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

1. Cloud Computing: Official Definition by the US National Institute of Standards and Technology (NIST)

Cloud computing is a model for

- enabling ubiquitous, convenient, on-demand network access
- to a shared pool of configurable computing resources, such as:
  - networks,
  - servers,
  - storage,
  - applications, and
  - services
- that can be rapidly provisioned and released with minimal management effort or service provider interaction.

2. Financial Aspect of Cloud Computing

- One of the important aspects of cloud computing 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.
- 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 option.
- However, if a peak need for computing occurs rarely:
  - then 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.

3. Resulting Questions

- Once the user knows its computational requirements, the first 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?

4. Why This Is Important and What We Propose

- One of the main purposes of cloud computing is to save user's money.
- However, most cloud users are computer folks with little knowledge of economics.
- As a result, often, they make wrong financial decisions about the cloud use.
- It is therefore important to come up with proper recommendations for using cloud computing.
- In this talk, we describe 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.

5. Case of Complete Information

- Let c_0 be the overall cost of buying and maintaining one unit (e.g., Teraflops).
- Then, if we buy computers with computational ability x_0, we pay $c_0 \cdot x_0$ for these computers.
- Let c_1 be a per-unit cost of computing in the cloud.
- Then, if we need to perform x computations in the cloud, we have to pay the amount $c_1 \cdot x$.
- Complete knowledge means that for each possible daily computation need x:
  - we know the probability $p(x)$ that we need x computations;
  - this probability $p(x)$ can be estimated by analyzing the previous needs;
  - for example, if we needed x computations in 10% of the days, this means that $p(x) = 0.1$.

6. Resulting Formula for the Cost

- We want to select the amount x_0 of computing power to buy.
- Then, 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.
- Th in-house cost is $c_0 \cdot x_0$.
- For each value $x > x_0$, the cost is $c_1 \cdot (x - x_0)$, the probability is $p(x) \approx \rho(x) \cdot \Delta x$.
- Thus, the overall cost is

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

7. Optimization: General Case

- We want to minimize the overall cost
$$C(x_0) = c_0 \cdot x_0 + c_1 \cdot \int_{x_0}^{\infty} (x - x_0) \cdot \rho(x)\, dx.$$
- Differentiating this expression w.r.t. x_0 and equating derivative to 0, we get $F(x_0) = 1 - \dfrac{c_0}{c_1}$.
- So, the optimal amount x_0 of computational power to buy is a quantile corresponding to $p = 1 - \dfrac{c_0}{c_1}$.
- 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.

8. Optimization: Example

- The user's need is usually described by the *power law* distribution: $F(x)=1-\left(\dfrac{x}{t}\right)^{-\alpha}$ for all $x\ge t$.
- In this case, $x_0=t\cdot\left(\dfrac{c_1}{c_0}\right)^{1/\alpha}$, and the resulting cost is $C(x_0)=c_0\cdot x_0\cdot\dfrac{1}{1-\dfrac{1}{\alpha}}$.
- The difference between the overall cost 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.
- Thus, the smaller should be the expected cost of using the cloud.
- When α increases, indeed $C(x_0)-c_0\cdot x_0\to 0$.

9. Case of Interval Uncertainty: Problem

- In practice, we rarely know the exact costs and probabilities.
- At best, we know the bounds on these quantities.
- So, we know:
  - the interval $[\underline{c}_0, \overline{c}_0]$ of possible values of c_0;
  - the interval $[\underline{c}_1, \overline{c}_1]$ of possible values of c_1, and
  - the interval $[\underline{F}(x), \overline{F}(x)]$ of possible values of $F(x)$ (a *p-box*).
- In this case, we only know that the cost $C(x_0)$ is between $\underline{C}(x_0)$ and $\overline{C}(x_0)$, where:

$$\underline{C}(x_0) = \underline{c}_0 \cdot x_0 + \underline{c}_1 \cdot \int_{x_0}^{\infty} (1 - \overline{F}(x))\, dx;$$

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

10. Case of Interval Uncertainty: Solution

- Natural requirements to decision making under interval uncertainty imply that we minimize

$$\alpha_H \cdot \underline{C}(x_0) + (1 - \alpha_H) \cdot \overline{C}(x_0).$$

- Here, α_H is Hurwicz's optimism-pessimism parameter:
  - $\alpha_H = 1$ corresponds to full optimism;
  - $\alpha_H = 0$ corresponds to full pessimism;
  - values $\alpha_H \in (0, 1)$ mean that we take both best-case and worst-case scenarios into account.
- The resulting optimal x_0 is a p-th quantile of

$$F_H(x) = \alpha_H \cdot \overline{F}(x) + (1 - \alpha_H) \cdot \underline{F}(x), \text{ where}$$

$$p = 1 - \frac{\alpha_H \cdot \underline{c}_0 + (1 - \alpha_H) \cdot \overline{c}_0}{\alpha_H \cdot \underline{c}_1 + (1 - \alpha_H) \cdot \overline{c}_1}.$$

11. When Is It Beneficial to Sign a Multi-Year Contract: Problem

- Let X denote the average yearly amount of computations to perform in the cloud.
- For a T-year contract, the price is $c_T < c_1$; shall we sign a contract?
- Computers improve year after year.
- So, the computing cost steadily decreases.
- Let $v < 1$ be a yearly decrease in cost.
- So, next year, computing in the cloud will cost $v \cdot c_1$ per computation, then $v^2 \cdot c_1$, etc.
- Payment delay is beneficial, since we can invest the money with interest.
- Thus, paying a next year is equivalent to paying $a \cdot q$ now, for some $q < 1$.

12. When Is It Beneficial to Sign a Multi-Year Contract: Solution

- If we pay year-by-year, we pay $v^{t-1} \cdot c_1 \cdot X$ in year t.
- This is equivalent to paying now the following amount:

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

- If we sign a contract, we pay $c_T \cdot X$ every year.
- This is equivalent to paying now the following amount:

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

- So, a multi-year contract is beneficial if

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

13. 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.