American Journal of Information Systems, 2018, Vol. 6, No. 1, 23-28
Available online at http://pubs.sciepub.com/ajis/6/1/4

DOI:10.12691/ajis-6-1-4

Optimal Allocation of QoS and Web Services in Cloud Computing

Jimbo Claver[1,*], Edris Hamraz[2], Jawad Azimi[3], Charles Owona[4]

[1]Department of Science and ICT, American University of Afghanistan & Waseda University, Tokyo, Japan (Joint Research Work)
[2]Department of ICT, American University of Afghanistan, Kabul, Afghanistan
[3]Japan International Cooperation Agency (JICA), Headquarter, Kabul, Afghanistan
[4]Department of Applied Mathematics, University of Yaounde, Cameron
*Corresponding author: jimbo.maths@gmail.com

Abstract Cloud computing is a great model of demand and supply in information communication and services. It represents a complex infrastructure and provides a dynamic, distributed, heterogeneous and autonomous platform for solving problems in business, science and technology. This paper proposes a cloud computing environment that supports dynamic application service composition model. We develop a Quality of Service (QoS) based framework for effective web services allocation. In computing, the service consumer is projected to provide the QoS requirements as part of service discovery query. The cloud as marketplace for trading instances of web services can be bought or leased by web applications. We found that using dynamic decision-making management approach relaying on dynamic portfolio allocation model mainly used in finance, one can achieve the purpose of saving resources by reducing costs of quality of services and eliminating risks related to different services simultaneously.

Keywords: *Cloud computing, Quality of Service (QoS), web services, information systems, workflow, data mining, modeling, simulation, dynamic allocation, optimization, decision management, risk, computing and applications*

Cite This Article: Jimbo Claver, Edris Hamraz, Jawad Azimi, and Charles Owona, "Optimal Allocation of QoS and Web Services in Cloud Computing." *American Journal of Information Systems*, vol. 6, no. 1 (2018): 23-28. doi: 10.12691/ajis-6-1-4.

1. Introduction

Cloud computing viewed as an upcoming model is a convenient communication and symbol of internet that represents a complex infrastructure integrating configurable computing resources, namely hardware, software, processing power applications, storage as a service among many computers [1,2]. With the development of new internet technologies available, the number of web services is increasing depending on the needs of users in different situations as well as the responses to different needs and business process [3,4]. The web service can be defined as a technology that offers software services or pathways that will trigger the shift in the ways distributed systems are created. This architecture works, in general, with two main entities: (1) the service provider and (2) the service consumer. The Quality of Service (QoS) and the web service are in most cases built to satisfy the customers' needs; however, the functional and non-functional characteristics of web services and its composition have raised great concerns in academia and industry.

In [5] describe the service oriented cloud computing system platform that enables web delivery of application based services with a set of common business domain and operational service. Some companies in [6,7] raise the issue with single user cloud provider, which has limited resources for use and lack interoperability among cloud providers. In [8] a reservoir, architecture and computational resources were portioned by a visualization layer into a virtual execution environment used for clouds. However, the existing models are mainly agent based computing of QoS parameters and service request monitors can fulfill the QoS requirements.

This paper proposes a cloud computing environment with optimal and dynamic allocation of service composition with ultimate goal to satisfy the best dynamic workflow environment, such as the cloud computing web service that has optimal and dynamic workflow that leads to a novel management system. In this work, we propose a new approach based on multi-objective portfolio selection in optimal service management in cloud computing.

The paper is organized as follows: In section 2, we develop the cloud computing architecture for web services; in section 3, we present the dynamic resource decision allocation in cloud computing system; in section 4, we discuss the web service portfolio allocation model; in section5, we present the multi-objective techniques in practice. In section 6, we highlight some steps toward the solution, and finally, we end this work with a short conclusion in section 7.

2. Cloud Computing Architecture

2.1. Cloud Computing Architecture

Cloud computing architecture of a cloud solution is a system, which comprises on premises in cloud resources,

middleware, services, and components of software, geolocation, and the relationships between them. The term also refers to documentation of a cloud architecture systems that facilitates communication between stakeholders, documentation of high-level decisions, and allows reuse of design components and patterns between projects.

When talking about a cloud computing system [for example in [9,10]], it's helpful to divide it into two sections: the front end and the back end, which are connected to each other through a network, usually the Internet. The front end is the side that a computer user or a client sees, and the back end is the "cloud" section of the system. The front end includes the client's computer (or computer network) and the application required to access the cloud computing system. It is important to mention that not all cloud computing systems have the same user interface and services like Web-based e-mail programs leverage existing Web browsers like Internet Explorer or Firefox; other systems have unique applications that provide network access to clients. On the back end of the system are the various computers, servers and data storage systems that create the "cloud" of computing services. In theory, a cloud computing system could include practically any computer program you can imagine, from data processing to video games. Usually, each application will have its own dedicated server. A central server administers the system, monitor traffic and client demands to ensure everything runs smoothly. It follows a set of rules called protocols and uses a special kind of software called middleware that allows networked computers to communicate with each other. Most of the time, servers don't run at full capacity which means that there's unused processing power going to be wasted. It's possible to fool a physical server into thinking it's actually multiple servers, each running with its own independent operating system. The technique is called server virtualization. By maximizing the output of individual servers, server virtualization reduces the need for more physical machines. If a cloud computing company, for example, has a lot of clients, there's likely to be a high demand for a lot of storage space. Some companies require hundreds of digital storage devices. Cloud computing systems need at least twice the number of storage devices it requires to keep all its clients' information stored. That's because these devices, like all computers, occasionally break down. A cloud computing system must make a copy of all its clients' information and store it on other devices. The copies enable the central server to access backup machines to retrieve data that otherwise would be unreachable. Making copies of data as a backup is called redundancy.

2.2. Cloud Computing for Web Services

As cloud computing is relatively new innovation to the world of technology, there are various open issues, which need to be resolved before cloud computing is fully accepted by the broad community. Before we will dive into the research methodology of this thesis, a deeper explanation is needed about cloud computing. We already had a discussion regarding the definition of cloud computing and now recall the definition. “Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, services) that can be rapidly provisioned and released with minimal management effort or service provider interaction” [11]. The above definition is supported by five key cloud characteristics; three delivery models and four deployment models. These supporting properties will be explained below, after which we will discuss various security issues and concerns related to cloud computing [11].

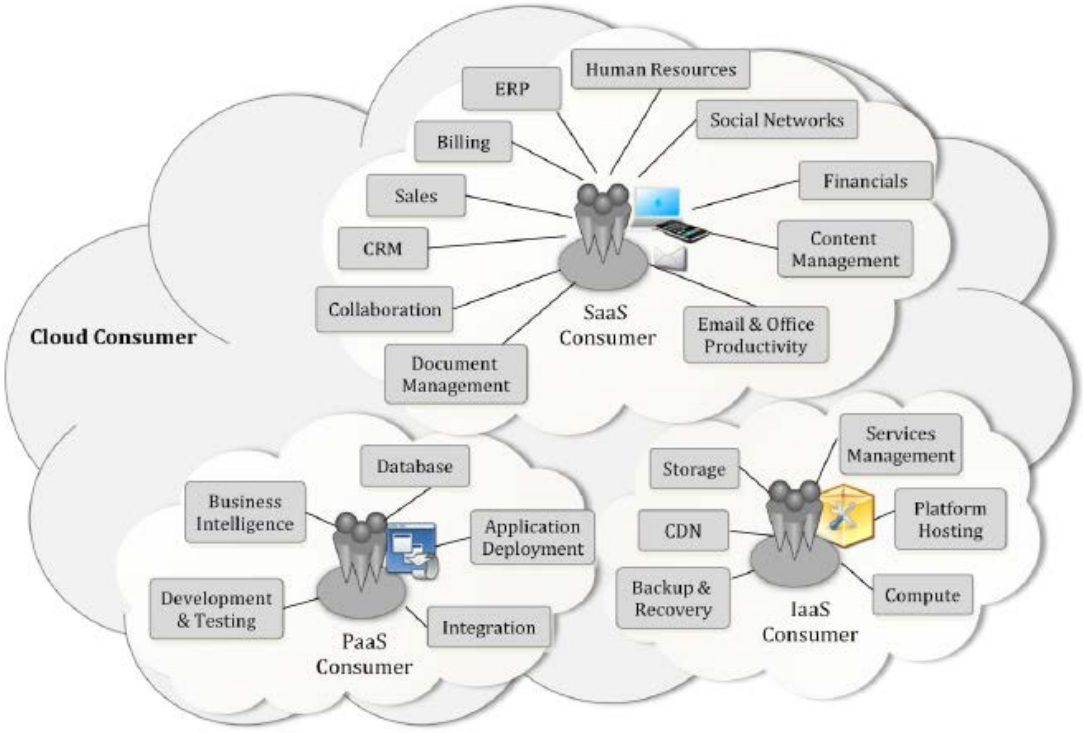

Figure 1. Architecture design of cloud computing systems: This figure shows the architecture of cloud computing, its infrastructure, applications and information about the software platform. (Source: NIST SP 500-292)

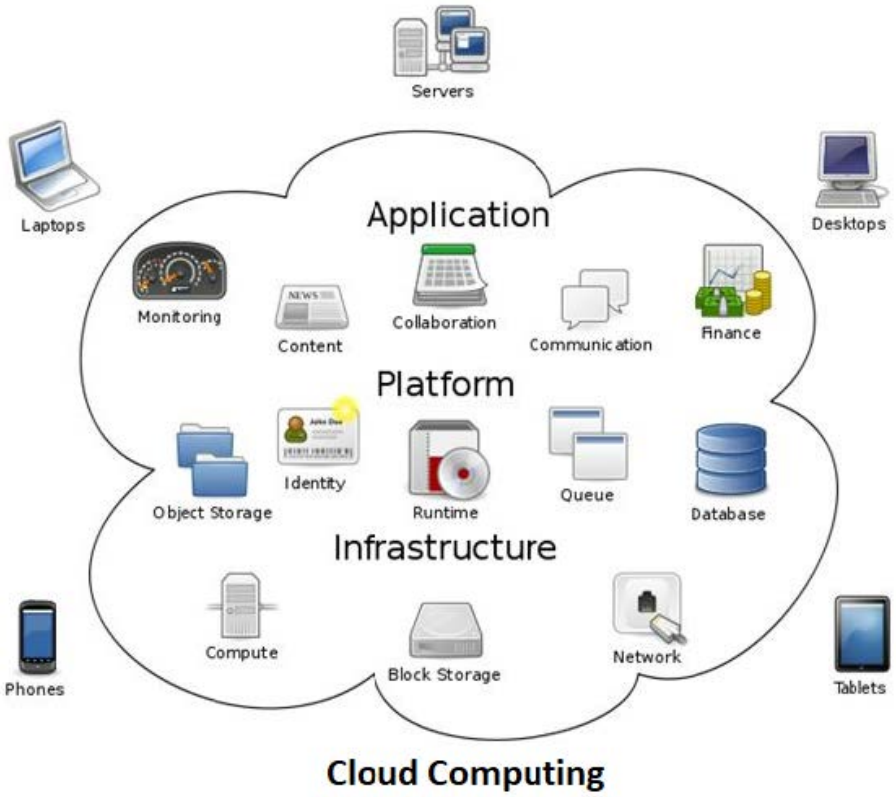

Figure 2. Cloud computing management system: This figure shows how cloud computing works with the peripheral devices connected to it. In the cloud computing system, we have application, platform, and infrastructure that communicate with the clients. (Source: Wikipedia.org./Cloud computing)

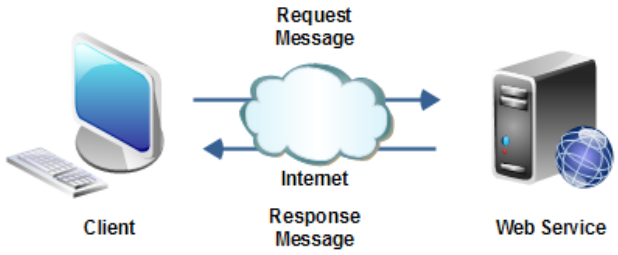

Figure 3. Cloud computing input-output system: This figure explains the communication structure between clients and web services. The client and web services are connected through the Internet. Whenever a request message is sent, the clients receive a response via Internet from web services. (Source: Quora.com)

The architecture of Cloud Computing represent a top-level architecture of cloud that depicts various cloud service delivery models. Cloud enhanced collaboration, agility, scalability and availability provide the potential for cost reduction through optimized and efficient computing. From an architectural perspective, given this abstracted evolution of technology, there is much confusion surrounding how the cloud is both similar and different from existing models and how these similarities and differences might impact the organizational, operational and technological approaches to cloud adoption as it relates to traditional network and information security practices [9,12,13].

Generally speaking, the web service composition provides a good idea of the cloud computing environment. We have a real time monitoring system quality and service quality information and resources [8,12,14]. In order to achieve customer expectation objectives, the system is managed automatically as to enable resource allocation with less cost and resource efficiency with module on the functional prospective, which is an internet facing self-description, self-contained basic building blocks of distributed computing with cross platform.

3. Dynamic Resource Decision Allocation

3.1. Resource Allocation

The process of assigning available resources to the needed cloud application is called resource allocation [15,16,17]. Cloud resources can be provisioned, fine-grained on demand and multiplexed manner. Cloud resources are requested by the cloud user when the application needs. Here, underutilization and overutilization of resources is avoided as much as possible, but the requested resources might not be available when a request is needed. The service provider has to make an allocation from other participating cloud data center. Resource Allocation Starves (RAS) service if the allocation is not managed precisely; on the other hand, resource provisioning solves the problem by allowing the cloud provider to manage the resources for each individual module by integrating cloud provider activities with allocation scarce resources within the limit of cloud environment to meet the needs of cloud applications called resource allocation strategy. For completing user jobs it requires to the type and amount of resource needed by each application. For optimal RAS, the order and time for allocation of resources is an input that should avoid the following criteria:

- Resource contention: it happens when two applications try to access the same resource at the same time.
- Resource fragmentation: it refers to a situation when there would be enough resource but cannot be allocated due to fragmentation into small entities for the needed application. The more resource fragmentation is raised; the better resources are isolated.
- Under provisioning: it occurs when the application is assigned with fewer number of resources than are demanded.

4. Web Service Portfolio Allocation

4.1. Concept of Modelling

Let us consider a web application that needs to allocate multiples instances of web services from a cloud based market. A cloud market is a place where are facilitated buying and selling instances of web services, which are offered with different prices and QoS. In [8,14], the auction–based approach was used to allocate all instances of web services from single to multiple providers that have the lowest price and optimal QoS. In this work, we use the same approach as in [15,16].

4.2. QoS Dynamics

The initial model presents the QoS (q_i) and various services S_i ; $i=1:n$ with constraints on weights as follows

$$\begin{cases} q_n = w_1 S_1 + w_2 S_2 + ... + w_n S_n \\ w_1 + w_2 + ... + w_n = 1 \end{cases} \quad (1)$$

The expected return of the web service portfolio E_i that is built by allocation of instances of web services from n providers is

$$E_p = \sum_1^n w_i \frac{q_i}{C_i} \quad (2)$$

and

$$\sum_1^n w_i = 1 \quad (3)$$

Where w_i represents the weight of the web services that is allocated from the service provider S_i and C_i the cost of the web services.

The risk of the Service Level Agreement (SLA) σ_i is quantified as the percentage between the numbers of SLAs that have been violated to total SLA delivered by the product provider S_i . The global risk of portfolio σ_p is calculated through the local risk σ_i associated to the web services S_i . We have

$$\sigma_p = \sqrt{\sum_1^n w_i^2 \sigma_i^2} \quad . \quad (4)$$

To optimize the global risk of the portfolio R_p and find the optimal solution, we must know how much weight w_i should be invested in each web service S_i to minimize the portfolio risk (construct a low risk portfolio). This is a portfolio optimization problem that is found in mathematical finance; it was first introduced in (Markowitz, 1954) and recently [15,16] have developed a new approach to portfolio allocation using Genetic Algorithm anfd stochastic constraints in some case. They have found that additional constraints might facilitate the finding of the optimal solution in a given search region. Several other techniques have, so far, been used to solve the portfolio optimization problem. In [8,9] generalized reduced gradient was used; the fitness function was introduced in order to solve the constrained optimization problem. [14] used the principle of the divided and conquer technique, which consists of decomposing the search tree in set of sub trees, assigning each sub-tree to a computing core. The issue with such an approach is that it cannot assure good local balancing between all services. [8,12] proposed series of innovative approaches to tackle the problem. Genetic algorithm was introduced and optimal solution was found using efficient computation. However, the search space of services and best terminal generation time remained unsolved. We recall that our ultimate goal in this work is to find the optimal percentage weight associated to each web service instance as to obtain the best allocation of all possible portfolios. We will use the same technique as in [15,16].

5. Multi - Objective Techniques

The main aim in this study is the cost efficiency and risk hedging of service allocations in cloud computing. In practice, we have to collect the services that are cost efficient and can reduce the risk. Consider portfolio consisting of k services below:

$S(i);\ i=1,.....,k$. Let $\sigma(i,j);\ j=1,.....,k$ be the covariance matrix of these services, then there is a set of services such that $0 \le w(i) \le 1$, $\sum_1^k w(i) = 1$ and

$$E(S(k)) = \sum_1^k S(i) w(i) \quad (5)$$

$$\sigma(k) = \sum_1^k \sigma(i,j) w(i) w(j) \quad (6)$$

where $E(S(k))$ is the expected return and $\sigma(k)$ the risk associate to k services. Thus, the portfolio selection problem is a multiobjective optimization problem.

5.1. Multi-Objective Optimization Approach

Portfolio optimization in dynamic asset allocation in cloud computing associated to Web service instances is a multi-objective problem.

A. Naïve Portfolio Optimization

In the Markowitz, portfolio is a set of securities chosen to maximize the expected return and minimize the expected risk. Typically risk is measured by the variances to obtain the best allocation. Assume $Z(i);\ \ i=1,.....,k$ the objective function on web services, we set $Z_1=-Z(i);\ \ i=1,.....,k$ the problem becomes

$$\begin{cases} \min Z_1 \\ s.t.\ \ Z(i)\le\alpha(i);\ \ i=1,.....,k \end{cases} \tag{7}$$

Where $Z_1=\sum_1^k Z(i)w(i);\ 0\le w(i)\le 1;\sum_1^k w(i)=1$; $\alpha(i)$ are parameters to be gradually decreased till no solution is found. The problem with this method is the choice of the thresholds $\alpha(i)$; in the case of equality, this method is guaranteed to give a Pareto optimal solution.

B- Fuzzy logic portfolio optimization

Another approach is to use the Fuzzy logic to study each objective function individually and find its minimum and maximum respectively.

$$\begin{cases} \min Z(i) \\ \max Z(i) \end{cases}. \tag{8}$$

Then determine

$$m(i)=\frac{\left(\min Z(i)-Z(i)\right)}{\left(\max Z(i)-\min Z(i)\right)}. \tag{9}$$

Thus, $0\le m(i)\le 1$.

Applying $\max\left\{\min\left\{m(i)\right\}, i=1,....,k\right\}$, this method is guaranteed to give a Pareto optimal solution although it is a bit difficult to apply for large number of objective functions.

6. Steps Toward the Solution

We begin by modifying the first objective so that all objectives functions are minimizing function, that is $Z'(1)=c-\sum_1^k w(i)S(i)$ where c is a real number.

We have two objectives:

(i) Minimize $Z'(1)$

(ii) Minimize $\sigma(k)$.

Using the Lagrange approach, the necessary condition for optimality is achieved.

$$L(\lambda)=\left\{c-\sum_1^k w(i)S(i)\right\}\left\{\sum_{i,j=1}^k w(i)w(j)\sigma(i,j)\right\}+ \\ +\lambda\left\{c-\sum_1^k w(i)S(i)\right\} \tag{10}$$

Set $\frac{\partial L(\lambda)}{\partial\lambda}=0$ where λ is the Lagrange multiplier.

Eliminating the Lagrange multiplier we obtain

$$2c=3\sum_1^k w(i)S(i)-E\left\{S(k)\right\}+ \\ +\frac{2\left\{\sum_1^k w(i)\sigma(i)\right\}\left\{c-\sum_1^k w(i)S(i)\right\}}{\sum_1^k w(j)\sigma(i,j)} \tag{11}$$

where $p=1,.....,k$.

Theorem 1. The optimal portfolio exists *if and only if* the optimal weights $\left\{w^*(1),w^*(2),.....,w^*(k)\right\}$ exist.

Theorem 2. The optimal portfolio exists *if and only if* the condition (7) is fulfilled.

Proof.

The proof of theorems 1 and 2 follows from the sub-section above.

C. Simulation study

We consider a model with two level of services with the following initial recorded data $S(2)=0.15$, $\sigma(1,1)=0.28$, $S(1)=0.24$. Applying the steps in section (6) it is straightforward that (12) is satisfied and the solution is presented in Table 1 below.

Table 1. In this Table, the efficient portfolios are derived; we obtained several efficient frontiers associated to web service levels

$w(1)$	$w(2)$	$E\left\{S(i)\right\}$	$\sigma_k\left\{S(i)\right\}$
0.02	0.06	0.19	0.21
0.03	0.34	0.18	0.13

7. Conclusion

The current research focused on cloud computing analysis combining cost efficiency quality of services and web services allocation. The cloud computing web services management system structure is similar to portfolio management system. Analyzing such architecture, the best solution corresponds to the best allocation of weight measures to all existing services as to construct a low risk quality of services. We found that dynamic decision making management approach relaying on dynamic portfolio allocation can achieve the purpose of saving resources.

Acknowledgements

We are grateful to all colleagues for discussions and suggestions, which helped us to improve the ideas in this paper. The work presented in this paper was supported by

the SAKURA Research Grant and AUAF Research and Professional Development Fund. We thank Hamidullah Hamidy for helping in reviewing and editing this work. This publication reflects only the authors' views and any remaining mistakes are ours.

Conflicts of Interest

None.

References

[1] Aumann, R.J., "Lecture note on game theory', Westview Press Inc., Boulder, Colorado," (1989).

[2] Buyya R., Yeo, Venugopal C. S. Broberg J, Brandic I., "Could computing and emerging IT platform vison," 2009.

[3] Backwell, D. & Girshick, M.A. "Theory of games and statistical decisions", John Wiley & Sons, New York, 1954.

[4] Buyya, R. C. Y., "*Cloud compuing and emerging IT platforms,"* 2009.

[5] Zeng, B. Benatallah, M. Dumas, J. Kalagnanam, and Q. Z. Sheng, "Quality driven web services composition: In proceedings of the 12th international conference on world wide web," 2003.

[6] Qia, L. W. Doua, X. Zhangc, and J. Chenc, "A QoS-aware composition method supporting cross-platform service invocation in cloud environment," *J. Comput. System Sci*., vol. 78, no. 5, pp. 1316-1329, 2012.

[7] Ying, Huang et.al., "A framework for building a low cost, scalable, and secured platform for web delivery business services" *IMB Journal of Research and Development,* 2010.

[8] Markowitch H. M., "Portfolio selection: Efficient diversification of investments", John Wiley & Sons, New York, 1959.

[9] Nallur v., Bahsoon, R., "Design of a market based mechanism for quality attribute tradeoff of service in cloud', *Proceeding of ACM Symposium on Applied Computing*, pp. 367-371, 2010.

[10] Venkatesa Kumar, V. and S. Palaniswami, "A dynamic resource allocation method for parallel data processing in cloud computing," *Journal of Computer Science* 8 (5): 780-788, 2012.

[11] Zhou, M. Q., Zhang, R. W. Xie, W. N. Qian, and A. Zhou, "Security and privacy in cloud computing: A survey 2010 sixth international," 2010.

[12] Litterwood, H. M, Papov P., Stringuini S., "Modeling software design diversity: A review of computing surveys," pp. 177-208, 2001.

[13] Zeng, L. B., Benatallah, A., Ngu, M. Dumas, Kalagnanam, J. and Chang, H. "QoS-aware middleware for web services composition," *IEEE Trans. Softw. Eng*., vol. 30, no. 5, pp. 311-327, 2004.

[14] Vinothina,V. R., Sridaran, "A survey on resource allocation strategies in cloud computing," *International Journal of Advanced Computer Science and Applications*, Vol. 3, No.6, 2012.

[15] Ferguson, T. S., "Mathematical statistics - A decision theoretic approach", Academic Press, New York, 1968.

[16] Jimbo, H.C, Jesus Pascal, Isidore Ngongo and Jawad Azimi, "Portfolio optimization with nonlinear transaction costs," *International Journal of Mathematical Modelling and Numerical Optimization*, 2018 (In press).

[17] Jimbo, H.C. Ouentcheu A, Bozeman R.E, "Portfolio optimization with the growth model," *Journal of Nonlinear and Convex Analysis*, pp.131-141, Editor: W. Takahashi and T. Tanaka, Publisher: Yokohama Publisher, Japan, 2003.

[18] Craven, M. and Jimbo, H.C, "An EA for portfolio selection over multiple Investment periods with exponential transaction costs", GECCO'13 Amsterdam, The Netherlands .ACM 978-1-4503-1964-5/13/07, 2013.

[19] Feller, W. "An introduction to probability theory and its applications', John Wiley & Son", New York, vol. 2, 1972.

[20] Jimbo, H.C, Isidore Ngongo, Gabriel Andjiga and Takeru Suzuki, "Portfolio optimization under cardinality constraints: A comparative study", *Open Journal of Statistics*, 7, 731-742, 2017.

[21] Jimbo, H.C, M. Craven, "Unconstrained optimization in stochastic cellular automata", *Journal of Nonlinear Analysis and Optimization*, Vol. 2, No 1, 113-122, 2011.

[22] Jimbo, H.C. M. Craven, "Optimizing stock investment portfolio with stochastic constraints", *Journal of Nonlinear and Convex Analysis volume* I. pp. 127-141, 2011.

[23] Kolmogorov, A.N., "Foundation of the theory of probability," (3rd eds). Phasis, Moscow, 1988.

[24] Mell, P. M., Grance, T., "The NITST Definition of Cloud Computing System," *ACM Library*, 2011.