Resource Allocation In Cloud-Fog Systems Using Genetic Algorithm

Sohaib Naseer Lateef AL-ABBASI
Department of electrical and computer engineering
Altinbas University
Istanbul, Turkey
203720359@ogr.altinbas.edu.tr

Galip CANSEVER
Department of electrical and computer engineering
Altinbas University
Istanbul, Turkey
galip.cansever@altinbas.edu.tr

2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA) | 978-1-6654-6835-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/HORA55278.2022.9800052

***Abstract*— This paper's proposal is that Cloud administrators are responsible for providing resources, but users are responsible for ensuring that information travels to and from the cloud safely and securely. The supply of resources has to be understood in layers, with each layer representing a distinct genre of goods that may be offered in various ways, and in this context, we suggest the employment of genetic algorithms and fog nodes to accomplish the ideal route to the clouds and back.**

Keywords—GA, AI, Cloud, Fog.

I. INTRODUCTION

It is possible to define cloud computing as a collection of interconnected servers, whether virtual or physical, or as a collection of computational resources made available on the internet as a service as a result of advances in ubiquitous smart devices and ubiquitous network technologies such as the internet of things, smart cities, intelligent grids, augmented / virtual reality, and n manned. Communications networks are continually changing, and this is well-known across the globe. As a result, an increasing number of new technologies and apps are being created with the goal of streamlining and simplifying the operations of businesses and individuals alike. Because of this ongoing growth, new opportunities in the communications industry are always emerging. One of these options is application migration, which allows for the decentralization of specific services. The term "Cloud Computing" comes to mind in this scenario. Servers in a cloud are used to give services to the users that use this sort of system. In order to be considered seriously and given the appropriate significance, every computer network carries a huge number of complications with it. The system's usage and how to improve it to minimize expenses and boost communication speed are two of these challenges. [4]

II. MATERIALS

A. Cloud Computing

The term CAD for breast cancer can be associated with two According to the definition of "cloud computing," Internet-based servers may be used 24 hours a day, seven days a week to handle requests. Any mobile device or location may offer Internet access to your information or service, regardless of device or location. Users from a wide variety of hosting providers all around the globe may count on them for regular service. As a result of these techniques, websites are more secure against hacker attacks and police raids by both local and federal authorities. This is a favorable development since cloud computing enables for an increase in the number of network-based services. It is in the best interests of both service providers and end customers to adopt this method, since it allows them to pay only for the services they really use. As a result, the customer may save money on wages or capital expenditures (premises, specialized material, etc.). Cloud computing (also known as cloud computing) is a new paradigm for delivering services that lets clients to access and react to a catalog of standardized services in a scalable and flexible fashion, known as cloud computing, in the business and technology sector.

B. Resource Allocation Problem

One type of optimization problems in resource allocation. The most studied resource allocation problem is the traveling salesman problem (TSP), which consists of creating the route of a traveling salesman, who leaves his home, seen once by all his clients, located in different locations, and finally comes home. Depending on how you order the visits to the clients, the traveler will travel more or less distance. The objective of the traveler is to travel the minimum distance to make all the visits. A widely applied resource allocation problem is the vehicle routing problem (VRP). In this problem, you have a fleet of vehicles that must make a series of deliveries to customers. Vehicles have a maximum load capacity. Depending on the deliveries assigned to each vehicle and the route each one takes, all deliveries may be made with a different number of vehicles and with a different total distance traveled. The objective of the fleet manager is to make all deliveries with a minimum cost (or minimum total distance traveled) The fleet assignment problem (FAP) of a passenger transport company, consists of calculating the time of departure of the journeys, and define the rotations (sequence of journeys ordered in time) carried out by each type of vehicle. Depending on the schedule of the journeys, the assignment of each journey to a type of vehicle, and the sequence of journeys made by each vehicle, different transport operations can be obtained. The goal of this problem is to maximize profit or minimize costs. [4, 6].

III. PROPOSED METHOD

The context of application of this approach is, at first, the allocation of tasks submitted by a user and computing resources offered by a provider, that is, IaaS services. The

978-1-6654-6835-0/22/$31.00 ©2022 IEEE

developed algorithm proposes to perform the allocation of a task list to a set of machine instances, in a one-to-one manner. This escalation must respect the QoS level defined in the SLA. Next, we will raise some hypotheses about the proposed modes of operation, organization of tasks and resources, necessary to create a good environment for reproducing experiments and results. For practical purposes, all available cloud instances have data processing-oriented characteristics. The idea of implementing different modes of operation was introduced, not only to validate the proposed solution, but also to show the versatility of the genetic algorithm in different types of contexts, simply by adjusting the Fitness Function.

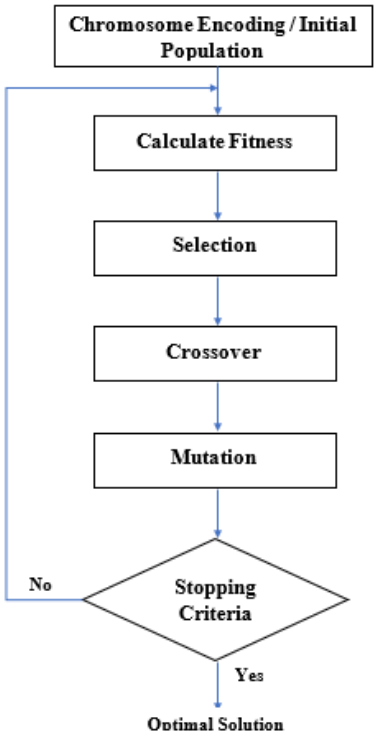

Fig1 Fitness score entities in GA

Now that we have an idea of how the allocator behaves in different modes of operation, we will see how the tasks and resources are structured. An application can be formed by several requirements, the ones chosen in this work are: Number of cores needed, memory (GB), price/hour and availability level. This minimum configuration must be submitted by the user. We will see later that the more parameters established, the more constraints will be created, and consequently the lower the performance of the algorithm. The choice of these requirements was based on the on-demand business model of Amazon Elastic Compute Cloud (EC2) [1], and Google Compute Engine (GCE) [5]. Both IaaS providers have their services based on offering resources via cloud instances, that is, the user can choose from a bank of machines (instances), with different configurations, the one most suitable to meet their needs, paying for the workstation. a fee per hour of use.

In a real case, due to the random nature of the heuristic, these constraints are usually confronted, automatically resetting the individual's score. Another concept that should be clear is that the coefficients chosen to determine the weight of the parameters in each method of operation were determined in a rather unrefined way, and that in order to utilize the full potential of the modes of operation, the user of the genetic approach must use a judicious methodology in determining their Aptitude Function, verifying the greatest needs and assigning a corresponding multiplication factor. Given that all individuals have been evaluated and now have a fitness score attached, the algorithm enters the selection phase, where it will choose two individuals by the "Selection Roulette" method for reproduction and creation of a new population.

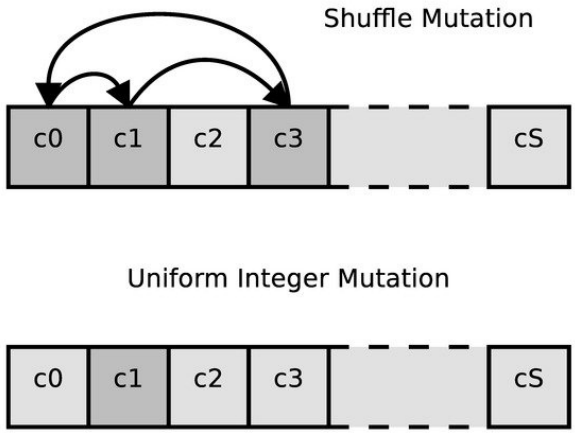

Fig.2 Selection Roulette or Shuffle mutation in GA

The algorithm calculates the total Fitness Score of a population and generates a random number between zero and this value. From there, the heuristic runs through the individuals and adds their scores, until the accumulated value exceeds the randomly generated number. In possession of the two individuals selected in the previous step, the AG performs a recombination of the two in order to generate an even fitter offspring, and as previously introduced, we call this process a crossover. The probability that recombination will occur is regulated by the Crossover Rate parameter, a value sensitive to the problem. The rate chosen for our algorithm is 90%, this value is well accepted in the literature and the algorithm behaved well with this parameter. If the heuristic does not perform a crossover, the selected chromosomes are simply copied to the next generation. Since we have implemented a standard genetic algorithm, the chosen crossover mode is based on the simplest model with only one division point.

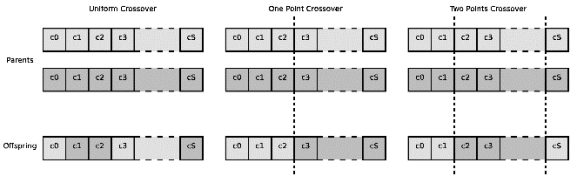

Fig.3 Crossover operators (representing tasks in cloud and fog

Tasks are the main units of our study, and as presented in the coding section, they are arranged in an ascending fashion on a chromosome. On the other hand, the machine instances provided by the clouds are randomly organized. Making a very basic analogy, a very common practice in real situations is that when a person gets lost from another, we say that the most effective strategy is for one person to stay still while the other looks, starting from that, looking for the most effective

way to find better combinations, we noticed that keeping the tasks fixed and recombining the machine instances increased the algorithm performance and considerably decreased the number of violated constraints compared to the crossover in both tasks and machines. The recombination model, therefore, was modeled as follows: The concept of the next step is very simple, the two offspring generated in the crossover process are submitted to the Mutation operator. This function traverses the chromosome gene by gene and according to the Mutation Rate, the machine instance of this chromosome can be replaced by any other instance within the cloud bank. As well as the crossover rate, the value of the mutation rate varies a lot depending on the problem to be optimized, even so, it usually has a low value for the heuristic to be viable, serving to generate genetic variability in the

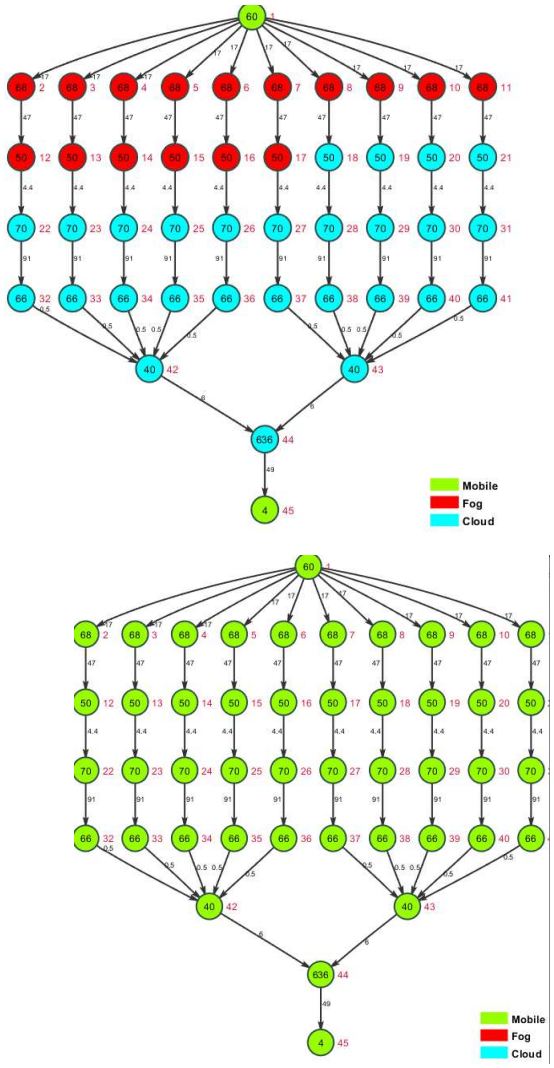

Figure.4 Cloud and fog nodes before and after resource allocation in GA

solution and prevent the algorithm from becoming stuck in a great location. The mutation rate chosen in our project is 5%, a low value, as indicated by several literatures [34], [23], and which obtained a satisfactory behavior for our challenge. An illustration of this process can be seen in Figure 4. Arriving at the final part of the main loop of the algorithm, we have a new population of descendants stored on a temporary basis that must be reinserted in the main structure for storing individuals. Once this relocation is done, the FS of all individuals is calculated again and checks if any of them has reached a minimum limit of satisfaction:

IV. Conclusion

The present work made it possible to identify the advantages and properties of a fog computing scenario based on the genetic algorithm GA optimization. The results have a certain advantage of the decentralized structure for applications that demand less response time. Structures located in the cloud are totally dependent on the state of the external network (i.e., variable characteristics related to the path between devices and servers in the cloud). Decreasing the frequency with which information needs to leave the devices' local network makes applications more efficient and less dependent on external entities located in the cloud. As future research opportunities within this theme, different hardware configurations can be explored both in the cloud and in the local structure of the fog. The assessment can be extended to other relevant metrics such as, for example, reliability, availability, privacy and cost of implementing and maintaining the structures involved.

V. REFERENECES

[1] F. Chen, J. Schneider, Y. Yang, J. Grundy and Q. He, "An energy consumption model and analysis tool for Cloud computing environments," Green and Sustainable Software (GREENS), 2012 First International Workshop on , vol., no., pp.45-50, 3-3 June 2012

[2] R. Li, Q. Zheng, X. Li, and J. Wu, ‘‘A novel multi-objective optimization scheme for rebalancing virtual machine placement’’, In the Proceedings of the IEEE 9th International Conference on Cloud Comput., San Francisco, CA, USA, pp. 710–717, 2016.

[3] Boru, D. Kliazovich, F. Granelli, P. Bouvry, A.Y. Zomaya, “Energy-Efficient Data Replication in Cloud Computing Datacenters”, IEEE Globecom workshop on Cloud Computing Systems, Networks and Applications, pp. 446-451, 2013.

[4] Ghribi, M. Hadji, D. Zeghlache, “Energy efficient vm scheduling for cloud data centers”, IEEE/ACM 13th International Symposium on Cluster, Cloud, and Grid Computing, p.p 671-678, 2013

[5] Hui, G., Zhang, J., Li, J., et al. (2016). Sucrose quantitative and qualitative analysis from tastant mixtures based on Cu foam electrode and stochastic resonance. Food Chemistry, 197, 1168–1176.

[6] Shi, P., Yuan, D., Han, D., et al. (2018). Stochastic resonance in a time-delayed feedback tristable system and its application in fault diagnosis. Journal of Sound and Vibration, 424, 1–14.

[7] L. Zhou, ‘‘On data-driven delay estimation for media cloud’’, IEEE Trans. Multimedia, Vol. 18, No. 5, pp. 905–915, 2016Bong, D., Ting, K., and Lai, K. (2008). Integrated approach in the design of car park occupancy information system (coins). IAENG International Journal of Computer Science, 35(1):7–14.

[8] Chatfield, K., Simonyan, K., Vedaldi, A., and Zisserman, A. (2014). Return of the devil in the details: Delving deep into convolutional nets. Computing Research Repository, abs/1405.3531.

[9] S. Pelley, D. Meisner, T. F. Wenisch, and J. W. VanGilder, “Understanding and absracting total data center power,” in WEED: Workshop on Energy Efficienct Design, 2009

[10] Yamapi, R., Yonkeu, R. M., Filatrella, G., et al. (2018). Effects of noise correlation on the coherence of a forced van der Pol type birhythmic system. Communications in Nonlinear Science and Numerical Simulation, 62, 1–17.

[11] Su, J., Sheng, Z., Xie, L., et al. (2019). Fast splitting-based tag identification algorithm for anti-collision in UHF RFID system. IEEE Transactions on Communications, 67(3), 2527–2538

[12] Zhao, F., & Guibas, L. J. (2004). Wireless sensor networks: an information processing approach. Burlington: Morgan Kaufmann.