Optimizing Cloud Use under Interval Uncertainty

Vladik Kreinovich and Esthela Gallardo

Department of Computer Science
University of Texas at El Paso, El Paso, TX 79968, USA
vladik@utep.edu, egallardo5@miners.utep.edu

Abstract. One of the main advantages of cloud computing is that it helps the users to save money: instead of buying a lot of computers to cover all their computations, the user can rent the computation time on the cloud to cover the rare peak spikes of computer need. From this viewpoint, it is important to find the optimal division between in-house and in-the-cloud computations. In this paper, we solve this optimization problem, both in the idealized case when we know the complete information about the costs and the user's need, and in a more realistic situation, when we only know interval bounds on the corresponding quantities.

1 Formulation of the Problem

What is cloud computing. The main idea behind cloud computing (see, e.g., [8, 17, 22, 27]) is that instead of performing all the computations on his/her own computer, a user can sometimes rent computing time from a computer-time-rental company. This, in effect, is what is known as *cloud computing.* Computations that use rented computer time are called *computing in the cloud.*

Renting is usually more expensive than buying and maintaining one's own computer, so if the user needs the same amount of computations day after day, cloud computing is not a good financial option. However, if a peak need for computing occurs rarely, it is often cheaper to rent the corresponding computation time than to buy a lot of computing power and idle it most of the time.

How much computation time should we rent? Once the user knows his/her computational requirements, the proper question is: should we use the cloud at all? if yes, how much computing power should we buy for in-house computations and how much computation time should we rent from the cloud company? how much will it cost?

Finally, if a cloud company offers a multi-year deal with fixed rates, should we take it or should we buy computation time on a year-by-year basis?

Why this is important. Surprisingly, while the main purpose of cloud computing is to save user's money, most cloud users are computer folks with little knowledge of economics. As a result, often, they make wrong financial decisions about the cloud use; see, e.g., [28]. It is important to come up with proper recommendations for using cloud computing.

What we do in this paper. In this paper, we provide the desired financial recommendations, first under the idealized assumption that we have a complete information, and then, in a more realistic situation of interval uncertainty.

2 How Much Computations to Perform In-House and How Much in Cloud: Case of Complete Information

Case of complete information: description. Let us first consider the idealized case when we have complete information about our needs and about all the costs.

This means, first, that we know the cost of keeping a certain level of computational ability in-house. Let us pick some time quantum (e.g., day or hour). Then, the overall cost of buying and maintaining the corresponding computers is proportional to these computer's computational ability – i.e., the number of computing operations (e.g., Teraflops) that these computers can perform in this time unit. Let c_0 denote the cost per unit of computations. Then, if we buy computers with computational ability x_0, we pay $c_0 \cdot x_0$ for these computers.

This also means that we know the cost of computing in the cloud. Let us denote this cost by c_1. So, if one day, we need to perform x computations in the cloud, we have to pay the amount $c_1 \cdot x$.

As we have mentioned, computing in the cloud is usually more expensive than computing in-house. Part of this extra cost is the cost of moving data, another part is the overhead to support the computing staff, marketing staff, etc. As a result, $c_1 > c_0$.

Complete knowledge also means that we know the user's needs. This means that for each possible computation need x, we know the probability that one of the days, we will need to perform exactly x computations. These probabilities can be estimated by analyzing the previous needs: if we needed x computations in 10% of the days, this means that the probability of needing x computations is exactly 10%.

The probability distribution is usually described either by a cumulative distribution function (cdf) $F(x) = \text{Prob}(X \le x)$, or by the probability density function (pdf) $\rho(x)$ for which the probability to be within an interval $[\underline{x}, \overline{x}]$ is equal to the integral $\int_{\underline{x}}^{\overline{x}} \rho(x)\,dx$, and the overall probability is 1: $\int \rho(x)\,dx = 1$.

The relationship between pdf and cdf is straightforward:

- $F(x)$ is the integral of pdf: $F(x) = \int_0^x \rho(t)\,dt$;
- vice versa, the pdf is the derivative of the cdf: $\rho(x) = \dfrac{dF}{dx}$.

What is the cost of buying x_0 computational abilities and doing all other computations in the cloud? We want to select the amount x_0 of computing power to buys, so that everything in excess of x_0 will be sent to the cloud. We want to select this amount so that the expected overall cost of computations is the smallest possible.

So, to find the corresponding value x_0, let us compute how much it will cost the user to buy x_0 equipment and to rent all other computation time. We already know that the cost of buying and maintaining an equipment with capacity x_0 is equal to $c_0 \cdot x_0$.

The expected cost of using the cloud can be obtained by adding the costs multiplied by the corresponding probabilities. We need computations in the cloud when $x > x_0$, For each such value x, we need to rent the amount $x - x_0$ in the cloud. The cost of such renting is $c_1 \cdot (x - x_0)$. The probability of needing exactly x computations is proportional to $\rho(x)$. To be more precise, the probability that we need between x and $x + \Delta x$ computations is equal to $c_1 \cdot (x - x_0) \cdot \rho(x) \cdot \Delta x$. The expected cost of using the cloud is therefore equal to the sum of such products, i.e., to the value $\sum c_1 \cdot (x - x_0) \cdot \rho(x) \cdot \Delta x$. In the limit, when $\Delta x \to 0$, this sum tends to the integral $\int_{x_0} c_1 \cdot (x - x_0) \cdot \rho(x)\, dx$. Thus, the overall cost is equal to the sum of the in-house and in-the-cloud costs:

$$C(x_0) = c_0 \cdot x_0 + c_1 \cdot \int_{x_0} (x - x_0) \cdot \rho(x)\, dx. \tag{1}$$

Let us use this cost expression to find the optimal value x_0. We want to find the value x_0 for which the cost expression (1) attains its smallest possible value. To find this minimizing value, we need to differentiate the expression (1) with respect to x_0 and equate the corresponding derivative to 0.

To make this differentiation easier, let us transform the expression (1) by using integration by parts $\int u\, dv = u \cdot v - \int v\, du$. Here, $\rho(x) = \dfrac{d(F(x) - 1)}{dx}$, so we can take $u = x - x_0$ and $v = F(x) - 1$. The product $uv = (x - x_0) \cdot (F(x) - 1)$ is equal to 0 on both endpoints $x = x_0$ and $x = \infty$, so we get

$$C(x_0) = c_0 \cdot x_0 - c_1 \cdot \int_{x_0} (F(x) - 1)\, dx.$$

Since $F(x) \le 1$, it is convenient to swap the signs and get the expression

$$C(x_0) = c_0 \cdot x_0 + c_1 \cdot \int_{x_0} (1 - F(x))\, dx. \tag{2}$$

The derivative of this sum is equal to the sum of the derivatives. The derivative of the second term can be obtained from the fact that the derivative of the integral is equal to the integrated function. Thus, the equation $\dfrac{dC(x_0)}{dx_0} = 0$ becomes $c_0 - c_1 \cdot (1 - F(x_0)) = 0$, i.e., equivalently,

$$F(x_0) = 1 - \frac{c_0}{c_1}. \tag{3}$$

This formula can be simplified even further if we take into account that for each $\alpha \in [0, 1]$, the value x for which $F(x) = \alpha$ is known as the α-th *quantile*.

For example, for $\alpha = 0.5$, we have the *median*, for $\alpha = 0.25$ and $\alpha = 0.75$, we have *quartiles*, for $\alpha = 0.1, 0.2, \ldots, 0.9$ we have *deciles*, etc.

So, we arrive at the following conclusion.

How many computations to perform in-house: optimal solution. If we know the costs c_0 and c_1 per computation in house and in the cloud, and we also know the probability distribution $F(x)$ describing the user's needs, then the optimal amount x_0 of computational power to buy is determined by the formula (3), i.e., x_0 is a quantile corresponding to $\alpha = 1 - \dfrac{c_0}{c_1}$.

Once we know the optimal value x_0, we can then compute the corresponding cost by using the formula (2).

Discussion. In the extreme case when $c_1 = c_0$, there is no sense to buy anything at all: we can perform all the computations in the cloud. As the cloud costs c_1 increases, the threshold x_0 increases, so when c_1 is very high, it does not make sense to use the cloud at all.

Example. The user's need is usually described by the *power law* distribution, in which, for some threshold t, we have:

- $1 - F(x) = 1$ for $x \le t$ and then
- $1 - F(x) = \left(\dfrac{x}{t}\right)^{-\alpha}$ for some $\alpha > 0$.

Power law is ubiquitous in many financial situations, see, e.g., [1–4, 9, 10, 14–16, 18–21, 23–26].

In this case, the formula (3) takes the form

$$\left(\frac{x_0}{t}\right)^{-\alpha} = \frac{c_0}{c_1}.$$

By raising both sides by the power $-1/\alpha$ and multiplying both sides by the threshold t, we conclude that

$$x_0 = t \cdot \left(\frac{c_0}{c_1}\right)^{-1/\alpha} = t \cdot \left(\frac{c_1}{c_0}\right)^{1/\alpha}. \tag{4}$$

Substituting this expression into the formula (2), we can compute the expected cost. This cost consists of two parts: $c_0 \cdot x_0$ and the integral; we will denote the integral part by I. Let us compute both parts and then add them up. Here,

$$c_0 \cdot x_0 = c_0 \cdot t \cdot \left(\frac{c_1}{c_0}\right)^{1/\alpha} = t \cdot c_0^{1-1/\alpha} \cdot c_1^{1/\alpha}. \tag{5}$$

Since $1 - F(x) = t^\alpha \cdot x^{-\alpha}$, the integral I takes the form

$$I = \int_{x_0} (1 - F(x))\, dx = c_1 \cdot t^\alpha \cdot \int_{x_0}^{\infty} x^{-\alpha}\, dx = c_1 \cdot t^\alpha \cdot \frac{x_0^{1-\alpha}}{\alpha - 1}.$$

Substituting the value (4) into this formula, we get

$$I = c_1 \cdot t^{\alpha} \cdot t^{1-\alpha} \cdot \left(\frac{c_1}{c_0}\right)^{(1-\alpha)/\alpha} \cdot \frac{1}{\alpha - 1},$$

i.e., to

$$I = t \cdot c_0^{1-1/\alpha} \cdot c_1^{1/\alpha} \cdot \frac{1}{\alpha - 1}. \tag{6}$$

By comparing (6) and (4), we can see that $I = c_0 \cdot x_0 \cdot \dfrac{1}{\alpha - 1}$, thus

$$C(x_0) = c_0 \cdot x_0 + I = c_0 \cdot x_0 \cdot \left(1 + \frac{1}{\alpha - 1}\right) = c_0 \cdot x_0 \cdot \frac{\alpha}{\alpha - 1}.$$

Dividing both the numerator and the denominator of this fraction by α, we get the final formula for the cost:

$$C(x_0) = c_0 \cdot x_0 \cdot \frac{1}{1 - \dfrac{1}{\alpha}}. \tag{7}$$

Discussion. The difference between the overall cost (7) and the in-house cost $c_0 \cdot x_0$ is the expected cost of using the cloud.

The larger α, the faster the probabilities of the need for computing power x decrease with x, and thus, the smaller should be the expected cost of using the cloud. And indeed, when α increases, the factor in (7) tends to 1, meaning that the cost of in-the-cloud computations tends to 0.

3 How Much Computations to Perform In-House and How Much in Cloud: Case of Interval Uncertainty

Formulation of the problem. In the previous section, we considered the idealized case when we know the exact costs c_0 and c_1 and the exact probabilities $F(x)$. In practice, we rarely know the exact costs and probabilities. At best, we know the bounds on these quantities, i.e.:

- we know the interval $[\underline{c}_0, \overline{c}_0]$ of possible values of in-house cost c_0;
- we know the interval $[\underline{c}_1, \overline{c}_1]$ of possible values of the in-the cloud cost c_1; and
- for each computation amount x, the know the interval of possible values $[\underline{F}(x), \overline{F}(x)]$ for the cdf $F(x)$; these bounds are also known as a *p-box*; see, e.g., [5–7].

How to select x_0 in case of interval uncertainty: analysis of the problem. For any selection of the value x_0, different values $c_0 \in [\underline{c}_0, \overline{c}_0]$ and

$c_1 \in [\underline{c}_1, \overline{c}_1]$, and for different functions $F(x) \in [\underline{F}(x), \overline{F}(x)]$, the formula (2) leads to different values of the the cost $C(x_0)$.

We do not know the probabilities of different values c_i or different functions $F(x)$, all we know is the bounds. In this case, the only information that we have about the cost $C(x_0)$ corresponding to a selection x_0 is that this cost belongs to the interval $[\underline{C}(x_0), \overline{C}(x_0)]$, where:

- the value $\underline{C}(x_0)$ is the smallest possible value of the cost, and
- $\overline{C}(x_0)$ is the largest possible value of the cost.

In such case of interval uncertainty, natural requirements leads to the following decision making procedure [11–13]:

- we select a parameter $\alpha \in [0, 1]$ that describes the user's degree of optimism-pessimism, and
- we select the alternative x_0 for which the combination $\alpha \cdot \underline{C}(x_0) + (1-\alpha) \cdot \overline{C}(x_0)$ is the smallest possible.

Here:

- the value $\alpha = 1$ (corresponding to full optimism) means that we only consider the best-case (optimistic) scenarios;
- the value $\alpha = 0$ (corresponding to full pessimism) means that we only consider the worst-case (pessimistic) scenarios;
- values α (between 0 and 1 means that we take both best-case and worst-case scenarios into account.

For the formula (2), it is easy to find the smallest and the largest value of $C(x_0)$: from the formula (2), we get

$$\underline{C}(x_0) = \underline{c}_0 \cdot x_0 + \underline{c}_1 \cdot \int_{x_0} (1 - \overline{F}(x))\, dx. \tag{8}$$

and

$$\overline{C}(x_0) = \overline{c}_0 \cdot x_0 + \overline{c}_1 \cdot \int_{x_0} (1 - \underline{F}(x))\, dx. \tag{9}$$

Thus, the above procedure means that we need to optimize the function

$$C_\alpha(x_0) = c_{0,\alpha} \cdot x_0 + c_{1,\alpha} \cdot \int_{x_0} (1 - F_\alpha(x))\, dx, \tag{10}$$

where we denoted

$$c_{0,\alpha} = \alpha \cdot \underline{c}_0 + (1-\alpha) \cdot \overline{c}_0; \tag{11}$$

$$c_{1,\alpha} = \alpha \cdot \underline{c}_1 + (1-\alpha) \cdot \overline{c}_1; \tag{12}$$

$$F_\alpha(x) = \alpha \cdot \overline{F}(x) + (1-\alpha) \cdot \underline{F}(x). \tag{13}$$

Differentiating the expression (10) with respect to x_0 and equating the derivative to 0, we conclude that $c_{0,\alpha} = c_{1,\alpha} \cdot (1 - F_\alpha(x_0))$, i.e., that

$$F_\alpha(x_0) = 1 - \frac{c_{0,\alpha}}{c_{1,\alpha}}. \tag{14}$$

Resulting recommendation. To find the optimal value x_0:

- we should first find the parameter α corresponding to the user's optimism-pessimism level;
- then, we compute the values $c_{0,\alpha} = \alpha \cdot \underline{c}_0 + (1-\alpha) \cdot \overline{c}_0$, $c_{1,\alpha} = \alpha \cdot \underline{c}_1 + (1-\alpha) \cdot \overline{c}_1$, and the function $F_\alpha(x) = \alpha \cdot \overline{F}(x) + (1-\alpha) \cdot \underline{F}(x)$;
- after that, we find the value x_0 for which $F_\alpha(x_0) = 1 - \dfrac{c_{0,\alpha}}{c_{1,\alpha}}$.

Once we find the optimal value x_0, we can use the formulas (8) and (9) to find the range of possible values of costs.

4 Auxiliary Question: When Is It Beneficial to Sign a Multi-Year Contract?

Formulation of the problem. Let us assume that we have an average yearly amount X of computations to perform in the cloud, and we expect the same amount for the few following years. For this year's computations, the cloud company offers us the rate of c_1 per computation; for a T-year contract, the price will be $c_T < c_1$. Shall we sign a contract?

Additional information that we need to make a decision. To decide which is more beneficial, we need to take into account two things:

- first, computers improve year after year, so the computing cost steadily decreases; let $v < 1$ be a yearly decrease in cost; this means that next year, computing in the cloud will cost $v \cdot c_1$ per computation, the year after that $v^2 \cdot c_1$, etc.;
- we also need to take into account that paying a certain amount a next year is less painful that paying the same amount a this year, since we could invest a, get interest, pay a next year, and keep the interest; from this viewpoint, paying a certain amount a next year is equivalent to paying $a \cdot q$ this year, where the discounting parameter $q < 1$ depends on the current interest rate.

Analysis of the problem. In the case of year-by-year payments:

- we pay the amount $c_1 \cdot X$ this year,
- we pay the amount $v \cdot c_1 \cdot X$ next year,
- we pay the amount $v^2 \cdot c_1 \cdot X$ the year after,
- $\ldots$, and
- we pay the amount $v^{T-1} \cdot c_1 \cdot X$ ins the last T-th year.

By using discounting, we find out that:

- paying $v \cdot c_1 \cdot X$ next year is equivalent to paying $q \cdot v \cdot c_1 \cdot X$ this year;
- paying $v^2 \cdot c_1 \cdot X$ in Year 3 is equivalent to paying $q^2 \cdot v^2 \cdot c_1 \cdot X$ this year;

- ..., and
- paying $v^{T-1} \cdot c_1 \cdot X$ in Year T is equivalent to paying $q^{T-1} \cdot v^{T-1} \cdot c_1 \cdot X$ this year.

Thus, year-by-year payments are equivalent to paying the following amount right away:

$$c_1 \cdot X + v \cdot q \cdot c_1 \cdot X + v^2 \cdot q^2 \cdot c_1 \cdot X + \ldots + v^{T-1} \cdot q^{T-1} \cdot c_1 \cdot X =$$

$$c_1 \cdot X \cdot (1 + q \cdot v + q^2 \cdot v^2 + \ldots + q^{T-1} \cdot v^{T-1}).$$

By using the formula for the sum of the geometric progression, we conclude that this cost is equal to

$$c_1 \cdot X \cdot \frac{1 - (q \cdot v)^T}{1 - q \cdot v}.$$

Alternatively, if we sign a contract, then we pay the same mount $c_T \cdot X$ every year. By using discounting, we find out that:

- paying $c_T \cdot X$ next year is equivalent to paying $q \cdot c_T \cdot X$ this year;
- paying $c_T \cdot X$ in Year 3 is equivalent to paying $q^2 \cdot c_T \cdot X$ this year;
- ..., and
- paying $c_T \cdot X$ in Year T is equivalent to paying $q^{T-1} \cdot c_T \cdot X$ this year.

Thus, these payments are equivalent to paying the following amount right away:

$$c_T \cdot X + q \cdot c_T \cdot X + q^2 \cdot c_T \cdot X + \ldots + q^{T-1} \cdot c_T \cdot X =$$

$$c_T \cdot X \cdot (1 + q + q^2 + \ldots + q^{T-1}).$$

By using the formula for the sum of the geometric progression, we conclude that this cost is equal to

$$c_T \cdot X \cdot \frac{1 - q^T}{1 - q}.$$

By comparing these two numbers, and dividing both sides of the resulting inequality by the common factor X, we arrive at the following conclusion.

When it is beneficial to sign a multi-year contract: recommendation. It is beneficial to sign a multi-year contract if

$$c_T \cdot \frac{1 - q^T}{1 - q} \leq c_1 \cdot \frac{1 - (q \cdot v)^T}{1 - q \cdot v}.$$

Acknowledgments

This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721.

References

1. J. Beirlant, Y. Goegevuer, J. Teugels, and J. Segers, *Statistics of Extremes: Theory and Applications*, Wiley, Chichester, 2004.
2. B. K. Chakrabarti, A. Chakraborti, and A. Chatterjee, *Econophysics and Sociophysics: Trends and Perspectives*, Wiley-VCH, Berlin, 2006.
3. A. Chatterjee, S. Yarlagadda, and B. K. Chakrabarti, *Econophysics of Wealth Distributions*, Springer-Verlag Italia, Milan, 2005.
4. J. D. Farmer and T. Lux (eds.). *Applications of statistical physics in economics and finance*, a special issue of the *Journal of Economic Dynamics and Control*, 2008, Vol. 32, No. 1, pp. 1–320.
5. S. Ferson, *Risk Assessment with Uncertainty Numbers: RiskCalc*, CRC Press, Boca Raton, Florida, 2002.
6. S. Ferson, V. Kreinovich, L. Ginzburg, D. S. Myers, and K. Sentz, *Constructing Probability Boxes and DempsterShafer Structures*, Sandia National Laboratories, Albuquerque, New Mexico, Report SAND2002-4015, 2013.
7. S. Ferson, V. Kreinovich, W. Oberkampf, and L. Ginzburg, *Experimental Uncertainty Estimation and Statistics for Data Having Interval Uncertainty*, Sandia National Laboratories, Report SAND2007-0939, 2007.
8. B. Furht and A. Escalanate, *Handbook of Cloud Computing*, Springer, New York, 2010.
9. X. Gabaix, G. Parameswaran, P. Vasiliki, and H. E. Stanley, "Understanding the cubic and half-cubic laws of financial fluctuations", *Physica A*, 2003, Vol. 324, pp. 1–5.
10. C. P. Gomez and D. B. Shmoys, "Approximations and Randomization to Boost CSP Techniques", *Annals of Operations Research*, 2004, Vol. 130, pp. 117–141.
11. L. Hurwicz, *Optimality criteria for decision making under ignorance*, Cowles Commission Discussion Paper, Statistics, No. 370, 1951.
12. V. Kreinovich, "Decision making under interval uncertainty (and beyond)", In: P. Guo and W. Pedrycz (eds.), *Human-Centric Decision-Making Models for Social Sciences*, Springer Verlag, 2014, pp. 163–193.
13. R. D. Luce and R. Raiffa, *Games and Decisions: Introduction and Critical Survey*, Dover, New York, 1989.
14. B. Mandelbrot, *The Fractal Geometry of Nature*, Freeman, San Francisco, California, 1983.
15. B. Mandelbrot and R. L. Hudson, *The (Mis)behavior of Markets: A Fractal View of Financial Turbulence*, Basic Books, 2006.
16. R. N. Mantegna and H. E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance*, Cambridge University Press, Cambridge, Massachusetts, 1999.
17. D. Marinescu, *Cloud Computing: Theory and Practice*, Morgan Kaufmann, Waltham, Massachusetts, 2013.
18. N. Markovich (ed.), *Nonparametric Analysis of Univariate Heavy-Tailed Data: Research and Practice*, Wiley, Chichester, 2007.
19. J. McCauley, *Dynamics of Markets, Econophysics and Finance*, Cambridge University Press, Cambridge, Massachusetts, 2004.
20. S. T. Rachev and S. Mittnik, *Stable Paretian Models in Finance*, Wiley Publishers, New York, 2000.
21. S. I. Resnick, *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Springer-Varlag, New York, 2007.

22. J. Rhoton, *Cloud Computing Explained*, Recursive Press, 2010.
23. B. Roehner, *Patterns of Speculation - A Study in Observational Econophysics*, Cambridge University Press, Cambridge, Massachusetts, 2002.
24. H. E. Stanley, L. A. N. Amaral, P. Gopikrishnan, and V. Plerou, "Scale invariance and universality of economic fluctuations", *Physica A*, 2000, Vol. 283, pp. 31–41.
25. S. V. Stoyanov, B. Racheva-Iotova, S. T. Rachev, and F. J. Fabozzi, "Stochastic models for risk estimation in volatile markets: a survey", *Annals of Operations Research*, 2010, Vol. 176, pp. 293–309.
26. P. Vasiliki and H. E. Stanley, "Stock return distributions: tests of scaling and universality from three distinct stock markets", *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 2008, Vol. 77, No. 3, Pt. 2, Publ. 037101.
27. A. T. Velte, T. J. Velte, and R. Elsenpeter, *Cloud Computing: A Practical Approach*, McGraw-Hill, New York, 2010.
28. A. Whittmann, "Does the cloud keep pace with Moore's law?", *Information Week*, 2012, No. 1327 (March 12, 2012), p. 42.