Resource allocation in a cloud under virus attacks

Eliran Sherzer, Gail Gilboa-Freedman and Hanoch Levy

1 Introduction

How should you place your resources on the cloud sustain potential failures, due to technical problems or malicious attacks? The question should be addressed within the domain of resource allocation strategies in cloud computing, which have been widely studied over the past years (see survey in [3]). In prior studies, the resource allocation problem was addressed under a framework where the servers are 100 percent reliable (see for example, [1], [2]). That is, if a server is allocated then it operates with probability 1.

As suggested above, however, system designers must account for server failures that may be caused by various reasons, such as viruses or breakdowns. In 2015 alone, 230,000 new malwares were launched every day. As a consequence, cyber security and network resilience are a major concern for internet service providers.

The maliciously caused failures and the occasional breakdowns lead us to divert from the traditional models of deterministic servers and propose a new model where resource allocation is conducted under the assumption whereby the number of operating servers is **stochastic**. Our model distinguishes between two values regarding the servers. The first is the number of allocated servers, which is the number of servers placed by the system operator; this number is deterministic. The second is the number of surviving servers after a possible failure. This is a random value since the outcome of a failure is uncertain. We refer to the number of the placed servers as *placement*, whereby these are our decision variables. The number of remaining servers is a random variable and it is referred to as the *supply* variable.

Service providers may use different types of servers within the same networks. For example, separate servers for Linux and Windows or separate servers for production and development. More common types of servers are: Web servers, Email servers, FTP servers and identity servers. Many networks on the internet employ a client-server networking model integrating websites and communication services. Usually, it comes with a limited types of servers. An alternative model peer-to-peer (P2P) networking allows all devices on a network to function as either a server or client as needed. Those networks are characterized with a multiple types of servers. For example, video streaming applications, where each movie is a different server type.

In this study, we focus on a multi-type system that the number of resources are limited. There is a random demand for each type, and the goal is to reach the maximum expected revenue of satisfied demand across all regions. We provide the model details in Section 2. The main challenge is to deal with a decision variables that are inherent in a stochastic factor.

We show that under a condition on the supply distribution, the objective function is concave. Thus one can use a greedy algorithm to solve the problem. Moreover, we show that this condition is both sufficient and necessary in order for the concavity property to hold (for any given demand).

What turns out to play a major role in our analysis is some kind of a cumulative function of the supply cdf, which we call *cumcum*. The cucmcum function of a random variable is defined as follows: Let X be a discrete random variable, with a support $\{0, 1, 2, ..., w\}$ and let $F_X(\cdot)$ be it's cdf. Then, the cucmum function of the random variable X at k is $\sum_{j=0}^{k} F_X(j)$ for $k \leq w$, and $\sum_{j=0}^{w} F_X(j)$ for $k > w$. We denote the cumcum function of X as $\not{F}_X(\cdot)$. We later show that the condition for the desired concavity property can be defined through the cumcum function.

2 Model

We consider a single-region system with multiple server types. The different types are numbered $1, 2, ..., t$. We refer to the set of servers placed for all types as a *placement*. We assume that there is a capacity of total servers m. Placement L is called feasible if the number of servers is not larger than m. We denote the number of requests for type i servers by D_i. In this model formulation, no assumptions are made regarding the distribution of D_i, nor, we assume independence between the demands.

Let L_i be the number of placed servers from type i. Let $S(L_i)$ be the stochastic supply of type i (i.e. the number of type i surviving servers after a possible failure). Whereas, the supply $S(L_i)$ is conditioned by the placement L_i. We note that the support of the supply distribution $S(L_i)$ is $\{0, 1, 2, ..., L_i\}$. The demand and supply cdf values are computed in $O(1)$ and

VALUETOOLS 2017, December 5–7, 2017, Venice, Italy

ACM ISBN 978-1-4503-6346-4/17/12…$15.00
https://doi.org/10.1145/3150928.3150969

the statistics are computed by an external data base.

2.1 Objective function We assume that there is a linear reward from matching a server to a demand. The linear reward parameter from a single match of type i is denoted by R_i. The number of rewards with respect to type i is $min\{S(L_i), d_i\}$, where $S(L_i)$ and d_i are the type i realizations of the supply and demand, respectively. Since we can't do optimization on realizations of random variables, we focus on the expected value of the total reward. Thus, the number of expected matching for type i is $\mathbb{E}[min\{S(L_i), D_i\}]$. Finally, we can construct the objective function:

$$\sum_{i=1}^{t} R_i \mathbb{E}[min\{S(L_i), D_i\}] \tag{2.1}$$

The main goal is to find a placement L such that the expression in (2.1) is maximized.

3 Proving concavity of the objective function for the use of a greedy algorithm

The objective function in (2.1) possess t separable addends. Whereas, each one is the expected value $\mathbb{E}[min\{S(L_i), D_i\}]$ for $i \in \{1, ..., t\}$. For concavity, each addend is required to be concave with it's corresponding placement variable $\{L_i\}$. Therefore, we focus on the expression $\mathbb{E}[min\{S(L_i), D_i\}]$. We show that the concavity property holds only for some of the supply distributions $S(\cdot)$. Moreover, we show that whether or not the concavity holds, depends on a very of simple condition regarding the cumcum function of the supply distribution. We prove that this condition is both sufficient and necessary. Our results holds for any demand distribution, whereas we do not make any assumptions on the demand. For our analysis, we use the following notations: Let $F_{S(L_i)}(\cdot)$ be the cdf of the *supply* given a placement of L_i servers. Finally, let $\Delta(L_i + 1, L_i) = \mathbb{E}[min\{S(L_i + 1), D\}] - \mathbb{E}[min\{S(L_i), D\}]$.

LEMMA 3.1. *A sufficient and necessary condition for any supply distribution $S(L_i)$, such that $\mathbb{E}[min\{S(L_i), D_i\}]$ is concave with L_i (for any D_i) is the following:*

$$2F_{S(L_i+1)}(k) - F_{S(L_i)}(k) - F_{S(L_i+2)}(k) \leq 0 \; \forall k. \tag{3.2}$$

Proof. Our proof is twofold. We first show that this statement is true for any realization d_i of D_i. Then, if it is true for any realization, then it's true for D_i as well. This is because, $\mathbb{E}[min\{S(L_i), D_i\}]$ is basically a weight sum of all realizations. Of course, a sum of concave functions is a concave function as well. We state that from basic probability, for any positive d_i: $\mathbb{E}[min\{S(L_i), d_i\}] = \sum_{j=1}^{d_i} \mathbb{P}(S(L_i) \geq j)$. From the definition of concavity we need to show that: $\Delta(L_i + 1, L_i) \geq \Delta(L_i + 2, L_i + 1)$. After some simple algebra we get:

$$2\sum_{j=0}^{d_i} \left(\mathbb{P}(S(L_i + 1) \leq j) - \mathbb{P}(S(L_i + 2) \leq j) - \mathbb{P}(S(L_i) \leq j)\right). \tag{3.3}$$

If it is true for all possible $d_i > 0$ then clearly it holds for any possible distribution of D_i. Finally, the expression in (3.3) is equivalent to the statement in the Lemma (where k is replaced by d_i).

THEOREM 3.1. *If the supply distribution $S_i(\cdot)$ meets the condition from Lemma 3.1, then the placement problem can be solved by a greedy algorithm (such as steepest descent).*

Proof. This is a direct result from the concavity property.

4 Conclusions

In this study we show that under a single condition on the supply distribution one can show that the objective function is concave, and hence, it is possible to allocate optimally via a greedy algorithm. From initial analysis, we know that the following distribution answer the condition: Binomial, Uniform, truncated Poisson and truncated Geometric. We also note, that it can easily be shown that by applying a steepest ascent algorithm, the problem can be solved in $O(m^2 t)$. If however, the condition doesn't hold, the problem can still be solved in polynomial time (yet, more expensive than the greedy algorithm) by dynamic programming. In conclusion, in this study, we introduced the concept of viruses attacks in a resource allocation model. Although, it is a very basic model, it is a solid foundation for a distributed geographic systems which are much more complicated.

References

[1] Y. Rochman, H. Levy, and E. Brosh. "Max percentile replication for optimal performance in multi-regional P2P VoD systems." Quantitative Evaluation of Systems (QEST), 2012 Ninth International Conference on. IEEE, (2012).

[2] Y. Rochman, H. Levy, and E. Brosh. "Resource placement and assignment in distributed network topologies". In IEEE INFOCOM (Turin, Italy, April 2013).

[3] V. Vinothina, R. Sridaran and P. Ganapathi. A "Survey on Resource Allocation Strategies in Cloud Computing". In *IJACSA* (2012) 3:97-102.