

Multi-Objective Firefly Simulated Annealing (MOFSA) Based Inter-Data Center Network Dimensioning Under Cloud Computing Environment

*¹K. Srilakshmi, ²P. Bhargavi,

^{1,2}Sri Padmavathi Mahila Visva Vidyalayam, Tirupati,

Email: srilakshmi21kakarla@gmail.com, pbhargavi18@yahoo.co.in

Received:12th June 2017, Accepted:25th June 2017, Published: 1st September 2017

Abstract

Cloud Computing (CC) presents a big chance for Inter-Data Center (IDC) network to modernize its infrastructure, systems and services. In addition these IDC network, the allotment and liberate of resources by various cloud customers bring about noteworthy energy utilization, less security and Virtual Machine (VM) position of at the IDC that consecutively enhances the Operational expenditures (Opex) of the operators. In order to resolve these troubles, this research merges energy effectiveness, virtual machine assignment, immigration restraints and Time of Use (ToU) awareness, and presents a new virtualization method, that is to say Time of Use Resource Provisioning (ToURP) awareness ToUP for an IDC network on top of an Internet Protocol (IP) on Wavelength Division Multiplexing (WDM) backbone. At first, we propose a Mixed Integer Linear Programming (MILP) formulation to model the best possible performance of ToURP. On the other hand additionally the traffic jammed between two backbone nodes, upstream user insists predestined to data centers and downstream data center insists beginning from numerous data centers. As the IDC network requires to be rearranged in polynomial time, we present a Multi-Objective Firefly Simulated Annealing (MOFSA) based algorithm for depiction of memory, storage, and communication resources in IDC network. This holistic technique is to sustainable ToURP, the toolkit is intended to give a base for a consistent and hopeful CC. They assess ToURP under diverse situations, and mathematical outcomes prove that noteworthy Opex savings could be attained whilst insists could be stipulated with less energy utilization in the IDC.

Key Words: *Cloud Computing (CC), Time of Use Resource Provisioning (ToURP), Service Level Agreement (SLA), Mixed Integer Linear PROGRAMMING (MILP), Inter-Data Center (IDC) Network, Multi-objective Firefly Simulated Annealing (MOFSA).*

1. Introduction

Cloud Computing (CC) is identified as presently one among the very trendy techniques existing; customers make use of the idea of “pay-as-you-go” for applications, computing and storage resources [1]. The ecological effect of data centers is important and the development forces of the Information Technology (IT) sector recommends professionals to deal with this issue. Since the quantity of data demanding applications amplify in the Cloud, use of huge amounts of physical servers boost radically not only for computing purpose but for as well replication intentions [2].

Consequently, amongst these confronts, energy competence in the Cloud appears to be an important problem because of the huge raise in Greenhouse Gas (GHG) discharges of data centers [3] and the Opex of data center operators [4], that is to say the electricity bills. In [5], the writers reveal the importance of the stability among storage, process and transport forces. Consequently energy effectiveness in CC contains quite a lot of facets for instance energy responsive workload and virtual machine placement on the physical servers [6], thermal observation of the data centers to deal with intra-data center workload allotment [7], and energy proficient Cloud service provisioning on top of the backbone network [8] all the way through intellectual exploitation of any cast/many cast-based communication protocols.

A few researchers [9-10] accept the truth that the energy utilization of entire server changes just about linearly with the Central Processing Unit (CPU) utilization. Beloglazov et al [11] presented an energy-competent resource allocation technique dependent upon energy utilization model of CPU. This technique efficiently decreased energy utilization whilst fulfilling the Quality of Service (QoS). However, it didn't think resources except CPU. Srikantaiah et al [12] learned the effects of system resources consumption on energy consumption and functioning. On the other hand these techniques they didn't examine the best possible operating points amid energy utilization and power utilization.

Numerous associated attempts take in hand data center's price deduction via energy pricing and power

utilization. In Sharma [13], the writers use fast active/inactive state changing over a collection of servers to adjust to changeable power restrictions. Rao, et al [14] searches to reduce on the whole prices for various data centers situated in dissimilar energy marketing areas. But in real time data with the purpose of accordingly enhances the power utilization and the electricity bill.

The very admired techniques amongst previous time changing tariffs, persuades a shift to Real-Time Pricing (RTP) [15], usually recognized to be hourly retail prices which traverse the clearing cost in either the day-ahead or real-time wholesale market. On the other hand, the resultant enhance in difficulty can be unwanted if customers are incapable to react to them proficiently or are miserable with probable enhances in bill instability. Additionally, all have concentrated on a restricted set of rate configurations, covering up simply choose two and three-tier Time of Use (ToU) pricing rates [16]. Resolve for a broader range of ToU configurations, and expanding to Critical Peak Pricing (CPP) [17] and RTP+. All the way through the utilization of a model this identifies best possible prices for numerous rate structures. From the above pricing techniques customers could as well decrease their electric bills by subsequent ToU rates. Initially, customers which priced flat rates might be disinclined to unite RTP based programs wherever electricity charges differ very often. And then, customers might not simply adjust to the flexibility of RTP tariffs.

In [18], ToU-awareness was expanded to enclose inter-data center traffic. In ToU awareness the Delay and Power is minimized by using Delay and Power Minimized Provisioning (DePoMiP) [8]. At firstly Holistic Model is presented to the VMs placement and migration restraints. At this time, the sub-optimality is carried out dependent upon the outcomes got from the resolution of the Mixed Integer Linear Programming (MILP) problem formulation method. One time the ToURP awareness based Multi-Objective Firefly Simulated Annealing (MOFSA) based meta-heuristic technique that intends at getting a DC workload sharing map. Performance measures in this contrast are electricity charge, power utilization, network resource utilization and path impediment.

2. Literature Review

Awada et al [19] introduced formulations and resolutions for Green Cloud Environments (GCE) to reduce its environmental effect and energy utilization underneath novel models by taking static and dynamic parts of cloud components. They examine to save 20% of energy utilization in cloud data centers. Li et al [20] presented an active energy proficient VM immigration and consolidation technique dependent upon a multi-

resource energy-proficient model. It proves superior energy competence in data center for CC.

Beloglazov et al [21] get better the consumption of computing resources and decrease energy utilization underneath workload self-governing excellence of service restraints. Ilyas et al [22] introduces a Relocate Energy Demand to Better Locations (RED-BL) structure for identifying a best possible plotting of workload to a previous collection of data centers whilst taking the price of workload relocation. RED-BL could decrease the electric bill by to the extent that 81% from the case while the workload is uniformly dispersed between data centers.

Zhang et al [23] present a new electricity bill capping technique which not just reduces the electricity charge, but as well imposes a cost plan on the monthly bill for cloud-scale data centers and guarantees the QoS. Tafani et al [24] present a distributed structure to make sure energy savings in the visual WDM backbone that transfer the traffic among nodes and numerous computational grids dependent upon anycast routing.

Dong et al [25] reduce the power utilization within the data center with three main issues. Primarily, via Linear Programming (LP) models and identify the best possible location of a data center. Second, reduce power utilization via the utilization of an LP model. Third examine via LP the issue of whether to situate data centers subsequent to renewable energy underneath diverse traffic situations and diverse network components' power utilization. Dong et al [26] examine presenting renewable energy to IP over WDM networks with data centers and assess the issue of whether to situate data centers subsequent to renewable energy or to send out renewable energy to data centres. The outcomes prove that enhancing the data centre sites underneath the multi-hop bypass heuristic with renewable energy could bring about 71% of the power utilization.

3. Proposed Methodology

This research primarily assesses the sub-optimality of ToURP meta-heuristic underneath small scale situations. At primarily Holistic Model is presented to the VMs placement and immigration restraints. At this point, the sub-optimality is carried out dependent upon the outcomes acquired from the resolution of the MILP problem formulation model. One time the ToURP awareness based Multi-Objective Firefly Simulated Annealing (MOFSA) based meta-heuristic technique that intends at getting a DC workload sharing map that ToURP minimizes the electricity charges for the data center operator and the network operator.

3.1 ToU-Aware Resource provisioning in Cloud

In this research take an incorporated architecture comprising the inter-data center network as well as the Smart Grid Communication Network (SGCN) [27]. The inter-data center network comprises an IP over WDM transport network in addition to the data centers related in the company of the backbone nodes. ToU-Resource aware Provisioning (ToRUP) accepts the Opex minimization technique. ToRUP is dependent upon the subsequent suppositions. All upstream DC, downstream DC, IDC and other demands are routed over an IP over WDM network comprising N nodes applying visual bypass technique [28]. Every backbone node is related with a data center, i.e., data center-'i' is related with node-i. Overhead of the access routers is not taken into account. Dependent upon the forecasted demand, a novel virtual backbone topology is developed to the physical topology. Resource kinds are additionally expanded into a vector of Resource provides that forms separate hardware parts.

3.2 An MILP Model for ToRUP

The MILP model for ToRUP intends to increase the Opex of the network and data center operators as denoted by the objective function in Equation (1).

$$\text{Minimize } \sum_{i \in DC} Op_n^i(T) + Op_{dc}^i(T) \quad (1)$$

Here $DC = (dc_1, \dots, dc_n)$ - Amount of data center nodes, $Op_n^i(T)$ - Opex of the network operator for the period T, $Op_{dc}^i(T)$ - Opex of the data center operator for T.

Opex Constraints : Opex of the network operator at node-i is the product of the energy utilization of the network component for the duration of the period T ($T_{end}-T_{start}$), Resource utilization of the network component for the duration of the period T ($T_{end}-T_{start}$) and the electricity charge for the period of T at the place of node-i as depicted in Equation (2). The energy utilization of the network component at node-i is described as the summation of the energy utilized by the IP router ports, transponders and the dynamic Erbium Doped Fiber Amplifiers (EDFAs) arranged on the unidirectional links from node-i to its adjacent.

$$\begin{aligned} Op_n^i(T) &= \sum_{j \in N_{dc}^{dc}} P_r \cdot C_{ij} \cdot EP_i(T) \cdot RU_i(T) (T_{end} \\ &\quad - T_{start}) \\ &\quad + \sum_{j \in N_i^p} (P_r \cdot W_{ij} + S_{ij} P_e \cdot f_{ij}) EP_i(T) (T_{end} \\ &\quad - T_{start}) \\ &\quad + \sum_{j \in N_i^p} (R_r \cdot W_{ij} + S_{ij} R_e) RU_i(T) (T_{end} \\ &\quad - T_{start}) \end{aligned} \quad (2)$$

Here C_{ij} =Capacity of a wavelength channel node-i to node-j, $EP_i(T)$ = ToRUP electricity charge at the place of node-i for the extent of period T, $EP_i(T)$ = ToRUP resource utilization at the place of node-i for the duration of period T N_i^p =node-i in the virtual (physical) topology, P_r = Power utilization of a transponder, W_{ij} = Amount of accessible channels in the virtual link, S_{ij} = Amount of Erbium Doped Fiber Amplifiers (EDFAs) in the physical link-ij, P_e = Power utilization of an EDFA, R_e = Power utilization of an EDFA, f_{ij} = Amount of fibers in the physical link-ij.

$$Op_{dc}^i(T) = Op_1(T) + Op_2(T) + Op_3(T) \quad (3)$$

$$\begin{aligned} Op_1(T) = &\left[\sum_{s \neq i, s, i \in DC} (\varphi_s^i + h(l_s, k_s^i)) \right. \\ &\quad \left. - h(l_i, k_i^s) \right] EP_i(T) \cdot (T_{end} \\ &\quad - T_{start}) \forall i \in DC \end{aligned} \quad (4)$$

$$\begin{aligned} Op_2(T) = &\left[\sum_{s \neq i, s, i \in DC} (\alpha_s^i + h(m_s, m_s^i)) \right. \\ &\quad \left. - h(m_i, m_i^s) \right] MU_i(T) \cdot (T_{end} \\ &\quad - T_{start}) \forall i \in DC \end{aligned} \quad (5)$$

$$\begin{aligned} Op_3(T) = &\left[\sum_{s \neq i, s, i \in DC} (\beta_s^i + h(r_s, r_s^i)) \right. \\ &\quad \left. - h(r_i, r_i^s) \right] RU_i(T) \cdot (T_{end} \\ &\quad - T_{start}) \forall i \in DC \end{aligned} \quad (6)$$

φ_s^i - Added cooling and processing power utilization presented to data center at node-i because of locating the prerequisites of the users at node-s, α_s^i - Memory utilization presented to data center at node-i owing to locating the insists of the users at node-s, β_s^i - Resource utilization presented to data center at node-i because of locating the insists of the users at node-s, $h(.)$ = function which gives the parameters(memory, resource , energy) presented to a data center because of workload immigration from

other data centers, l_i =primary Load at node ‘i’, m_i =preliminary Memory at node ‘i’, r_i =primary resource at node ‘i’, ℓ_{MAX} =Maximum load of a data center, m_{MAX} =Maximum Memory of a data center, r_{MAX} =Maximum Resource of a data center, k_s^i = Proportion of the workload immigrated from data centers to data center-i to the primary load of the data center-s for the duration of present virtualization period, m_s^i = Proportion of the memory utilized immigrated from data centers to data center-i to the preliminary memory of the data center-s for the duration of current virtualization period, r_s^i = Proportion of the resource consumed immigrated from data centers to data center-i to the preliminary memory of the data center-s for the time of present virtualization period. Equation (3) creates Opex of the data center operator for the data center related to node-I, the resource utilization of the data centers time interval (T), memory utilization of the data centers at time interval (T).

IDC Constraints: IDC traffic from data center-s to data center d uses the virtual link-ij, and afterward the insist on the equivalent link is anticipated to be a function of the immigrated workload proportion from data centers to data center-d pass through that link. On the other hand, this multiplication brings about non-linearity in the MILP formulation.

IP Layer Flow Conservation Constraints: MILP creates a flow management restraint at the source node for upstream DC and IDC insists. The more aspects explanation of these restrictions is conversed in the current work [29]. The restraints in Eqs (7-10) guarantee flow management in the IP layer for normal traffic among the core nodes and the downstream DC traffic.

$$\sum_{j \neq d, j \in DC} (\Omega_{ds}^{DOWN} \cdot Y_{jd\ down}^{ds} + \lambda_{jd}^{sd}) \quad (7)$$

$$- \sum_{j \neq d, j \in DC} (\Omega_{ds}^{DOWN} \cdot Y_{jd\ down}^{ds} + \lambda_{dj}^{sd}) = |\Lambda_{sd} + \Omega_{ds}^{DOWN}|, \forall V, s \neq d \quad (8)$$

$$\sum_{j \neq s, j \in DC} (Y_{down}^{ds} \cdot \gamma_{ij\ down}^{ds} + \lambda_{jj}^{sd}) \quad (9)$$

$$- \sum_{j \neq s, j \in DC} (Y_{down}^{ds} \cdot \gamma_{ij\ down}^{ds} + \lambda_{jj}^{sd}) = 0, \forall V, (s, d, i) \in DC \quad (10)$$

Y_{down}^{ds} = Size of the insist from data center-s to node-d, Ω_{ds}^{DOWN} =Downstream demand from DC-d to node-s, Λ_{sd} = Non-DC demand size from node-s to node-d, λ_{jj}^{sd} = Binary variable is one among the

normal demand from node-s to node-d uses the virtual link-ij. Correspondingly the restraints in Eqs (11-14) make sure flow management in the IP layer with resource utilization for usual traffic among the core nodes and the downstream DC traffic.

$$\sum_{j \neq d, j \in DC} (\Psi_{ds}^{DOWN} \cdot \delta_{jd\ down}^{ds} + \lambda_{jd}^{sd}) \quad (11)$$

$$- \sum_{j \neq d, j \in DC} (\Psi_{ds}^{DOWN} \cdot \delta_{jd\ down}^{ds} + \lambda_{dj}^{sd}) = |\Lambda_{sd} + \Psi_{ds}^{DOWN}|, \forall V, s \neq d \quad (12)$$

$$\sum_{j \neq s, j \in DC} (Y_{down}^{ds} \cdot \delta_{ij\ down}^{ds} + \lambda_{jj}^{sd}) \quad (13)$$

$$- \sum_{j \neq s, j \in DC} (Y_{down}^{ds} \cdot \delta_{ij\ down}^{ds} + \lambda_{jj}^{sd}) = 0, \forall V, (s, d, i) \in DC \quad (14)$$

Y_{down}^{ds} = Size of the resource requirement demand from data center-s to node-d, Ψ_{ds}^{DOWN} = Resource requisite downstream insist from DC-d to node-s, Λ_{sd} = Non-DC demand size from node-s to node-d, $\delta_{jd\ down}^{ds}$ = Binary variable is one among IDC resource prerequisite insist from data center-s to data center-d uses virtual link-ij.

Manycast Constraints for Upstream DC and IDC Demands:

Attaining adequate amount of targets by an upstream DC insist is guaranteed by Eqs (15)-(16). Eq (17) guarantees that an upstream DC demand could use not more than one virtual link previous to attaining the target node d. With the purpose of assisting many casting, the nodes have to contain multicast ability as created by Eq. (18). Consequently, an upstream DC demand could be performed over the identical virtual links up to node-j, here the insist is divided into several virtual links.

$$\sum_{i \in DC} \sum_{i \neq d} Y_{up}^{sd} \cdot \gamma_{jd\ up}^{sd} \leq D_{max}^s \cdot \Omega_{ds}^{UP}, \forall s, d \in DC \quad (15)$$

$$\sum_{i \in DC} \sum_{i \neq d} Y_{up}^{sd} \cdot \gamma_{jd\ up}^{sd} \geq D_{min}^s \cdot \Omega_{ds}^{UP}, \forall s, d \in DC \quad (16)$$

$$\sum_{i \in DC} \sum_{i \neq d} \gamma_{jd\ up}^{sd} \leq 1, \forall s, d \in DC \quad (17)$$

$$\sum_{i \in DC} \gamma_{ij\ up}^{sd} \leq 1, \forall s, i, j \in DC \quad (18)$$

Attaining adequate amount of destinations by an upstream DC demand with resource requisite is guaranteed by Eqs. (19)-(22). Eq (21) guarantees that an upstream DC resource requisite demand could make use of not more than one virtual link previous to attaining the target node d. To make possible many

casting, the nodes ought to have multicast ability as denoted by Eq (22).

$$\sum_{i \in DC} \sum_{i \neq d} Y_{up}^{sd} \cdot \delta_{jd_{up}}^{sd} \leq D_{max}^s \cdot \Psi_{ds}^{UP}, \forall s, d \in DC \quad (19)$$

$$\sum_{i \in DC} \sum_{i \neq d} Y_{up}^{sd} \cdot \delta_{jd_{up}}^{sd} \geq D_{min}^s \cdot \Psi_{ds}^{UP}, \forall s, d \in DC \quad (20)$$

$$\sum_{i \in DC} \sum_{i \neq d} \delta_{jd_{up}}^{sd} \leq 1, \forall s, d \in DC \quad (21)$$

$$\sum_{i \in DC} \delta_{jd_{ij_{up}}}^{sd} \leq 1, \forall s, i, j \in DC \quad (22)$$

Flow Conservation Constraint in the Optical Layer: Eq (23)-(24) indicate the flow management restraints in the ocular layer. Eq (23) guarantees the subsequent three states: i) the source node of a virtual link doesn't contain any inward wavelength channels; ii) The target node of a virtual link doesn't contain any departing wavelength channels. iii) A virtual link doesn't have any loops. Eq (24) guarantees the amount of lightpaths pass through a physical link_{mn} is restricted by the whole channel capacity of the fibers amid node-m and node-n.

$$W_{mn}^{ij} - W_{nm}^{ij} = \begin{cases} -C_{ij} & m = i \\ C_{ij} & m = j \\ 0 & else \end{cases}, \forall m, n, i, j, \in DC \quad (23)$$

$$\sum_{i \in DC} \sum_{j \in DC} W_{mn}^{ij} - W \cdot f_{mn} \leq 0, \forall m, n \in DC \quad (24)$$

Cyber-Physical Constraints: In Eq. (25), the probable power utilization overhead of the upstream demand of node-s to data center-d is created by presuming the latent amplify in the workload of the data center.

$$\Phi_s^d = \sum_{j \neq d, j \in DC}^{DC} \Theta_{s,d} \cdot \gamma_{jd_{up}}^{sd} + \sum_{j \neq d, j \in DC}^{DC} X_{s,d} \cdot \delta_{jd_{up}}^{sd} + \sum_{j \neq d, j \in DC}^{DC} R_{s,d} \cdot \gamma_{jd_{up}}^{sd} \quad (25)$$

$\Theta_{s,d}$ indicates the possible power consumption overhead that will be presented to data center 'd' if the insists inward at node-s are located in data center-'d', $X_{s,d}$ indicates the memory utilization overhead that will be provided to data center 'd', - $R_{s,d}$ indicates

the memory consumption overhead that will be presented to data center 'd', Therefore, in case any inward link of noded is used by the upstream demand started at node-s, power utilization overhead of the upstream data center insist of node-s to data center-d is $\Theta_{s,d}$, or else it is zero since the data center was not chosen among the set of candidate destinations

3.3 Holistic Approach

The above stated restraints for instance the memory, power and resources are developed by the holistic approach are talked about along these lines. The research model [30] is compiled of the three major dimensions which stand for system resources, cloud applications, and communication fabric. The resource kinds are additionally expanded into a vector of Resource Supplies that models separate hardware parts. Capacities are utilized to signify abilities of an element. For instance, a quad-core Intel CPU is a resource supply which is owned by the Computing resource sort, with a Capacity vector comprising four components which signify the cores of the processor [31].

The facet of cloud applications explains an influence of the set up software stack. This facet is additionally split into three levels: Cloud Applications, VMs, and Physical Nodes. The Cloud Applications layer represents the features of the load of the data center; the VMs layer explains the organization and the pattern of the virtual machines hosted by the servers. Lastly, the Physical Nodes layer explains hardware characteristics of the servers.

The Resource Provider is a unit which creates its resources accessible to the Resource Consumer. The method of resource allocation is in the allocating of the Resources accessible at the Resource Providers with the Resource insists of the Resource Consumers. A suitable allotment ought to meet up two fundamental states: the demands could not go beyond capability of the resource provider and ought to not be allotted to a stipulation with a lower structural design. The offered abstraction permits making a random amount of intermediate layers which play as Resource Providers and Resource Consumers together. The last layer is created by physical nodes. The intermediate layer contains VMs.

The VMs could preserve resources of the physical hosts and let them accessible for task calculation. Lastly, cloud applications are implemented at the uppermost layer. They are concluding Resource Consumers denoting the datacenter workload. Jobs in a system are collected together into cloud applications which stand for running services. Correspondingly, every VM contains a kind which stands for its performance. Every cloud application contains a

definite set of VM kinds that could operate it. Lastly, every task is described with input and output sizes which ought to be passed over the network correspondingly previous and subsequent to the task execution.

Multiple Resource Types: Data center servers could comprise several kinds of resources for instance Central Processing Unit (CPU), memory, storage, and networking. Another dispute for overtly mapping different kinds of resources is related with management interfaces of recent hypervisors and cloud management systems that assign and merge VMs in line with their insists and accessible resources of physical servers. Resource is as well a super-class of the subsequent three classes: Resource Spec, Dc Resource, and Res Demand. The Resource Spec class illustrates hardware requirement. It as well piles up the distinctive name of the hardware part kind and a vector of power conditions accessible for the part signified with integer numbers. Dc Resource is a class utilized to signify instances of Resource Spec that are aimed at components of servers. It stocks up the pointer to the Resource Spec that describes its activities. Additionally, an integer field denotes a power state of the resource. The CPU class is inherited from the Dc Resource class and broadens its functionality by organizing listings of jobs implemented and allocated to a CPU. Lastly, Res Demand class, executed for the demand side, denotes resource needs. For Res Demand, the capability parameter describes the prerequisite for the resource or the number of the untreated workload.

3.4 MOFSA -Based Meta-Heuristic for ToRUP

So as to guarantee polynomial time resolution for the enhancement model stated above, present a MOFSA based meta- heuristic that attains an IDC traffic matrix in the initial step, and in the next step rearranges the inter-data center network to stipulation every kind of demands in a ToRUP way.

Step-1 - IDC Workload Migration

The initial step is an MOFSA based technique. The notation Map is the major workload distribution matrix while TempMap indicates an impermanent workload distribution matrix. O_{temp} and $O_{current}$ point to the primary Opex and present Opex, likewise while $Op_{dc}^i(T)$ is the Opex of data center-i. T , B and t_{cool} are the MOFSA associated settings representing the annealing temperature, Boltzmann constant and cooling rate of the system, correspondingly. In this research, a hybrid meta-heuristic technique dependent upon the Simulated Annealing (SA) algorithm and

Firefly Algorithm (FA) is proposed that is known as ‘MOFSA’ briefly. With the purpose of acquiring superior and more vigorous resolutions, the Taguchi technique is utilized for calibration the parameters of this method.

Encoding Scheme: An appropriate encoding method that is pinpointing the features of workload immigration contains substantial persuade on the functioning of a meta-heuristic methodology.

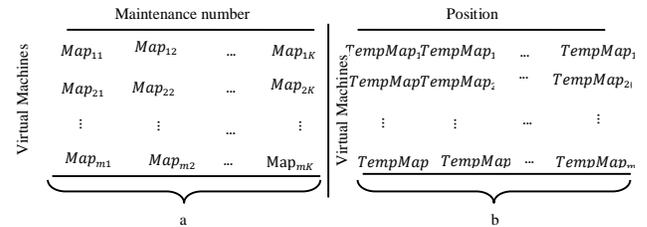


Figure. 1 Encoding scheme of the solution

This comprises two part encoding method was utilized with the intention of revealing a workload immigration. This process has been demonstrated in Figure 1. Section (a) explains the transverse of the workload immigration matrix (Map) is 100 by a Map_{ij} matrix. Section (b) indicates the TempMap represents a provisional workload distribution matrix that is specified by a $TempMap_{ij}$ matrix. Every row of this matrix reveals the series of workload on the machine equivalent to that row. Every column of this matrix reveals the amount of machines in the workload distribution.

Initialization: At first, every cell of the diagonal of the workload migration matrix (Map) is 100 (that is to say $Map_{ij}=100$) representing that every data center at first maintains 100% of its workload within itself. For IDC workload sharing, $\{dest1, dest2\}$ denotes the list of target data centers, and its size is restricted to two on behalf of simplicity, whilst κ_{MIN} signifies the lower bound for the proportion of workload which should be hosted in the present data center. After that, the left shifting operator alters and regulates the outcomes of O_{temp} and $O_{current}$ get superior solutions.

Neighborhood Search Based Workload Structure: To set up the neighborhood of first search structure, the move operator in the firefly algorithm is utilized. In addition, to set up the neighborhood of second search structure, the three general operators of swap, insert, and inverse are taken that are altered relating to the problem states. The O_{temp} and $O_{current}$ to attain healthier solutions in the novel solution are progressed by using the left-shifting operator. In the FA, the progress of firefly (IDC) ‘i’ in the direction of the more

gorgeous (brighter) firefly j (IDC) is identified via the subsequent relation [32]:

$$x_i = x_i + \beta_0 e^{-\gamma r_{ij}^2} (X_j - X_i) + \alpha \left(rand - \frac{1}{2} \right) \quad (26)$$

$\beta_0 e^{-\gamma r_{ij}^2}$ is called the magnetism function whose value reduces with rising distance amid two fireflies (r_{ij}). β_0 is the attractiveness at $r_{ij}^2 = 0$, and γ is the fixed light absorption coefficient in the atmosphere. Expression a ‘rand’ is for the randomization function which produces random numbers with uniform distribution in the [0,1] interval. By utilizing the Cartesian distance, the distance amid the two fireflies of i and j is acquired from the subsequent relation [32]:

$$r_{ij} = \left\| x_i - x_j \right\| = \sqrt{\sum_{k=1}^d (x_{ik} - x_{jk})^2} \quad (27)$$

Here x_{ik} is the k^{th} component of the i^{th} firefly. From the distance amid two machines temporary distribution matrix, TempMap is determined. One time the temporary distribution matrix, TempMap, a novel Opex value is computed by Eqs (1-6). If this Opex value is below the present Opex (i.e., $O_{\text{temp}} < O_{\text{current}}$ and $F \geq 1$), the novel placement is believed, and the primary Opex (O_{current}) is brought up to date to be equal to the present transitory Opex (O_{temp}). Or else (i.e., $O_{\text{temp}} \geq O_{\text{current}}$ and $F \in [0, 1]$), a random number, r , is produced in the interval [0,1). If $r < F$, the novel placement is agreed, and the present Opex is maintained consequently.

Swap, Insert, and Inverse Operators: For this reason, the typical insert and inverse operators are customized. In every operator, primarily, one IDC is arbitrarily chosen with their workload migration matrix. In the swap operator, two places are selected arbitrarily, and the places of the two available data centers workloads are substituted with one another in every data centers. In the insert operator, one migration job and one location are arbitrarily chosen, and the chosen job is inserted in the chosen place in all IDC. In the inverse operator, two points of the series are selected arbitrarily, and the datacenter places amid these two jobs are inverted in every IDC. Or else, if $r \geq F$, the solution is discarded. At every iteration, the algorithm chooses a row and a column arbitrarily from the migration matrix, Map, and provisionally sets the chosen cell at zero if not the consequent cell is on the diagonal. In case the chosen cell is situated on the diagonal, the algorithm provisionally sets its value to κ_{MIN} . After that, the changed value of the cell is dispersed amongst $DC_{\text{MAX}} = 2$ cells and the consequent cell. In the last step, the temperature is

brought up to date by the cooling rate, t_{cool} , (i.e., $T \leftarrow t_{\text{cool}} \times T$). SA converges in case either of the subsequent two states holds: *i*) For a temperature T , solutions do not get better for a sensible amount of iterations, and *ii*) Annealing temperature is equivalent to under the ground temperature ($T \leq T_{\text{ground}}$), i.e., the frozen condition.

Step-2 - ToRUP-Aware Backbone Virtualization

Routing the IDC and upstream DC demands, ToURP orders the candidate data centers dependent upon the subsequent two criterion: *i*) Minimum Opex, *ii*) Minimum path delay. ToURP chooses $\lceil DC_{\text{min}}/2 \rceil$ of the target data centers dependent upon the minimum Opex ranking, DC_{min} indicates the minimum amount of data centers to be attained so as to provision into numerous unicast demands. In relation to the cost assignment function, ToRUP intends to choosing the virtual links which are created by low physical cost ($Cost_{\text{mn}}^{\text{phy}}(T)$) and have accessible virtual link capabilities.

The physical link price comprises the subsequent three elements: *i*) ToRUP charges at the consequent position of node- n for the period of T , *ii*) Power utilization of the network element along the link, *iii*) Length of the fiber-link. In the equation, W_{mn} represents the residual wavelength capability in the physical link $_{\text{mn}}$. In relation to the physical price assignment function, ToRUP intends to choosing the links with short length, low power utilization, resource, memory and which are aimed at the nodes having less ToRUP costs.

$$Cost_{ij}^{dc}(T) = \begin{cases} \sum_{link_{mn} \in link_{ij}} Cost_{mn}^{phy}(T) & C'_{ij} > 0 \\ \infty & else \end{cases} \quad (2)$$

$$Cost_{mn}^{phy}(T) = \begin{cases} EP_i(T)RU_i(T) \left[\begin{matrix} Lf_{mn} \cdot (P_r \cdot W_{mn} \cdot S_{mn} + P_e \cdot V) \\ (R_r \cdot W_{mn} \cdot S_{mn} + R_e \cdot W_m) \end{matrix} \right] & 9 \end{cases} \quad (2)$$

In the equation link- $_{ij}$ represents a virtual link that is created by a set of physical links represented by link- $_{mn}$.

4. Numerical Results

Afterward assess the performance of ToRUP in relation to on the whole Opex in the Cloud, Opex of the data center and network operators, power utilization and path delay of every demand kinds. An IP over WDM network is treated as the transmission medium underneath two topologies, that is to say the NSFNET [33] with 14 nodes and 21 links, as well as the European Optical backbone Network (EON) [34]

that contains 28 nodes and 41 links. Four diverse demand kinds for instance Downstream Data Center (DC) demands, Upstream DC demands, Inter-Data Center (IDC) demands and Non-DC demands are measured in this part.

4.1 MILP-based model vs. ToRUP

In this part, confirm ToRUP by contrasting the heuristic and meta- heuristic outcomes to the outcomes attained underneath the MILP. In these test set-up, to get outcomes in reasonable time, set $K_{MIN} = 1$ in the MILP and the ToRUP heuristic with the intention that the IDC restraints are removed.

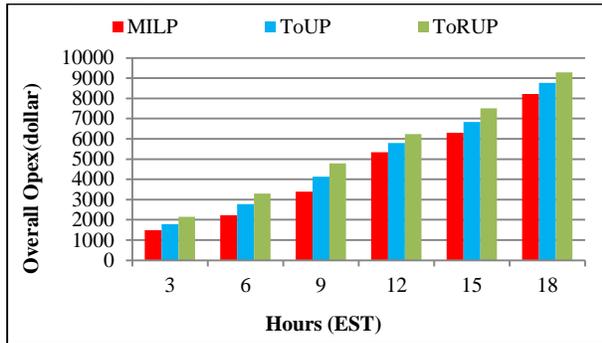


Figure. 2 Cumulative Opex comparison vs. models

As depicted in the figure 2, ToRUP shows better performance at Opex when compared to ToUP and MILP formulation. From the table 1 it shows that the Overall Opex value of the proposed ToRUP is 9285 dollars for estimation time of 18 hrs, whereas the other ToUP and MILP methods are 8763 and 8214 dollars respectively. It concludes that the proposed system has higher overall Opex cost than existing system.

Table 1. Overall Opex comparison vs. models

Hours(EST)	Overall Opex(Dollars)		
	MILP	ToUP	ToRUP
3	1487	1789	2156
6	2235	2781	3289
9	3389	4125	4792
12	5341	5798	6235
15	6293	6827	7518
18	8214	8763	9285

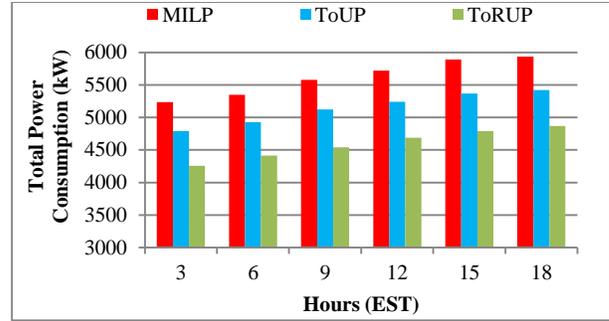


Figure. 3 Power consumption comparison vs. models

Figure 3 demonstrates the power consumption results of ToRUP, ToUP and the MILP .It concludes that the power utilization results of ToRUP is lesser when compared to MILP formulation and ToUP. From the table 2 it shows that the power consumption value of the proposed ToRUP is 4871 kW for estimation time of 18 hrs, whereas the power consumption value of other ToUP and MILP methods are 5421 kW and 5935 kW respectively. It concludes that the proposed system has lesser power consumption than existing methods.

Table 2. Total Power consumption vs. models

Hours(EST)	Total Power consumption(kW)		
	MILP	ToUP	ToRUP
3	5234	4789	4258
6	5348	4927	4415
9	5581	5124	4543
12	5721	5243	4689
15	5894	5369	4792
18	5935	5421	4871

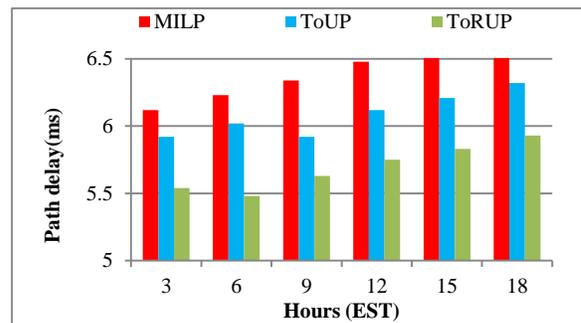


Figure. 4 Average path delay comparison vs. models

In Figure 4 illustrates the Average path delay comparison results of proposed ToRUP and existing ToUP and MILP methods. From the table 3 it demonstrates that the average path delay value of the presented ToRUP is 5.93 ms for estimation time of 18 hrs, whereas the power consumption value of other

ToUP and MILP methods are 6.32 ms and 6.75 ms respectively. It concludes that the proposed system has lesser power consumption than existing methods.

Table 3. Average path delay consumption vs. models

Hours(EST)	Average path delay(ms)		
	MILP	ToUP	ToRUP
3	6.12	5.92	5.54
6	6.23	6.02	5.48
9	6.34	5.92	5.63
12	6.48	6.12	5.75
15	6.64	6.21	5.83
18	6.75	6.32	5.93

5. Conclusion and Future Work

In this research, we analyzed the Time-of-Use resource provisioning (ToRUP) - aware design of a virtual IDC network to reduce the electricity bills of the operators and decrease the energy utilization all through the CC Environment. On the other hand the ToURP awareness based Multi-Objective Firefly Simulated Annealing (MOFSA) based meta-heuristic algorithm that intends to get a DC workload sharing map. Primarily, we propose a MILP formulation to form the best possible conduct of ToURP. On the other hand additionally the traffic amid two backbone nodes, upstream user demands intended to data centers and downstream data center insists beginning from a lot of data centers. Power consumption value of the proposed ToRUP is 4871 kW for estimation time of 18 hrs, whereas the power consumption value of other ToUP and MILP methods are 5421 kW and 5935 kW respectively. Furthermore, we have proved that ToRUP brings about additional power utilization savings in the Cloud while every data center is forced to maintain some quantity of its original workload. The upcoming work encloses modeling and designing VM immigration procedures, developing new task and VM scheduling techniques for the model, and enhancements of the models implementation.

References

- [1] M.T. Khorshed, A.S. Ali, and S.A. Wasimi, "A survey on gaps, threat remediation challenges and some thoughts for proactive attack detection in cloud computing", *Future Generation computer systems*, Vol.28, No.6, pp. 833-851, 2012.
- [2] J.W. Lin, C.H. Chen, and J.M. Chang, "QoS-Aware Data Replication for Data-Intensive Applications in Cloud Computing Systems",

- IEEE Transactions on Cloud Computing*, Vol. 1/1, pp. 101–115, 2013.
- [3] L. Parolini, B. Sinopoli, B.H. Krogh, and Z. Wang, "A Cyber-Physical Systems Approach to Data Center Modeling and Control for Energy Efficiency," *Proceedings of the IEEE*, Vol. 100/1, pp. 254–268, 2012.
- [4] Y. Guo, Z. Ding, Y. Fang, and D. Wu, "Cutting down electricity cost in internet data centers by using energy storage", *IEEE Global Telecomm. Conf. (GLOBECOM)*, pp. 1–5, 2011.
- [5] J. Baliga, R.W.A. Ayre, K. Hinton, and R.S. Tucker, "Green Cloud Computing: Balancing Energy in Processing, Storage, and Transport," *Proceedings of the IEEE*, Vol. 99/1, pp. 149–167, 2011.
- [6] A. Beloglazov, J. Abawajy, and R. Buyya, "Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing", *Future Generation Computer Systems*, Vol.28, No.5, pp. 755–768, 2012.
- [7] H. Viswanathan, E. K. Lee, and D. Pompili, "Self-organizing sensing infrastructure for autonomic management of green datacenters", *IEEE Network*, Vol. 25/4, pp. 34–40, 2011.
- [8] B. Kantarci and H.T. Mouftah, "Designing an energy-efficient cloud network", *IEEE Journal of Optical Communications and Networking*, Vol.4/11, pp. B101–B113, 2012.
- [9] A. Gandhi, M. Harchol-Balter, R. Das, and C. Lefurgy, "Optimal power allocation in server farms. In *ACM SIGMETRICS Performance Evaluation Review*, Vol.37, No.1, pp. 157-168, 2009.
- [10] G. Chen, W. He, J. Liu, S. Nath, L. Rigas, L. Xiao, and F. Zhao, "Energy-Aware Server Provisioning and Load Dispatching for Connection-Intensive Internet Services", *In NSDI*, Vol.8, pp. 337-350, 2008.
- [11] A. Beloglazov, and R. Buyya, "Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers", *Concurrency and Computation: Practice and Experience*, Vol.24, No.13, pp.1397-1420, 2012.
- [12] S. Srikantaiah, A. Kansal, and F. Zhao, "Energy aware consolidation for cloud computing", *Conference on Power aware computing and systems*, Vol.10, pp. 1-5, 2008.

- [13] N. Sharma, S. Barker, D. Irwin and P. Shenoy, "Blink: managing server clusters on intermittent power", *In ACM SIGPLAN Notices*, Vol.46, No.3, pp.185-198, 2011.
- [14] L. Rao, X. Liu, L. Xie, and W. Liu, "Minimizing electricity cost: Optimization of distributed internet data centers in a multi-electricity-market environment", *Proceedings IEEE INFOCOM*, pp.1-9, 2010.
- [15] H. Alcott, "Real-time pricing and electricity market design", *Working Paper*, 2013.
- [16] E. Celebi and J.D. Fuller, "Time-of-use pricing in electricity markets under different market structures", *IEEE Transaction on Power Systems*, Vol.27, No.3, pp. 1170–1181, 2012.
- [17] T. Hahn, Z. Tan, and W. Ko, "Design of time-varying rate considering CO₂ emission," *IEEE Transactions on Smart Grid*, Vol.4, No.1, pp.383-389, 2013.
- [18] B. Kantarci and H.T. Mouftah, "Time of use (ToU)-awareness with inter-data center workload sharing in the cloud backbone," *In IEEE International Conference on Communications (ICC)*, pp.SA–GC04.1, pp.1–6, 2013.
- [19] U. Awada, K. Li, and Y. Shen, "Energy consumption in cloud computing data centers", *International Journal of Cloud Computing and services science*, Vol.3, No.3, pp.145, 2014.
- [20] H. Li, G. Zhu, C. Cui, H. Tang, Y. Dou, and C. He, "Energy-efficient migration and consolidation algorithm of virtual machines in data centers for cloud computing", *Computing*, Vol.98, No.3, pp.303-317, 2016.
- [21] A. Beloglazov, R. Buyya, Y.C. Lee, and A. Zomaya, "A taxonomy and survey of energy-efficient data centers and cloud computing systems", *Advances in computers*, Vol.82, No.2, pp.47-111, 2011
- [22] M.S. Ilyas, S. Raza, C.C. Chen, Z.A. Uzmi, and C.N. Chuah, "RED-BL: Energy solution for loading data centers", *IEEE INFOCOM*, pp. 2866–2870, 2012.
- [23] Y. Zhang, Y. Wang, and X. Wang, "Electricity bill capping for cloud-scale data centers that impact the power markets", *41st International Conference on Parallel Processing (ICPP)*, pp. 440-449, 2012.
- [24] D. Tafani, B. Kantarci, H.T. Mouftah, C. McArdle, and L.P. Barry, "Distributed management of energy-efficient light paths for computational grids", *IEEE Global Communications Conference (GLOBECOM)*, pp. 2924-2929, 2012.
- [25] X. Dong, T. El-Gorashi, and Elmoghani, J.M. "Green IP over WDM networks with data centers", *Journal of Lightwave Technology*, Vol.29, No.12, pp.1861-1880, 2011.
- [26] X. Dong, T. El-Gorashi, and J.M.H. Elmoghani, "Use of renewable energy in an IP over WDM network with data centres", *IET optoelectronics*, Vol.6, No.4, pp.155-164, 2012.
- [27] X. Lu, W. Wang, and J. Ma, "An empirical study of communication infrastructures towards the smart grid: Design, implementation, and evaluation," *IEEE Trans. on Smart Grid*, Vol. 4/1, pp.170–183, 2013.
- [28] G. Shen and R.S. Tucker, "Energy-minimized design for IP over WDM networks", *IEEE/OSA J. of Optical Communications and Networking*, Vol. 1, pp. 176–186, 2009.
- [29] B. Kantarci, and H.T. Mouftah. "Inter-data center network dimensioning under time-of-use pricing. *IEEE Transactions on Cloud Computing*, Vol.4, No.4, pp. 402-414, 2016.
- [30] M. Guzek, D. Kliazovich, and P. Bouvry, "A holistic model for resource representation in virtualized cloud computing data centers", *International Conference on Cloud Computing Technology and Science (CloudCom)*, Vol.1, pp. 590-598, 2013.
- [31] M. Guzek, S. Varrette, V. Plugaru, J.E. Sanchez, and P. Bouvry, "A Holistic Model of the Performance and the Energy-Efficiency of Hypervisors in an HPC Environment", *in Proc. of the Intl. Conf. on Energy Efficiency in Large Scale Distributed Systems (EE-LSDS)*, 2013.
- [32] X.S. Yang, "Firefly algorithms for multimodal optimization", *In International symposium on stochastic algorithms*, pp. 169-178, 2009.
- [33] B. Kantarci, and H.T. Mouftah, "Energy-efficient cloud services over wavelength-routed optical transport networks", *In Global Telecommunications Conference*, pp. 1-5, 2011.
- [34] S. De Maesschalck, D. Colle, I. Lievens, M. Pickavet, P. Demeester, C. Mauz, and J. Derkacz, "Pan-European optical transport networks a availability-based comparison", *Photonic Network Communications*, Vol.5, No.3, pp. 203-225, 2003.