ policy can be outperformed tends to decrease as the optimal L value tends to increase (except for small K). For instance, for $\rho = 0.6$, the $L = K - 1$ policy is best for $6 \leq K \leq 61$, while setting $L = K$ is best for all other K values. When the load ρ decreases to 0.5 the greedy strategy is optimal for all K.

Additional experiments indicate that increasing the targeted loss probability ϵ also decreases the set of K values for which the greedy strategy can be outperformed. For instance, for $\rho = 0.6$ and $\epsilon = 10^{-6}$ the $L = K - 1$ policy only outperforms the greedy one for $7 \leq K \leq 33$.

Simulation experiments, not reported here, indicate that the accuracy of the mean field approximation is quite good for finite values of n. More specifically, for $\lambda = 6$ and 7, we simulated the system with $n = 100$ server sites, each equipped with $K = 10$ servers and determined the additional probe rate for $L = K, K - 1$ and $K - 2$, The value of d was determined by the mean field model, such that the loss rate (in the mean field) equaled 10^{-9}. The relative error of the mean field model observed in these experiments was less than 2 percent in all cases.

4. REFERENCES

[1] D. L. Eager, E. D. Lazowska, and J. Zahorjan. A comparison of receiver-initiated and sender-initiated adaptive load sharing. *Perform. Eval.*, 6(1):53–68, 1986.

[2] N. Gast and B. Gaujal. A mean field model of work stealing in large-scale systems. *SIGMETRICS Perform. Eval. Rev.*, 38(1):13–24, 2010.

[3] T. Kurtz. *Approximation of population processes.* Society for Industrial and Applied Mathematics, 1981.

[4] M. Mitzenmacher. The power of two choices in randomized load balancing. *IEEE Trans. Parallel Distrib. Syst.*, 12:1094–1104, October 2001.

[5] M. Mitzenmacher, A. Richa, and R. Sitaraman. The power of two random choices: a survey of techniques and results. *Handbook of Randomized Computing*, 1, 2001.