

QoS-aware Resource Management in SDN-Based InterClouds : A Software Cybernetics Perspective

Suleiman Onimisi Aliyu, Feng Chen and Ying He

Software Technology Research Laboratory (STRL),

De Montfort University,

Leicester, UK.

soaliyu@hotmail.com fengchen@dmu.ac.uk, ying.he@dmu.ac.uk

Abstract— Software Defined Networking (SDN) paradigm is introducing new architectural approaches for many unresolved issues of networking. These new approaches are imperative in emerging scenarios where user requirements keep growing, the required bandwidth keeps increasing, and so does the variety of applications (e.g. Big data analytics) that suggest the network plays a more prominent role. As such, our research aims to provide a novel self-tuning (adaptive) resource management technique in SDN-based InterCloud environments. Essentially, this workshop paper presents a policy-based quality of service (QoS) control framework described using principles in coalition games with externalities or partition form games (PFGs). More specifically, we model QoS-aware resource management in the SDN-based InterCloud platform as an adaptive control problem. This on-going work outlines a proposed dynamic programming and anytime approach (Integer partition based) to solve the multi-criteria optimisation problem of a Markov decision process (MDP) model for system dynamics in SDN-based InterClouds.

Keywords—*Software Cybernetics; Software Defined Networks; InterCloud Computing; Adaptive Resource Management; QoS Control; Coalition games with externalities.*

I. INTRODUCTION

It is becoming clear that this era is of software pervasiveness. Increasingly, modern products and services embed software or are customised, optimised or managed using software [1]. Dynamic environments, rapidly changing requirements, unpredictable and uncertain operating conditions require a new mode of application development and deployment. Software and services need to become smarter, self-organised, sustainable, resource efficient, robust and safe to meet stakeholder demands [2]. As Cloud systems are complex, large-scale, and heterogeneous distributed systems, management of their resources is a challenging task. They need automated and integrated intelligent strategies for provisioning of resources to offer services that are secure, reliable, and cost-efficient.

According to the Global Intercloud technology forum [3], the InterCloud is a cloud model that guarantees performance and availability of services, by facilitating on-demand reassignment of resources and transfer of workload through interworking of cloud systems and using standard interfacing. A rapid surge in Internet of Things (IoT) or big data applications which produce and process copious amounts of

data may require resources that can only be provided by a federation of Clouds. More so, considering cost implications, it is unlikely proprietary cloud infrastructure vendors will establish data centers at all geographic locations to satisfy the low-latency requirements of future consumer-centric IoT applications. To mitigate these bottlenecks of current Cloud models, essentially, the InterCloud, an interconnected global “Cloud of clouds”, is increasingly perceived as a viable paradigm for the future of cloud computing or data centers.

Software Defined Networking (SDN) has emerged as a promising approach for providing flexible network programmability, particularly for data centers. It facilitates dynamic configuration, operation and monitoring of a network. SDN heralds unprecedented simplification in network management and innovation by service providers, and its control model presents itself as a viable solution to the issues in network virtualisation. [4].

Nevertheless, managing resources in SDN-based InterCloud computing, an important functional requirement for emerging IoT or big data applications, has received very little research attention.

This paper discusses a proposed self-adaptive approach to QoS-aware resource management in SDN-based InterCloud application environments. Specifically, in this research, we present a novel policy-based QoS control framework based on theoretical principles in partition form games with positive and negative externalities, denoted PFG⁺ and PFG⁻ respectively. In the presence of changing environments or requirements, the fundamental questions of research in software cybernetics remain pertinent, when or how (quantitatively) can software, its processes or behavior be adapted to meet objectives or goals. The contribution of this work to literature includes:

- An adaptive QoS control architecture for SDN-based InterCloud platforms using a feedback control mechanism related to software cybernetics.
- A distributed control approach for QoS-aware resource management in SDN-based InterClouds.

The rest of this paper is organized as follows. Section II discusses resource management policies. Section III presents the exact policy-based QoS framework for resource management in the control pane of the SDN-based platform. Section IV describes the proposed self-adaptive management technique in SDN-based InterClouds. Section V concludes the paper and highlights future work.

II. BACKGROUND - POLICIES FOR RESOURCE MANAGEMENT

Software Cybernetics is a field that addresses key issues and research questions related to: (i) formalism and quantifying feedback mechanisms in software processes and software, (ii) adapting concepts in control theory to software processes and software, (iii) applying principles in software theory or engineering to control systems and processes. The performance and quality of service (QoS) is important for network software [5], [6]. The network should adjust its operation and service scenarios in accordance with the changes in traffic flow. This is again a control problem of software and a largely unexplored research topic, although there have been some research efforts in this area. Due to the distribution of services on the Internet, the Quality of Service (QoS) and dependability have been a concern which may prevent web services being applied in mission-critical applications. Among other techniques, testing is critical to assure the QoS of services [7].

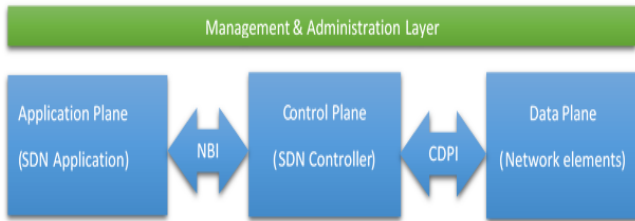


Fig.1. A Generic architecture for Software-defined Networks (SDN).

Reference [8] applied software cybernetics to manage adaptation behaviour of complex network systems, in which model-based software techniques were employed to assess the quality of adaptation in a network system in the presence of uncontrollable external environment conditions.

The constant growth in IT is making communication networking more and more complex to manage. One very promising solution is Software Defined Networking (SDN) that decouples the data and control planes, having a centralised or distributed controller (s) for the network. This gives chances to control and manage the network as desired, thus opening many new possibilities [9].

An SDN controller (the CNG Manager) was proposed by [10] to achieve dynamic and on demand InterCloud networking that can handle any kind of underlying networking technology when multiple Cloud providers are involved. The novelty of their solution consists in enabling the control of connectivity between distributed resources acquired from multiple cloud providers.

A core functionality required for any man-made system is resource management. Necessarily, managing resources affect the basic criteria for evaluating computer systems which include cost, functionality, and performance. Necessarily, there are about four mechanisms or enabling techniques for implementing resource management policies in cloud computing [11], these are:

- *Control theory:* Predicts local behaviour, using feedback to guarantee system stability and predict transient behaviour.

- *Machine learning (ML):* A major advantage of applying machine-learning techniques is the absence of a performance model of the system. ML techniques can be applied to coordinating several autonomic system managers.
- *Utility-based:* Utility-based approaches require a performance model and a mechanism to correlate user-level performance with cost.
- *Market-oriented/economic mechanisms:* Mechanisms don't necessarily need a system model, such as combining auctions for bundles of resources.

An SDN-based management and control framework for fixed backbone networks was presented in [12]. The method provides support for both static and dynamic resource management applications. Its architecture is compatible with the generic ONF SDN model and comprises three layers which interact with one other via a set of interfaces.

In Interconnected Cloud environments, autonomic policies are of high interest because of the scale of the system, the barrage of service requests, the large consumer population and the variability of the load. The ratio of the mean to the peak resource demand can be very significant [13]. An important feature of cloud computing is elasticity, described as provisioning and de-provisioning resources in the computational cloud to fit variations in their workload. Control loops and thresholds-based rules are concepts in autonomic computing that ease the construction of elastic cloud computing solutions.

Adami et al. [9] built a system to enable QoS control and routing in Software Defined Networks. When the OpenFlow controller installed a rule for a flow, it also took care of placing it in the right queue. The experimental results showed the system behaving as expected, managing in a more efficient way the network resources and giving guarantees about traffic handling.

Bari et al. [14] presented the design and implementation of PolicyCop, an autonomic QoS policy enforcement framework based on SDN. PolicyCop provides an interface for specifying QoS policies and exploits the northbound API of SDN controller to enforce them. PolicyCop takes advantage of control applications to monitor the compliance of the policies and autonomically adapts the control plane rules with changing traffic conditions.

A game theoretic approach was adopted by [15] to facilitate the formation of InterCloud coalitions and his work devised a four-stage cloud to cloud interaction protocol and strategies for InterCloud agents. Mathematical proofs showed the InterCloud coalition formation strategies converge to a sub-game perfect equilibrium and every cloud agent in an InterCloud coalition received a payoff equal to its shapely value.

III. FRAMEWORK FOR POLICY-BASED QoS-AWARE RESOURCE MANAGEMENT IN SDN-BASED INTERCLOUDS

A. Resource Coalitions and QoS Policies.

Coalition formation is necessary for solving applications in ubiquitous computing where it is necessary to adapt changing resources and environments [16]. Assume the set A comprises singleton resource pools denoting the processing elements (e.g. VMs) for the application environment. Assume the pools also come from q participating cloud providers in the data plane of the SDN-based InterCloud environment, then, $A = \{\rho_1, \rho_2, \rho_3, \dots, \rho_q\}$.

Then by definition, a resource coalition R in the SDN-based InterCloud platform is a non-empty subset of A ($R \subseteq A$) [17]. Theoretically speaking, the set of all possible coalitions is given by 2^q . By Implication, this set also includes the empty set (\emptyset) as a coalition, the empty coalition. The set A is also a coalition, called the grand resource coalition comprising resource pools from all q cloud providers. To adjust resource capacity dynamically, we propose application resource provisioning behavior in the control plane of SDN-based InterCloud platforms is done programmatically. Using open standards like OpenFlow [18], resources are pooled via a centralised entity called SDN controller or network packet broker, to constitute resource coalitions. Thus, a subset R of A , called a feasible resource coalition is defined using set notation as:

$$R = \{\mathcal{R} \mid \mathcal{R} \subseteq A, \mathcal{R} \neq \emptyset\} \quad (1)$$

A coalition structure or QoS policy in the control plane of the SDN-based environment is defined as a partition (comprising resource coalitions, see (1)) of A , that is, $\Pi = \{R_1, R_2, \dots, R_k\}$, where $1 \leq k \leq q$. Necessarily, two simple rules guide the process for generating feasible QoS policies.

- (i) $R_n \neq \emptyset$, for any $n \in 1, 2, \dots, k$ and $\bigcup_{n=1}^k R_n = A$
- (ii) $R_l \cap R_h = \emptyset$, for any $l, h \in 1, 2, \dots, k$ and $l \neq h$

The state space in several applications is derived from an important activity called mapping, which assigns each coalition structure or policy to a range of payoffs for each resident resource coalition. A situation whereby the reward of a resource coalition is independent of the behavior of ‘external’ resource coalitions is a mapping referred to as characteristic function. However, in many realistic applications of coalition formation, the feasible payoffs for resource coalitions is affected by ‘externalities’ hence the notion of partition function [19]. We denote $v(R, \Pi)$ as the value or measured (monitored) performance of an embedded resource coalition R in the execution of QoS policy Π . Furthermore, we can compute the value (performance) V of a QoS policy Π as,

$$V(\Pi) = \sum_{R \in \Pi} v(R, \Pi) \quad (2)$$

B. QoS Policy Management Graph

The problem of finding a policy Π that maximises QoS can be described as a search in a coalition structure generation (CSG) graph [20], a graph in which every node represents a

QoS policy. The policy management graph for SDN-based InterClouds is based on a novel representation of the policy search space called Integer partition (IP) graph [21], [22]. Using the IP representation, every node in the policy management graph represents a policy subspace that comprises feasible policies, nodes are categorised into k levels (partition spaces) denoted as I_1, I_2, \dots, I_k [17]. Partition space I_n contains the policies comprising n resource coalitions. For instance, if the number of cloud providers or ‘singleton’ resource coalitions in the SDN-based InterCloud environment is $k = 6$, then, in the worst case, the policy search space is the set G of all integer partitions (policy subspaces) of k is given as,

$$G = \{[6], [3] [3], [2] [4], [1] [5], [2] [2] [2], [1] [2] [3], [1] [1] [4], [1] [1] [2] [2], [1] [1] [1] [3], [1] [1] [1] [1] [2], [1] [1] [1] [1] [1] [1]\} \quad (3)$$

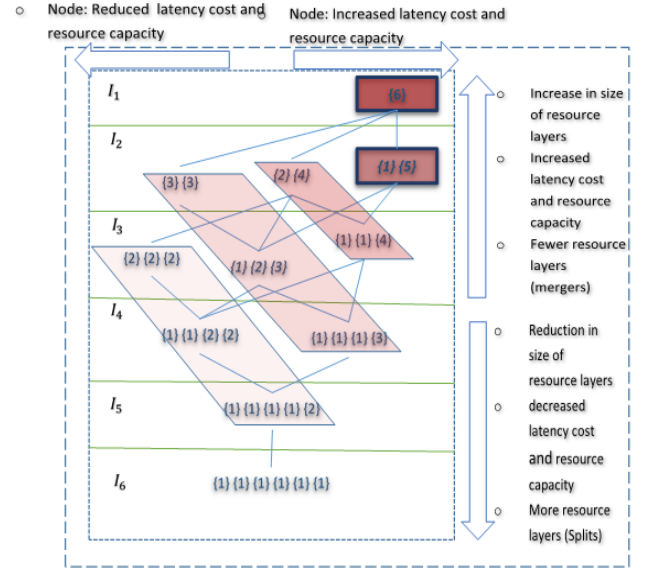


Fig.2. IP-based QoS Policy Management graph for 6 processing elements (data centers).

C. Research Problem

The Coalition Structure Generation (CSG) problem, an important activity of coalition formation, involves partitioning the set of network processing elements into exhaustive and disjoint coalitions (QoS policies) following (i) and (ii) to maximise social welfare [20] [21]. More so, in many real-world application environments, there are significant externalities from coalition formation in which the performance of one coalition may be affected by the formation of another distinct coalition [19], [22].

IV. ADAPTIVE QoS-AWARE RESOURCE MANAGEMENT IN SDN-BASED INTERCLOUDS

The game-theoretic system dynamics model for adaptive QoS-aware management is described as a Markov decision process (MDP) denoted as $(G, T, Q, \hat{\eta})$.

Where G is the IP-based QoS Policy management graph. T is dimensionless measured time with a strict total ordering. Q is

the set of states in the SDN-based InterCloud environment. The state transition function $\dot{q}: T \rightarrow Q$.

Feedback loops enable generic mechanisms for controlling the execution, maintenance, and evolution of software and computing systems [23]. As shown in Fig.3, we propose a control architecture that relies on a feedback loop mechanism (monitor) for implementation of QoS-aware resource management in SDN-based InterCloud platforms.

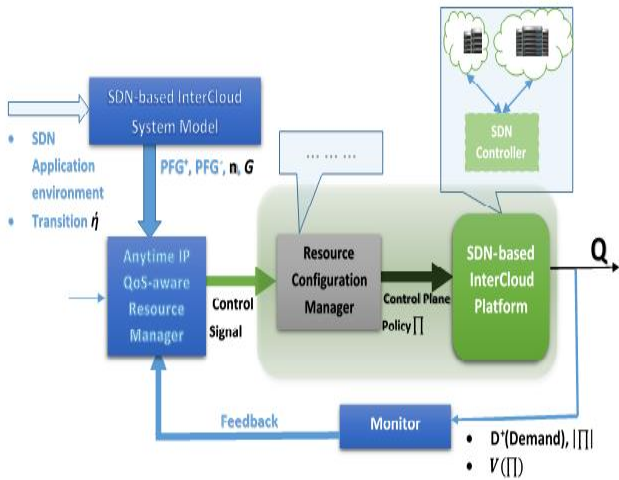


Fig.3. QoS control architecture using feedback for adaptive resource management in SDN-based InterClouds.

A. IP Anytime Algorithm for Adaptive QoS Policy Control

The IP algorithm proposed in [21], [22] is based on an integer partition representation of the search space that groups policies into policy subspaces based on the sizes they contain. Using this representation, it is possible to establish bounds (upper and lower) on the value (performance) of the best policy in every policy subspace P_I^A of $I \in I_n$ through search. More precisely, let Max_s and Avg_s be the maximum and average value (reward) of policies of size s ($s \leq n$). For all Integer partitions $I \in I_n$, it is possible to compute the upper bound $UB_I = \sum_{s \in I} max_s$ and lower bound $LB_I = \sum_{s \in I} Avg_s$, respectively [17], [22]. These bounds are then used to guarantee worst case scenarios on the quality of the best policy found so far and also to prune redundant policy subspaces. As for the remaining policy subspaces, they are searched one after another unless a value is found that is higher than the upper bound of other subspaces, in which case we abandon searching that policy subspace.

The IP-based anytime algorithm proceeds by performing partial search to establish bound on the optimal QoS policy in SDN-based InterCloud platform by observing the maximum (best) value of each resource coalition R in some policy Π .

In games with externalities, every policy subspace (Integer partition) $G_I: I \in I^n: |I| \leq 2$ must be searched in PFG⁺ class settings (Fig. 4 & 6a) and the number of policies searched (or search cost) is 2^{n-1} with a worst-case ratio bound $\beta = n$. Positive externalities may occur in real-world applications where there are overlapping goals between resource coalitions in the SDN-

based InterCloud environment. For negative externalities, it is necessary to search every partition $G_I: I \in \{[n], [n-1,1], [n-2,1,1], \dots, [1,1,\dots,1]\}$ (Fig.5 & 6b). Here, the number of policies searched (or search cost) is $2^n - n$ and $\beta = \lceil n/2 \rceil$ (proof in [21]). The negative externalities are typical in applications where resources are shared. Particularly, as resource coalitions move towards their goals, consuming more resources in the pool, they tend to move the world further from other resource coalitions.

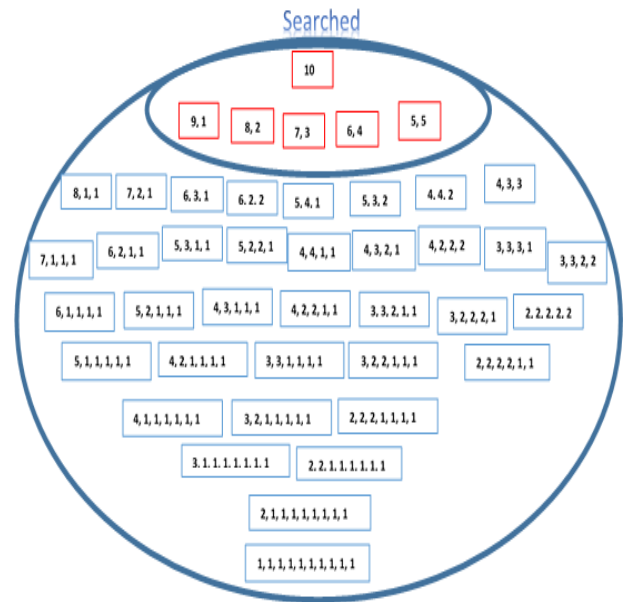


Fig. 4. Partial search for PFGs with positive externalities (10 resource Clouds).

B. Optimisation of QoS Policies

The goal is to search (test process) for the most effective and efficient deployment of application resources (via QoS policies) across Clouds and prevent a deterioration of application throughput in the SDN-based InterCloud environment. This approach proposes to make optimal decision in the presence of trade-off between the performance and cost of QoS policies in the SDN-based InterCloud. The aim is then to find a Pareto optimal solution for the coalition structure generation (or CSG) problem comprising the two prominent objectives (value or performance of QoS policies and its behavioral complexity cost). The worst case behavioral (communication) complexity cost for QoS policies shows a linear trend (centralised architecture) as depicted in Fig.2. In many situations, the system administrator's utmost importance for resource management is the value (performance) of QoS policy over its communication complexity cost. Any automated solution process for the coalition structure generation problem should capture these preferences.

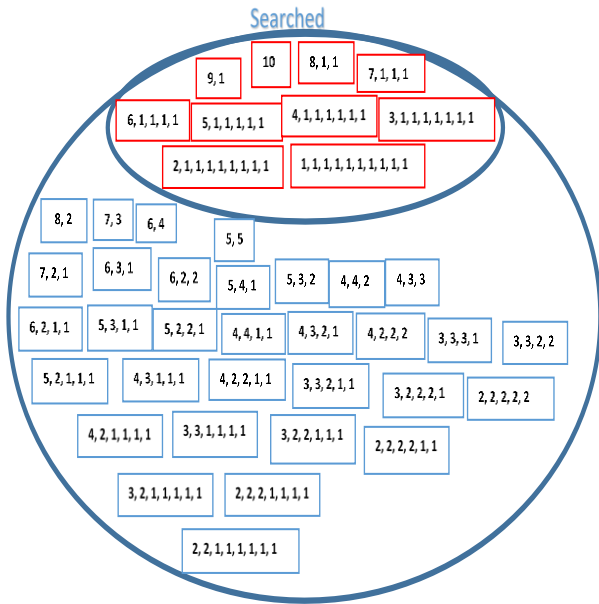


Fig. 5. Partial search for PFGs with negative externalities (10 resource Clouds)

Accordingly, this effort proposes an a priori multi-criteria optimisation (MCO) method called a lexicographic method, which assumes that objectives can be arranged in a hierarchy of preference or importance [24]. Particularly, the lexicographic method comprises two single-objective optimisation problems for the coalition structure generation problem. Intuitively, the performance (Value) of QoS policies is given the highest priority whereas, cost attributed to behavioral complexity is given the least priority. The bi-criteria optimisation problem can be solved using a scalarisation approach called the weighted sum method.

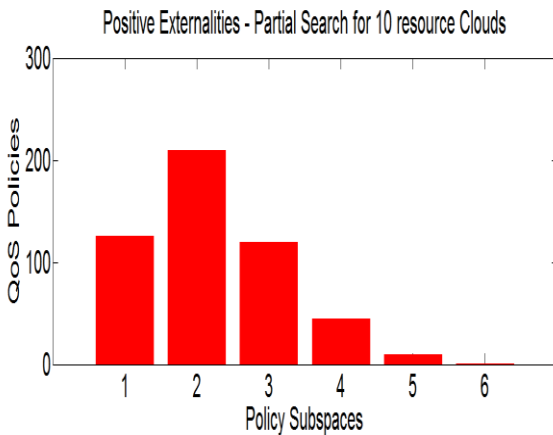


Fig.6a. Partial Search for PFG⁺.

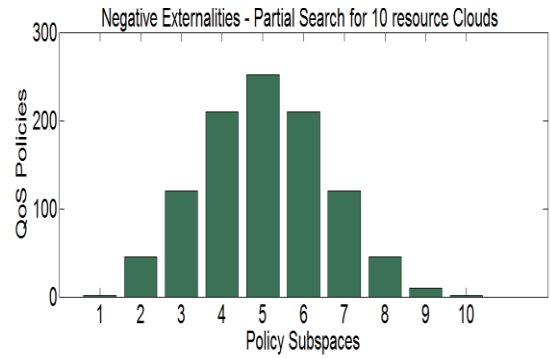


Fig. 6b. Partial Search for PFG⁻.

V. CONCLUSION AND FUTURE WORK

This paper has proposed a QoS control architecture for adaptive resource management in SDN-based InterClouds using a feedback control loop mechanism in software cybernetics. Accordingly, we have presented a policy-based QoS management framework using theoretical framework in partition form games (PFGs). This approach models system dynamics in SDN-based InterCloud environments as a Markov decision process (MDP). Furthermore, we implement QoS policy optimisation as the solution to a lexicographic ordering coalition structure generation maximisation problem and adopt the IP-based anytime algorithm to solve it in both positive and negative externalities. In future work, we plan to architect an SDN-based InterCloud platform and use it to describe emerging application scenarios like the Internet of Things. Simulated test-beds would be implemented for the application environment and experiments carried-out to benchmark or evaluate our approach.

REFERENCES

- [1] P. Bellavista et al. Software Engineering Key Enabler for Innovation. NESSI Working Group: Networked European Software and Services Initiative. White Paper. 2014.
- [2] H. Yang, F. Chen, and S. Aliyu. "Modern software cybernetics: New trends." (2017): 169-186.
- [3] Use Cases. "Functional Requirements for Inter-Cloud Computing." In *Global Inter-Cloud Technology Forum, GICTF White Paper*. 2010.
- [4] E. Haleplidis, K., Pentikousis, S., Denazis, J. H., Salim, D., Meyer, and O., Koufopavlou. *Software-defined networking (SDN): Layers and architecture terminology*. No. RFC 7426. 2015.
- [5] K.Y. Cai, J. W. Cangussu, R. A. DeCarlo, and A.P. Mathur. "An overview of software cybernetics." In *Software Technology and Engineering Practice, 2003. Eleventh Annual International Workshop on*, pp. 77-86. IEEE, 2003.
- [6] J. W. Cangussu, K.Y. Cai, S. D. Miller, and A.P. Mathur. "Software cybernetics." *Wiley Encyclopedia of Computer Science and Engineering* (2008).
- [7] X. Bai, Y., Chen, and Z., Shao. "Adaptive web services testing." In *Computer Software and Applications Conference, 2007. COMPSAC 2007. 31st Annual International*, vol. 2, pp. 233-236. IEEE, 2007.
- [8] K. Ravindran. "Software cybernetics to manage adaptation behavior of complex network systems." In *Computer Communication and Networks (ICCCN), 2014 23rd International Conference on*, pp. 1-8. IEEE, 2014.
- [9] D., Adami, L., Donatini, S., Giordano, and M., Pagano. "A network control application enabling software-defined quality of service." In

Communications (ICC), 2015 IEEE International Conference on, pp. 6074-6079. IEEE, 2015.

- [10] M. Mechtri, I. Houidi, W. Louati, and D. Zeglache. "SDN for inter cloud networking." In *Future Networks and Services (SDN4FNS), 2013 IEEE SDN for*, pp. 1-7. IEEE, 2013.
- [11] D.C. Marinescu. *Cloud computing: theory and practice*. Newnes, 2013.
- [12] D., Tuncer, M., Charalambides, S., Clayman, and G., Pavlou. "Adaptive resource management and control in software defined networks." *IEEE Transactions on Network and Service Management* 12, no. 1 (2015): 18-33.
- [13] A.N. Toosi, R. N. Calheiros, and R. Buyya. "Interconnected cloud computing environments: Challenges, taxonomy, and survey." *ACM Computing Surveys (CSUR)* 47, no. 1 (2014): 7.
- [14] M.F., Bari, S. R. Chowdhury, R., Ahmed, and R., Boutaba. "PolicyCop: An autonomic QoS policy enforcement framework for software defined networks." In *Future Networks and Services (SDN4FNS), 2013 IEEE SDN For*, pp. 1-7. IEEE, 2013.
- [15] K. M. Sim. Agent-based interactions and economic encounters in an intelligent InterCloud. *IEEE Transactions on Cloud Computing*, 2015, 3(3), 358-371.
- [16] B. Horling, and V. Lesser. "A survey of multi-agent organizational paradigms." *The Knowledge Engineering Review* 19, no. 4 (2004): 281-316.
- [17] S.O., Aliyu, F. Chen, and H., Li. "A Self-Tuning procedure for resource management in InterCloud Computing." In *Software Quality, Reliability and Security Companion (QRS-C), 2016 IEEE International Conference on*, pp. 326-333. IEEE, 2016.
- [18] Foundation, Open Networking. "Software-defined networking: The new norm for networks." *ONF White Paper* 2 (2012): 2-6.
- [19] Thrall, Robert M., and William F. Lucas. "N-person games in partition function form." *Naval Research Logistics (NRL)* 10, no. 1 (1963): 281-298.
- [20] T., Sandholm, K., Larson, M., Andersson, O., Shehory, and F., Tohmé. "Coalition structure generation with worst case guarantees." *Artificial Intelligence* 111, no. 1-2 (1999): 209-238.
- [21] T. Rahwan, S.D. Ramchurn, V.D. Dang, A. Giovannucci, and N.R. Jennings. Anytime optimal coalition structure generation, 2007, In *AAAI*, 7, pp. 1184-1190.
- [22] T. Rahwan, T. Michalak, M. Wooldridge, and N.R. Jennings. Anytime coalition structure generation in multi-agent systems with positive or negative externalities, 2012, *Artificial Intelligence*, 186, 95-122.
- [23] H., Giese, Y. Brun, J. Di Marzo Serugendo, C. Gacek, H. Kienle, H. Müller, M. Pezze, and M. Shaw. "Engineering self-adaptive and self-managing systems." *Software Engineering for Self-Adaptive Systems. LNCS 5525* (2009): 47-69.
- [24] K.W., Jee, D.L., McShan, and B.A., Fraass. "Lexicographic ordering: intuitive multicriteria optimization for IMRT." *Physics in Medicine and Biology* 52, no. 7 (2007): 1845