Design of energy-efficient cloud systems via network and resource virtualization

Burak Kantarci,1 Luca Foschini,2 Antonio Corradi2 and Hussein T. Mouftah1,∗†

1School of Electrical Engineering and Computer Science, University of Ottawa, ON, Ottawa, Canada
2Dipartimento di Informatica—Scienza e Ingegneria, University of Bologna, Bologna, Italy

SUMMARY

Data centers play a crucial role in the delivery of cloud services by enabling on-demand access to the shared resources such as software, platform and infrastructure. Virtual machine (VM) allocation is one of the challenging tasks in data center management since user requirements, typically expressed as service-level agreements, have to be met with the minimum operational expenditure. Despite their huge processing and storage facilities, data centers are among the major contributors to greenhouse gas emissions of IT services. In this paper, we propose a holistic approach for a large-scale cloud system where the cloud services are provisioned by several data centers interconnected over the backbone network. Leveraging the possibility to virtualize the backbone topology in order to bypass IP routers, which are major power consumers in the core network, we propose a mixed integer linear programming (MILP) formulation for VM placement that aims at minimizing both power consumption at the virtualized backbone network and resource usage inside data centers. Since the general holistic MILP formulation requires heavy and long-running computations, we partition the problem into two sub-problems, namely, intra and inter-data center VM placement. In addition, for the inter-data center VM placement, we also propose a heuristic to solve the virtualized backbone topology reconfiguration computation in reasonable time. We thoroughly assessed the performance of our proposed solution, comparing it with another notable MILP proposal in the literature; collected experimental results show the benefit of the proposed management scheme in terms of power consumption, resource utilization and fairness for medium size data centers. Copyright © 2013 John Wiley & Sons, Ltd.

Received 21 March 2013; Accepted 12 July 2013

1. INTRODUCTION

Cloud computing architectures have gained increasing attention in recent years, and several vendors are looking at them for feasible solutions to optimize the usage of their own infrastructures [1,2]. Among several advantages, these solutions offer pools of virtualized computing resources, paid on a pay-per-use basis, so the initial investment and maintenance costs are drastically reduced.

At the current stage, several management issues still deserve additional investigations, especially in cloud architectures for large-scale geographically distributed data centers that introduce high economical investments for the cloud owner. Above all, the power required by both IT resources (e.g. servers and network elements) and non-IT equipment (e.g. cooling systems and uninterrupted power supplies) introduces a significant economical overhead. Recent figures estimate that the ICT as a whole makes a direct contribution to global CO₂ emissions that exceed 2% [3].

Hence, also driven by the emerging green computing research, cloud providers are looking for efficient techniques that dynamically reconfigure the IT infrastructure to reduce the total power consumption of data centers. Toward this goal, careful usage of virtualization techniques attempts to reduce power consumption by consolidating the execution of multiple virtual machines (VMs) on

*Correspondence to: Hussein T. Mouftah, School of Electrical Engineering and Computer Science, University of Ottawa, ON, Ottawa, Canada.
†E-mail: mouftah@site.uottawa.ca

Copyright © 2013 John Wiley & Sons, Ltd.
the same physical host through the introduction of appropriate placement functions in charge of detail-
ing the real VM-to-server mappings, namely, VM placement.
At the same time, only a few proposals [4,5] explicitly considered the effects of network requirements
and constraints on VM placement. In fact, moving the services into the cloud will increase the associated
transport energy and introduces several challenging issues [6]. First, VM consolidation is difficult because
it has to balance the exploitation of available resources and the avoidance of possible performance
degradation due to excessive resource consolidation [4,7,8]. In addition, several existing proposals and
algorithms addressing the energy optimization problem typically focus only either on VM placement
on a single data center, namely intra-data center management [9–12], or on the power saving problem
across different data centers, namely inter-data center management (without considering intra-data center
VM placement issues) [13–15]. Finally, although a few interesting emerging solutions are proposing
cooling and thermal-aware job scheduling to enhance power savings of the data centers [16–18], VM
placement solutions able to jointly increase both energy efficiency of data centers and the transport
medium connecting them are still widely unexplored.

The paper addresses all the above issues by proposing a novel solution that exhibits three main
original contributions. First, we introduce a new holistic mixed integer linear programming (MILP)
problem to calculate the best inter and intra-data center VM placement strategy in large-scale cloud
systems, which has been briefly introduced previously [19]. In particular, our problem considers
constraints on different resources, namely, computing (such as CPU and memory), energy and
networking resources, at both intra and inter-data center scales with the final goal of optimizing also
the energy versus network delay trade-off over the core backbone. Second, we assess the effectiveness
of our solution with a widely accepted dataset by benchmarking it with another MILP-based solution in
the literature [20]. Moreover, since the general holistic MILP model may require too long a time to find
a solution, we further partition the problem into two sub-problems: in the first intra-data center step, we
use a subset of the constraints in the MILP model where intra-data center VM placement is formulated;
and then, in the second inter-data center step, we use the output of that MILP formulation to route the
demands towards data centers. Furthermore, since the second inter-data center VM placement step is
still a heavy one, we adopt a previously proposed heuristic to solve it [20], so further lowering overall
computation time. Third, we thoroughly assessed the performance of the proposed solution. Collected
numerical results confirm that our MILP-based two-step solution outperforms the other existing and
benchmarked MIPL-based solution by guaranteeing significant power savings in the cloud network
and by fairly distributing the power consumption among backbone nodes.

The paper is organized as follows. In Section 2 we present the needed background material and discuss
related work. In Section 3 we describe our new model, and in Section 4 we present a comprehensive set of
experimental results to assess its technical soundness. Finally, conclusions and directions of future work
end the paper.

2. RELATED WORK

Cloud power efficiency is a timely hot topic that is receiving increasing attention due to both environmental
and economical issues [1,3]. In the following, without any pretense of being exhaustive for the sake of space
limitations, we present a selection of some important proposals dealing with cloud power efficiency at intra/
inter-data center deployment scales.

Starting with intra-data center solutions, Kim et al. [9] address the problem of cloud resource
provisioning for real-time services. They exploit a priori knowledge of VM workloads to drive
resource allocation, and strive to increase power efficiency as a trade-off between task completion
times and number of powered-on servers. Similarly, given a particular workload, another recent
proposal presents a mathematical model to find the exact number of physical servers to ensure power
efficiency, considering also dynamic CPU frequency scaling to further reduce power consumption
[10]. However, the authors assume that cloud jobs can be partitioned among different physical servers
and that assumption, viable in particular scenarios, does not fit well the general VM allocation
problem. Mistral is a recent proposal that adopts the A*-search technique to optimize both VM
performance and power consumption and to find complex reconfiguration actions [11]. Mistral
considers indicators about both the stability of the next data center configuration in the decision process and the transient costs associated with run-time reconfigurations; the presented experimental results, obtained in a real cloud testbed, make the proposed solution extremely solid. With a similar practical perspective, we have also done several experiments to assess how traffic demands between VMs deployed on the same physical influence host CPU and memory overhead, to avoid excessive overhead due to local communications [21,22]. Finally, Mann et al. [12] focuses on reducing the total energy consumption associated with network elements, namely, powered-on switches and patches. The authors present a new network-aware virtual machine placement (NAVP) problem that strives to consolidate as many traffic demands as possible over the same set of network links, in order to reduce the total data center energy consumption.

Focusing on inter-data center VM placement, job submission in cloud computing has its roots in the grid computing anycast or manycast paradigms. Anycast refers to selecting a single destination from a set of candidate destinations [13], while in manycast a subset of destinations is selected from the set of candidate destination nodes [14]. Focusing on energy-efficient design of the backbone for Internet and data center demands, Dong et al. [13] propose the optimal locations of a limited number of data centers in the transport network and a demand provisioning algorithm to reconfigure the network by virtual topology mapping. In brief, it provisions regular Internet traffic and the downstream traffic originating from the data centers via unicast routing, while for upstream data center demands carrying the job submissions it adopts the anycast paradigm. Along the same direction, Buysse et al. [15] propose an optimization model to obtain energy efficiency switching of the appropriate network elements and serving the demands via anycast routing. An evolutionary algorithm has also been proposed for energy-efficient provisioning of manycast demands over wavelength-routed transport network [23]. Finally, very recently, Kantarci and Mouftah [20] have proposed MILP formulations for energy-efficient cloud network design. In that paper, we use the MILP formulation to schedule the demands on the nearest data centers as a benchmark scheme for our novel intra and inter-data center VM placement energy-efficient scheme.

3. INTER AND INTRA-DATA CENTER ENERGY-EFFICIENT VM PLACEMENT

In this section, we define our MILP formulation to address the energy-efficient VM placement problem in intra and inter-data center management scenarios by balancing computing resources, network delays and energy savings.

3.1. Main design guidelines and system model

We propose a VM placement solution to minimize the transport energy, as well as the energy consumption in the data centers by periodically reconfiguring the virtualized backbone network topology and by optimizing computing resource allocation in the data centers. In our model, cloud computing demands, namely data flows due to needed data traffic exchanges generated from and addressed to VMs placed at a data center, are provided together with the regular Internet demands over an IP-over-wavelength division multiplexing (WDM) network. In IP-over-WDM network scenarios, IP routers are the major power consumers in the backbone network [13]; hence, if we are able to reconfigure the virtualized backbone topology in order to bypass them it is possible to minimize network energy consumption by adopting design approaches such as those proposed elsewhere [13,20] for VM placement in the data centers. It is worth noting that network virtualization is to accommodate the high volume of data transmission between the end-to-end virtual services [24] and support end-to-end quality of service [25]. Furthermore, data center network virtualization has several advantages such as QoS support, ease of deployment and management, as well as resilience against security threats for cloud systems [26]. However, in this paper we limit our focus to the virtualization of the core network.

Our holistic solution makes a step forward by combining the benefits of that minimum energy-aware scheduling between data centers [20] (inter-data center VM placement) and the best (network-aware) VM placement within data centers (intra VM placement) [22,27]. We assume that the overhead of cloud
demands (also referred to as submitted jobs) is known in advance and might be forecast, in terms of power consumption at each data center, in the backbone network; then, based on the forecast cloud demand profile, we reconfigure the virtualized backbone topology to comply with those VM performance objectives in terms of exchanged traffic and needed computing resources.

With a closer view to technical details, the backbone transport medium is assumed to be an IP-over-WDM network where each backbone node is associated with a data center. The upstream traffic initiated from a user traverses access routers and aggregates at the core nodes. Three types of traffic are assumed: (i) upstream data center traffic that is originated at a user and has to be processed at one or more data centers; (ii) downstream data center traffic that is originated from multiple data centers and destined to the end users; and (iii) regular Internet traffic that flows between two backbone nodes.

Each data center is equipped with $H$ physical hosts and $M$ VMs. Each VM $v_m$ has a CPU and a memory capacity of $CPU_{CAP}$ and $MEM_{CAP}$. In order to consolidate the workload of the VMs among physical hosts, each $v_m$ also has a bandwidth capacity of $BW_{CAP}$. Each job submission arrives to the cloud with a predefined memory requirement ($MEM_{REQ}$), a CPU requirement ($CPU_{REQ}$) and a bandwidth requirement ($BW_{REQ}$).

3.2. Holistic VM placement via MILP formulation

As mentioned above, the proposed scheme adopts the cloud network recon configuration scheme [20] and the VM placement approach [22] for the virtualization of the backbone network. The resulting MILP formulation follows the steps shown in Figure 1(a); in particular, we call it ‘holistic’ because it will jointly determine both the data centers and the corresponding VM placement.

Starting from the objective function, equation (1) expresses the power consumption (in watts) associated with node $i$ of the cloud backbone as the sum of three components: (i) power consumption of the associated data center ($DC_i$); (ii) power consumption in the IP layer; (iii) power consumption in the WDM layer; since we want to minimize energy consumption, our main goal is to minimize that sum.

With a closer view to single equation components, $N_i^v$ ($N_i^p$) denotes the set of neighboring nodes of node $i$ in the virtual (physical) topology, while $P_C$ and $C_I$ denote the power consumption of an IP router port and the number of lightpaths on the virtual link $ij$, respectively. Also, $P_r$, $W_{ij}$, $P_{edfa}$ and $f_{ij}$ denote, respectively, the power consumption of a transponder, the number of channels in the physical link $x_{itij}$, the physical distance between node $i$ and node $j$, the power consumption of an erbium-doped fiber amplifier and the number of fibers in the physical link-$ij$:

$$
\min \sum_{i} DC_i + \sum_{N_i^v} P_rC_{ij} + \sum_{j \in N_i^p} (P_rW_{ij} + S_{ij}P_{edfa}f_{ij}), \quad \forall i
$$

Equations (2)–(27), instead, represent either formulations of single components that we need to evaluate (such as equation (2)) or constraints of our optimization model. In the constraint set, $O_{down}$ and $O_{up}$ denote the downstream demand from data center $s$ to node $d$ and the upstream traffic (i.e., job submission) from node $s$ to data center $d$, respectively. Also, $\gamma^{up}$ is a binary variable denoting whether there is downstream traffic from data center $s$ to node $d$ traversing the virtual link $ij$, while $\lambda^{down}$ denotes the regular traffic demand traversing the virtual link-$ij$ and destined from node $s$ to node $d$. $A_{ad}$ identifies the regular traffic demand from node $s$ to node $d$, whereas $\gamma^{sd}$ is the possible demand from node $s$ to data center $d$, and $\gamma^{ad}$ is a binary variable denoting whether there is traffic from node $s$ to data center $d$ traversing the virtual link $ij$. In the manycast constraints, $D_{max}^{up}$ and $D_{min}^{up}$ denote the maximum and minimum number of destinations for the upstream traffic from node $s$. As for the physical properties of the transport medium, $W_{ij}^{mm}$ and $L_{ij}$ denote the number of wavelength channels on the virtual link-$ij$ traversing the physical link $mn$ and the shortest distance from node $i$ to node $j$, respectively. In order to formulate the thermal properties of the data centers, we use the $DC_{cool}$ and $DC_{proc}$ to indicate the cooling power consumed at data center $d$ and the processing power consumed at data center $d$, respectively, whereas $\Theta_{sd}$ is used to denote the power consumption overhead introduced to data center $d$ by the job submitted by node $s$.

Equation (2) formulates the power consumption of data center $d$ upon provisioning the demands from several nodes, where $DC_{cool}$ and $DC_{proc}$ denote the current cooling and
processing powers consumed at data center \(d\), and \(\Theta_{s,d}\) refers to the power consumption overhead of the job submitted by node \(s\) to data center \(d\):

\[
DC_d - \sum_{s \in V} \sum_{i \neq d} \Theta_{s,d} \gamma_{sd} = DC_{d}^{\text{cool}} + DC_{d}^{\text{proc}}, \quad \forall d \in V
\]  

(2)

Equations (3)–(8) present the flow conservation constraints in the IP layer. Equation (4) ensures that the virtual light-tree originating at a source node (node \(s\)) has sufficient number of leaves, i.e., the desired number of destinations are reached, which is between the predefined minimum \((D_{s,\text{min}})\) and the maximum \((D_{s,\text{max}})\) values. Equation (5) formulates the flow conservation constraint of the upstream DC traffic for the intermediate nodes. The constraints in equations (6)–(7) ensure flow conservation in the IP layer for regular traffic between the core nodes and the downstream DC traffic. Finally, equation (8) ensures that the aggregated downstream DC traffic flow destined to a node (node \(d\)) is equal to the downstream DC demand at the corresponding node. Equations (9)–(10) denote the flow conservation constraints in the optical layer. Equation (9) avoids source node of a virtual link having any incoming wavelength channels whereas the destination node of a virtual link has an outgoing wavelength channel. Furthermore, by equation (10), the number of lightpaths traversing a physical link \(mn\) is bounded by the total channel capacity of the fibers between node \(m\) and node \(n\):
\begin{align*}
D^s_{\min} & \cdot O_s^{\uparrow} \leq \sum_{d \in V} \sum_{j \in V} \gamma^d_{up} \cdot \eta^d_{ji} - \gamma^d_{up} \cdot \eta^d_{ji} \leq D^s_{\max} \cdot O_s^{\uparrow}, \quad \forall s \in V \\
-D^s_{\min} \cdot O_s^{\uparrow} & \geq \sum_{d \in V} \sum_{j \in V} \gamma^d_{up} \cdot \eta^d_{ji} - \gamma^d_{up} \cdot \eta^d_{ji} \geq - D^s_{\max} \cdot O_s^{\uparrow}, \quad \forall s \in V \\
\sum_{s \neq i} \sum_{d \neq s} \sum_{j \neq i} \left( \left( \gamma^d_{up} \cdot \eta^d_{ji} \right) - \gamma^d_{up} \cdot \eta^d_{ji} \right) &= 0, \forall i \in V \\
\sum_{j \neq d, j \in V} \left[ \left( \Omega^{\downarrow} \cdot \eta^d_{di} + \rho^d_{ji} \right) - \left( \Omega^{\downarrow} \cdot \eta^d_{di} + \rho^d_{ji} \right) \right] &= \Lambda^d + \Omega^{\downarrow} \cdot \Omega^{\downarrow}, \quad \forall (s, d) \in V, s \neq d \\
\sum_{j \neq s, j \in V} \left[ \left( \Omega^{\downarrow} \cdot \eta^d_{sj} + \rho^d_{ij} \right) - \left( \Omega^{\downarrow} \cdot \eta^d_{sj} + \rho^d_{ij} \right) \right] &= 0, \quad \forall (s, d, i) \in V, d, s \neq i \\
\sum_{j \neq d, j \in V} \Omega^{\downarrow} \cdot \eta^d_{dj} &= \Omega^{\downarrow} \cdot \Omega^{\downarrow}, \quad \forall s, d \in V \\
W_{nm}^{ij} - W_{nm}^{ij} &= \begin{cases} -C_{ij} & m = i \\ C_{ij} & m = j \\ 0 & \text{else} \end{cases}, \quad \forall m, n, i, j \in V \\
\sum_{i \in V} \sum_{j \in V} W_{nm}^{ij} - W_{fm} \leq 0, \forall m, n \in V 
\end{align*}

In equation (11), where $C$ denotes the capacity of a wavelength channel, it is guaranteed that a virtual link has sufficient capacity to provision all types of demands traversing it. Equations (12) and (13) ensure that an upstream DC demand can reach a sufficient number of destinations which is bounded by $D^s_{\min}$ and $D^s_{\max}$. Equation (14) constrains an upstream request to reach a destination $d$ by using at most one virtual link prior to reaching the destination. Equation (15) ensures that the backbone nodes are multicast capable; thus an upstream DC request can flow over the same virtual links up to node $j$, and at node $j$ the demand can be split into multiple virtual links:

\begin{align*}
\sum_{s \in V} \sum_{d \in V} \gamma^d_{ji} + \gamma^d_{up} \cdot \eta^d_{ji} + \gamma^d_{down} \cdot \eta^d_{ji} \leq C \cdot C_{ij}, \quad \forall i, j \in V \\
\sum_{i \neq d} \gamma^d_{up} \cdot \eta^d_{ji} \leq D^s_{\max} \cdot O_s^{\uparrow}, \quad \forall s \in V \\
\sum_{i \neq d} \gamma^d_{up} \cdot \eta^d_{ji} \geq D^s_{\min} \cdot O_s^{\uparrow}, \quad \forall s \in V \\
\sum_{i \neq d} \gamma^d_{up} \cdot \eta^d_{ji} \leq 1, \quad \forall s, d \in V \\
\sum_{j \neq i, j \in V} \gamma^d_{up} \cdot \eta^d_{ji} \leq 1, \quad \forall s, i, j \in V \\
\sum_{i \in V} \gamma^d_{up} \cdot \eta^d_{ji} \leq \mathrm{CPU}_{\mathrm{CAP}}^j \forall j, s, d, \left[ l \in N^d_s \right] \\
\eta^d_{ji} \leq \gamma^d_{up} \cdot \eta^d_{ji} \forall i, j, s, d, \left[ l \in N^d_s \right] \\
\eta^d_{ji} \leq \mathrm{CPU}_{\mathrm{REQ}} \forall i, j, s, d, \left[ l \in N^d_s \right] \\
-\eta^d_{ji} + \mathrm{CPU}_{\mathrm{REQ}} \cdot \gamma^d_{up} + \gamma^d_{up} \cdot \eta^d_{ji} \leq 1, \quad \forall i, j, s, d, \left[ l \in N^d_s \right] 
\end{align*}
\[ \sum_i \tau_{up}^{sd} \pi_{ij}^{sd} \leq \text{MEM}_{CAP}^{j}, \quad \forall i, j, d, l \in N^d \]  
\[ \pi_{ij}^{sd} \leq \gamma_{ldw}, \quad \forall i, j, d, l \in N^d \]  
\[ \pi_{ij}^{sd} \leq \text{MEM}_{REQ}^{i} \chi_{ij}^{sd}, \quad \forall i, j, d, l \in N^d \]  
\[ -\pi_{ij}^{sd} + \text{MEM}_{REQ}^{i} \chi_{ij}^{sd} + \gamma_{ldw} \leq 1, \quad \forall i, j, d, l \in N^d \]  
\[ \sum_i \tau_{up}^{sd} \kappa_{ij}^{sd} \leq \text{BW}_{CAP}^{d}, \quad \forall i, j, d, l \in N^d \]  
\[ \kappa_{ij}^{sd} \leq \gamma_{ldw}, \quad \forall i, j, d, l \in N^d \]  
\[ \kappa_{ij}^{sd} \leq \text{BW}_{REQ}^{i} \chi_{ij}^{sd}, \quad \forall i, j, d, l \in N^d \]  
\[ -\kappa_{ij}^{sd} + \text{BW}_{REQ}^{i} \chi_{ij}^{sd} + \gamma_{ldw} \leq 1, \quad \forall i, j, d, l \in N^d \]  

For simplicity, we assume that all VMs have the same CPU, memory, and bandwidth requirements. Equations (17)–(19) denote the constraints regarding virtual machine CPU requirements and their physical host mappings, whereas \( \chi_{ij}^{sd} \) is a binary variable, and it is one if \( \text{vm}_i \) is placed into the physical host \( j \) in the data center \( d \). In the equations, \( \eta_{ij}^{sd} \) refers to \( \tau_{ldw}^{sd} \cdot \text{CPU}_{REQ}^{i} \chi_{ij}^{sd} \) which is linearized by equations (17)–(19). Similarly, equations (20)–(23) and equations (24)–(27) denote the memory requirements and bandwidth requirement constraints in the data centers, respectively. In the equations, \( \pi_{ij}^{sd} \) refers to \( \gamma_{ldw} \cdot \text{MEM}_{REQ}^{i} \chi_{ij}^{sd} \) which is linearized by equations (21)–(23), whereas \( \kappa_{ij}^{sd} \) refers to \( \gamma_{ldw} \cdot \text{BW}_{REQ}^{i} \chi_{ij}^{sd} \) which is linearized by equations (25)–(27).

### 3.3. Two-step MILP problem solution and heuristics for network topology reconfiguration

Finding an optimal solution for our MILP-based holistic problem without splitting it would require a non-negligible time because it is a heavy optimization problem. Hence, in order to solve it in a reasonable time, we propose a two-step decomposition of our holistic formulation; the basic idea is to determine, in the first step, the data centers satisfying the intra-data center VM placement constraints, and then, in the second step, to route cloud demands towards those data centers with respect to the manycast paradigm as shown in Figure 1(b).

Figure 1(b) illustrates the main steps we follow. The first intra-data center VM placement step executes an MILP which only considers the constraints in equations 16–27 to obtain VM-physical server mappings along with the potential power consumption of the data centers. Indeed, the data centers that do not meet the constraints will not be considered in the second step. The second inter-data center VM placement step, instead, uses the output of the first step to form the set of candidate data centers towards which upstream demands can be routed based on the manycast paradigm.

Finally, to further optimize the second inter-data center VM placement step, we claim the need to adopt heuristics; in particular, we use the delay and power minimized provisioning (DePoMiP) heuristic we proposed previously [20]. In particular, DePoMiP receives, as the output obtained at the end of the first step: the set of demands (i.e. upstream data center demands, downstream demands originating from data centers and regular demands that are neither destined to data centers nor originated from data centers); an initial virtual topology; and the list of candidate data centers for each cloud demand. DePoMiP pops the demands from the demand list and checks their demand types. If a demand is an upstream data center demand, it is provisioned based on manycast. Based on the set of candidate data centers obtained in the first step, the heuristic selects a subset of the corresponding list so that the manycast problem is mapped onto a multiple unicast connection provisioning problem. For each candidate destination data center, the heuristic attempts to route the demand over the virtual topology, and in case of success it updates the virtual link costs. In the case of a failure in routing over the virtual topology, the algorithm
adds a new virtual link from source to the destination node, and routes the newly added virtual link over
the physical topology. As the aim of this paper is not reconfiguration of the cloud backbone but proposing
a holistic framework for inter and inter-data center VM placement, we do not present the details of the
heuristic and refer the reader to Kantarci and Mouftah [20], where destination data center selection, virtual
link cost assignment and physical link cost assignment are explained in detail.

4. PERFORMANCE EVALUATION

Let us open this section with a theoretical analysis of the runtime overhead needed to compute the MILP
holistic formulation (see Section 3.3) by comparing it with the overhead needed for the computation of
the two-step MILP holistic problem for different problem sizes in terms of number of physical hosts and
variables (see Figure 2). In particular, we consider the six-node topology in Figure 3(a) and the NSFNET
topology in Figure 3(b) in the cloud backbone. The y-axis of the figure denotes the number of variables in
the two topologies, while the x-axis denotes the size of the data centers in terms of the product of the number
of VMs per physical hosts and the total number of physical hosts in a data center. As seen in the figure,
when the problem is aimed to be solved by the holistic MILP formulation, the number of variables to be
solved varies between $10^6$ and $10^7$ when the number of VMs per data center varies between 2000 and
20,000 under the six-node topology. When the backbone network is larger, such as the NSFNET topology,
the number of variables to be solved varies between $10^7$ and $10^8$ as the number of VMs per data center
increases. The two dashed curves denote the number of variables to be solved if inter-data center VM place-
ment is solved via our two-step MILP problem solution. As seen in the figure, the number of variables to be
solved is degraded by 100 for both cases. Furthermore, the impact of the backbone topology on the number
of variables to be solved is not as significant as it is under the optimal solution.

Hereby, we can conclude that solving the holistic MILP formulation may lead to time-wise infeasibility,
and that motivates the need for our two-step MILP holistic problem solution; in the remainder of this
section, we will focus on the two-step MILP holistic problem solution only (or the short-form holistic
MILP) to thoroughly assess its performance.

The performance evaluation results are organized into two subsections. The first subsection
compares our proposed two-step solution with the solution proposed in Kantarci and Mouftah [20],
taken as a benchmark MILP-based solution. Differently from our two-step holistic solution, Kantarci
and Mouftah [20] select the closest data center that can accommodate VM placement requests taking
into account CPU, memory and bandwidth capacities of physical hosts in the target data centers. In
other words, the objective of this benchmark MILP model is to find the shortest manycast tree and

![Figure 2. Run-time overhead of MILP holistic formulation and two-step MILP holistic problem in
terms of problem size](image-url)
the shortest path (both in kilometers) to serve incoming cloud demands. More formally, the benchmark solution has the objective function formulated in equation (efeq:constminslt), and the performance metrics of the comparison are power consumption, average VM utilization, maximum VM index, physical CPU load, and CPU load fairness among nodes:

\[
\min \sum_{m \in V} \sum_{n \in V} \sum_{i \in V} \sum_{j \in V} \lambda_{mnij} = \Lambda_{mn} + \gamma_{mnij}^{up} + \gamma_{nmij}^{down} / C_{Li} \cdot j (28)
\]

The second subsection, instead, investigates the impact of the DePoMiP heuristic, by comparing its performances with that obtainable by the two-step MILP holistic problem solution.

Figure 3. Two topologies considered in the cloud backbone: (a) six-node topology; (b) NSFNET topology.
We conclude this initial part by detailing the configurations used for all our experiments. We assume each data center provides 1120 physical hosts where at most five VMs run on each physical host. Moreover, we assume each data center is initially used at [0.1–0.8]% while an upstream demand is assumed to introduce a utilization increase between 0.025% and 0.2%. We adopt the experimental settings as in Corradi et al. [22]; the physical hosts have dual-core CPUs with 4 GB memory, whereas a VM includes one virtual CPU (VCPU) and 512 MB memory capacity; also, we have considered three VCPU load levels ($\rho_{\text{VCPU}}$) for the VMs as 20%, 60% and 80%.

We assume the power consumption of the network equipment, $P_{\text{edfa}}$, $P_{t}$ and $P_{r}$ to be 8, 73 and 1000 W respectively, while each fiber link in the backbone is assumed to operate at 40 Gbps with 16 channels. To be coherent with Kantarci and Mouftah [20], we set to 2 the desired number of destination data centers to be reached for upstream demands, while the size of the destination set varies between 3 and 4. For idle and fully utilized cases, the data center is assumed to consume 168 and 319.2 kW for processing (100 and 280 kW for cooling) power to be coherent with Kantarci and Mouftah [20]. We have used two topologies, namely the six-node topology in Figure 3(a) and the NSFNET topology in Figure 3(b) for performance evaluation. In the six-node topology each node experiences the same load profile for every demand type throughout the day, as shown in Table 1, whereas in the NSFNET topology each set of node experiences a different demand profile due to different time zones, as shown in Table 2.

### 4.1. Evaluation of the proposed two-step MILP holistic problem solution

Figure 4(a, b) illustrates the power consumption of the overall cloud network and the data centers under the holistic and benchmark solutions under the six-node and the NSFNET topologies, respectively. Since the data centers are assumed to have initial loads, the initial power consumption of the cloud system is illustrated as a benchmark in both plots. Both figures show that, starting from the on-peak hours of the days, the proposed solution introduces lower power consumption in the data centers. Furthermore, a comparison of the results for the two topologies and traffic patterns shows that the

#### Table 1. Demand profile in the six-node topology throughout the day (Gbps)

<table>
<thead>
<tr>
<th>Demand type</th>
<th>01–03</th>
<th>04–06</th>
<th>07–09</th>
<th>10–12</th>
<th>13–15</th>
<th>16–18</th>
<th>19–21</th>
<th>22–24</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>90</td>
<td>90</td>
<td>110</td>
<td>80</td>
<td>100</td>
</tr>
<tr>
<td>Upstream data center</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>18</td>
<td>18</td>
<td>22</td>
<td>16</td>
<td>20</td>
</tr>
<tr>
<td>Downstream data center</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>135</td>
<td>135</td>
<td>165</td>
<td>120</td>
<td>150</td>
</tr>
</tbody>
</table>

#### Table 2. Demand profile in the NSFNET throughout the day (Gbps)

<table>
<thead>
<tr>
<th>Hours</th>
<th>01–03</th>
<th>04–06</th>
<th>07–09</th>
<th>10–12</th>
<th>13–15</th>
<th>16–18</th>
<th>19–21</th>
<th>22–24</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular Internet demands</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>90</td>
<td>90</td>
<td>110</td>
<td>80</td>
<td>100</td>
</tr>
<tr>
<td>[8,10,11,12,14]</td>
<td>50</td>
<td>30</td>
<td>30</td>
<td>70</td>
<td>90</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>[6,7,9]</td>
<td>100</td>
<td>40</td>
<td>30</td>
<td>30</td>
<td>90</td>
<td>90</td>
<td>110</td>
<td>80</td>
</tr>
<tr>
<td>[4,5,13]</td>
<td>80</td>
<td>30</td>
<td>30</td>
<td>50</td>
<td>100</td>
<td>90</td>
<td>110</td>
<td>80</td>
</tr>
<tr>
<td>Upstream data center demands</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>18</td>
<td>18</td>
<td>22</td>
<td>16</td>
<td>20</td>
</tr>
<tr>
<td>[8,10,11,12,14]</td>
<td>10</td>
<td>6</td>
<td>6</td>
<td>14</td>
<td>18</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>[6,7,9]</td>
<td>20</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>18</td>
<td>18</td>
<td>22</td>
<td>16</td>
</tr>
<tr>
<td>[4,5,13]</td>
<td>16</td>
<td>6</td>
<td>6</td>
<td>10</td>
<td>20</td>
<td>18</td>
<td>22</td>
<td>16</td>
</tr>
<tr>
<td>Downstream data center demands</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>135</td>
<td>135</td>
<td>165</td>
<td>120</td>
<td>150</td>
</tr>
<tr>
<td>[8,10,11,12,14]</td>
<td>75</td>
<td>45</td>
<td>45</td>
<td>105</td>
<td>135</td>
<td>150</td>
<td>150</td>
<td>150</td>
</tr>
<tr>
<td>[6,7,9]</td>
<td>150</td>
<td>60</td>
<td>45</td>
<td>45</td>
<td>135</td>
<td>135</td>
<td>165</td>
<td>120</td>
</tr>
<tr>
<td>[4,5,13]</td>
<td>120</td>
<td>45</td>
<td>45</td>
<td>75</td>
<td>150</td>
<td>135</td>
<td>165</td>
<td>120</td>
</tr>
</tbody>
</table>
The proposed solution can cope with the heterogeneity of the demand profile (in Figure 4b), while the benchmark solution further increases the power consumption in the cloud.

In Figure 5(a, b), the holistic solution is compared with the benchmark solution in terms of average CPU load per active physical host throughout the day under the two topologies. Here, the cloud system is tested under various VCPU load levels for the VMs. The trends in both figures show that at a certain time the average number of VMs per physical host varies with the VCPU load. This behavior is due to the fact that under a certain physical CPU load the number of VMs per physical host is a function of the VCPU load; in other words, a certain physical CPU load can be induced by a lower number of VMs with higher VCPU load levels [22]. We illustrate this phenomenon in Figure 6(a, b), where we compare our proposed solution to the benchmark solution in terms of maximum VM index ($i_{\text{max}}^{\text{VM}}$) per physical host. With maximum VM index, we denote the maximum number of VMs that are placed on a physical host in a data center, more formally $i_{\text{max}}^{\text{VM}}$ can be expressed via equation 29. Thus $i_{\text{max}}^{\text{VM}}$ provides information on the physical host utilization in the data centers:

$$i_{\text{max}}^{\text{VM}} = \arg\max \left( \sum_{i} x_{ij}^d, \forall d, j \right)$$

proposed solution can cope with the heterogeneity of the demand profile (in Figure 4b), while the benchmark solution further increases the power consumption in the cloud.

In Figure 5(a, b), the holistic solution is compared with the benchmark solution in terms of average CPU load per active physical host throughout the day under the two topologies. Here, the cloud system is tested under various VCPU load levels for the VMs. The trends in both figures show that at a certain time the average number of VMs per physical host varies with the VCPU load. This behavior is due to the fact that under a certain physical CPU load the number of VMs per physical host is a function of the VCPU load; in other words, a certain physical CPU load can be induced by a lower number of VMs with higher VCPU load levels [22]. We illustrate this phenomenon in Figure 6(a, b), where we compare our proposed solution to the benchmark solution in terms of maximum VM index ($i_{\text{max}}^{\text{VM}}$) per physical host. With maximum VM index, we denote the maximum number of VMs that are placed on a physical host in a data center, more formally $i_{\text{max}}^{\text{VM}}$ can be expressed via equation 29. Thus $i_{\text{max}}^{\text{VM}}$ provides information on the physical host utilization in the data centers:

$$i_{\text{max}}^{\text{VM}} = \arg\max \left( \sum_{i} x_{ij}^d, \forall d, j \right)$$

Figure 4. Overall and data center power consumption under the benchmark and holistic MILP solutions: (a) under six-node topology; (b) under NSFNET.
Figure 5(a, b) shows that the proposed solution does not increase the VM utilization when compared with the benchmark solution. Although this is an expected behavior due to the assumption of the CPU requirements of the VMs and the capacities of the physical hosts, when the VMs and physical hosts are differentiated with respect to the CPU requirements and the capacities the holistic solution can be expected to introduce further performance enhancement in terms of VM utilization per physical host.

In Figure 7(a, b) our proposed solution is compared with the benchmark solution in terms of average physical CPU load per data center throughout the day. As the figure shows, for most time slots the adoption of the VM placement in Corradi et al. [22] and the adoption of data center selection in Kantarci and Mouftah [20] produces a CPU load on the physical hosts to be significantly less under the holistic solution. Furthermore, under the NSFNET topology, the holistic solution takes advantage of heterogeneous demand profile in a larger region, and can place the VMs in the data centers where CPU load per active physical host will be lower upon placement of the corresponding VMs.

Figure 5. Node-by-node VM utilization under the benchmark and holistic solutions: (a) in six-node topology; (b) in NSFNET
We have further evaluated the holistic solution in terms of CPU load fairness, having formulated the CPU load fairness index (CLFI) by adopting Jain’s fairness index figure. As shown in equation (30), CLFI is a function of the average CPU load per physical host in the data center $i$ ($CPU_{Load}^i$) and the number of data centers. In Figure 8a, b), the CLFI of the holistic solution is evaluated by considering the average CPU load per physical host for each data center at each time slot of the day under the two topologies. Due to uniformly distributed demand profile in the six-node topology, the holistic solution introduces over 0.94 CLFI throughout the day (Figure 8a), whereas under the NSFNET the holistic solution introduces over 0.81 CLFI to various data centers in the cloud network (Figure 8b). It can be said that the proposed solution introduces feasible CLFI to the data centers under both topologies. However, demand profile and backbone topology characteristics appear as the factors that affect the network-wide CLFI:

$$CLFI = \left( \frac{\sum_{i=1}^{N} CPU_{Load}^i}{N} \right)^2 / \left( N \cdot \sum_{i=1}^{N} CPU_{Load}^i \right)$$

(30)

4.2. Impact of the heuristic solution for inter-data center step computation

The two-step MILP problem presented above can still lead to very long times as the network size increases and data centers are scaled up. Therefore, heuristics can provide faster results. In particular, we

Figure 6. Maximum VM index in the cloud under the holistic solution: (a) in six-node topology; (b) in NSFNET.
have evaluated our two-step solution under the energy-aware provisioning heuristic in Kantarci and Mouftah [20], and compared its performance to the results obtained under the MILP, where the cloud backbone lies over the NSFNET topology.

Figure 9 shows that the holistic solution, under the energy-aware heuristic, leads to very similar average VM utilization values in comparison to that under the holistic solution, confirming that it is possible to adopt our heuristic in placing the VMs in between the data centers.

Figure 10(a–c) illustrates the difference between the two-step MILP and heuristic solutions in terms of maximum VM indices under different VCPU load levels. It is worth noting that the difference denotes the ratio of the absolute difference between the maximum VM indices to the maximum VM index of the MILP-based solution. As seen in the figure, as the CVPU load level gets higher (i.e. $\rho_{\text{VCPU}} = 60\%$ in Figure 10b and $\rho_{\text{VCPU}} = 80\%$ in Figure 10c), the results under the heuristic-based solution gets closer to those under the MILP-based solution with respect to the locations, as well as the time slots. At the same time, the heuristic-based solution is still able to limit the maximum VM index to that under the MILP-based solution under low VCPU load levels, as seen in Figure 10(a) (i.e. $\rho_{\text{VCPU}} = 20\%$).

Figure 7. Physical CPU load under the benchmark and holistic solutions: (a) in six-node topology; (b) in NSFNET

Copyright © 2013 John Wiley & Sons, Ltd. Int. J. Network Mgmt DOI: 10.1002/nem
Figure 8. CLFI under the holistic solution: (a) in six-node topology; (b) in NSFNET

Figure 9. Average VM utilization under the two-step MILP, and by employing the heuristic solution
<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>01−03</td>
<td>04−06</td>
<td>07−09</td>
<td>10−12</td>
<td>13−15</td>
<td>16−18</td>
<td>19−21</td>
<td>22−24</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>0.2</td>
<td>0.4</td>
<td>0.6</td>
<td>0.8</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 10. Difference between maximum VM indices under the two-step MILP and heuristic solutions: (a) $\rho_{\text{VCPU}} = 20\%$; (b) $\rho_{\text{VCPU}} = 60\%$; (c) $\rho_{\text{VCPU}} = 80\%$
Figure 11 illustrates the average physical CPU load difference between the MILP-based and heuristic-based solutions throughout the day. In most time slots of the day (except 7:00–9:00 am EST), the difference is less than 1% compared to the two-step MILP solution. Furthermore, even between the 7:00 and 9:00 am time slot, the power-aware heuristic can introduce physical CPU loads that are less than 5% of the MILP-based solution. The reason for a higher difference during this time slot is that the demand profile is more uniform rather than being heterogeneous and, further, the corresponding time slot is an off-peak time for three of the four regions. Therefore, the heuristic selects the first feasible solution for the upstream demands although there could be other destinations where placing the incoming workload may lead to less increase in power consumption, as well as the physical CPU load level in the data center. Figure 12 presents a more detailed comparison as the difference between the physical CPU loads is presented with respect to node locations and time slots. As seen in the figure, for most node ID–time slot pairs the difference between the physical CPU loads of MILP-based and heuristic-based solutions is insignificantly low, and of the order of $10^{-3}$.

![Figure 11. Difference between physical CPU loads under the MILP and heuristic solution](image1)

![Figure 12. Node-by-node and hourly difference between physical CPU loads under the MILP and heuristic solution](image2)
5. CONCLUSION AND FUTURE WORK

Smart VM placement algorithms are essential to exploit data center computing resources efficiently. In this paper, we study the problem of energy-efficient intra and inter-data center VM placement. To achieve this goal, we formulate a MILP-based holistic optimization problem and we propose a two-step MILP solution and a heuristic to solve it in reasonable time. After an initial presentation of our new MILP formulation, we showed that it allows considerable power savings to be obtained when applied to large-scale cloud systems consisting of multiple medium-size data centers geographically distributed and connected via a backbone network. Experimental results indicate that our VM placements increase not only energy efficiency but it also permit achievement of better fairness.

Furthermore, we have evaluated the impact of a previously proposed power-aware provisioned heuristic on the performance of the proposed inter and intra-data center VM placement framework. The numerical results show that adoption of the power-aware heuristic in Kantarci and Mouftah [20], along with the holistic solution towards inter and intra-data center VM placement, can introduce promising results in terms of both average and maximum VM utilization, physical CPU load and CPU load fairness.

Therefore, we propose a two-step inter and intra-data center placement approach where, in the first step, data centers satisfying the memory, CPU capacity and bandwidth constraints are determined, and in the second step the demands are routed towards the selected data centers based on the manycast paradigm.

Several research directions are left for future work. First, we intend to extend our formulation to consider also groups of multiple VMs with high and correlated traffic such as in the case of VMs collaborating in the same MapReduce data-processing task. Second, we are developing an OpenStack-based prototype for intra and inter-data center management to support our holistic VM placement approach, in order to assess the proposed optimization model in a real-world testbed.

REFERENCES

DESIGN OF ENERGY-EFFICIENT CLOUD SYSTEMS


AUTHORS’ BIOGRAPHIES

Burak Kantarci is a postdoctoral fellow at the School of Electrical Engineering and Computer Science of the University of Ottawa. Dr. Kantarci received the M.Sc. and Ph.D. degrees in Computer Engineering from Istanbul Technical University in 2005 and 2009, respectively, and he completed the major content of his PhD thesis at the University of Ottawa between 2007 and 2008 under the supervision of Prof. Hussein Mouftah. He was the recipient of the Siemens Excellence Award in 2005 for his contributions to the optical burst switching research. He has co-authored over six-dozen technical papers in established journals and flagship conferences, and he has contributed to eight book chapters. He is the co-editor of Communication Infrastructures for Cloud Computing. He has been serving in the Technical Program Committees of Ad Hoc and Sensor Networks Symposium, Optical Networks Symposium and Green Communication Systems Track of the Selected Areas in Communications Symposium of IEEE GLOBECOM and IEEE ICC conferences. He is also co chairing the International Workshop on Management of Cloud Systems. Dr. Kantarci is a Senior Member of the IEEE.

Luca Foschini graduated from University of Bologna, Italy, where he received PhD degree in computer science engineering in 2007. He is now an assistant professor of computer engineering at the University of Bologna. His interests include distributed systems and solutions for pervasive wireless computing environments, system and service management, context-aware services and adaptive multimedia, and management of Cloud computing systems. He is member of IEEE and ACM.

Antonio Corradi graduated from University of Bologna, Italy, and received MS in electrical engineering from Cornell University, USA. He is a full professor of computer engineering at the University of Bologna. His research interests include distributed and parallel systems and solutions, middleware for pervasive and heterogeneous computing, infrastructure support for context-aware multimodal services, network management, mobile agent platforms. He is member of IEEE, ACM, and Italian Association for Computing (AICA).

Hussein T. Mouftah joined the School of Information Technology and Engineering (now School of Electrical Engineering and Computer Science) of the University of Ottawa in 2002 as a Tier 1 Canada Research Chair Professor, where he became a University Distinguished Professor in 2006. He has been with the ECE Dept. at Queen’s University (1979-2002), where he was prior to his departure a Full Professor and the Department Associate Head. He has six years of industrial experience mainly at Bell Northern Research of Ottawa (became Nortel Networks). He served IEEE ComSoc as Editor-in-Chief of the IEEE Communications Magazine (1995-97), Director of Magazines (1998-99), Chair of the Awards Committee (2002-03), Director of Education (2006-07), and Member of the Board of Governors (1997-99 and 2006-07). He has been a Distinguished Speaker of the IEEE

Copyright © 2013 John Wiley & Sons, Ltd. Int. J. Network Mgmt DOI: 10.1002/nem
Communications Society (2000-07). Also he served IEEE Canada (Region 7) as Chair of the Awards and Recognition Committee (2009-12). He is the author or coauthor of 9 books, 60 book chapters and more than 1300 technical papers, 12 patents and 140 industrial reports. He is the joint holder of 12 Best Paper and/or Outstanding Paper Awards. He has received numerous prestigious awards, such as the 2008 ORION Leadership Award of Merit, the 2007 Royal Society of Canada Thomas W. Eadie Medal, the 2007-2008 University of Ottawa Award for Excellence in Research, the 2006 IEEE Canada Mc-Naughton Gold Medal, the 2006 EIC Julian Smith Medal, the 2004 IEEE ComSoc Edwin Howard Armstrong Achievement Award, the 2004 George S. Glinski Award for Excellence in Research of the U of O Faculty of Engineering, the 1989 Engineering Medal for Research and Development of the Association of Professional Engineers of Ontario (PEO), and the Ontario Distinguished Researcher Award of the Ontario Innovation Trust. Dr. Mouftah is a Fellow of the IEEE (1990), the Canadian Academy of Engineering (2003), the Engineering Institute of Canada (2005) and the Royal Society of Canada RSC: The Academy of Science (2008).