WHITEPAPER

Inter-Cloud Workload Optimization

A SCALABLE HIERARCHICAL APPROACH

How to optimize workload distribution, save energy costs and reduce carbon emissions in an Inter-Cloud environment

Agostino Forestiero(1,2), Raffaele Giordanelli(1), Carlo Mastroianni(1,2), Michela Meo(3), Giuseppe Papuzzo(1,2)

(1) Eco4Cloud – www.eco4cloud.com
(2) ICAR-CNR, Institute for High Performance Computing and Networking of the Italian National Research Council, Italy – www.icar.cnr.it
(3) Department of Electronics and Telecommunications at Politecnico di Torino, Italy

Inter-Cloud Workload Optimization

1. Executive summary

Optimal distribution of applications and Virtual Machines to servers is still an open problem, especially in large and dynamic systems such as inter-cloud instance

In this whitepaper, we first examine the problem of efficient distribution of the workload in the “Inter-Cloud”, i.e., a scenario in which several data centers or cloud instances are interconnected and applications can be routed or migrated among them. Section 2 introduces the **ECOINTERCLOUD** architecture for workload distribution in an Inter-Cloud environment. Section 3 discusses the business goals that can be achieved by using the **ECOINTERCLOUD** solution. Section 4 includes a high-level technical description of the algorithms used for the assignment and the migration of Virtual Machines, and reports the efficiencies that can be achieved in a specific scenario. Finally, Section 5 discusses technical issues related to inter-DC VM migration.

The optimal distribution of applications and Virtual Machines to servers is still an open problem, especially in large and dynamic systems. The problem is even more complex in geographically distributed data centers, whose adoption is rapidly increasing. They are deployed by major cloud service providers, such as Amazon, Google, and Microsoft, to match the increasing demand for resilient and low-latency cloud services, or to interconnect heterogeneous data centers owned by different companies in so called “Inter-Cloud” scenarios, and architectural frameworks are being introduced by major vendors (e.g. Cisco’s Intercloud Fabric or VMware’s vCloud Air) to build highly secure hybrid clouds and extend private data centers to public clouds as needed, on demand, bi-directionally to the location(s) of choice and with consistent network and security policies. In such environments, data centers offer different and time-varying energy prices, and workload variability is experienced both within single sites and across the whole infrastructure. Moreover, the Inter-Cloud scenario includes the “hybrid Cloud” case, in which private and public are interconnected and integrated.

The dynamic migration of workload among data centers has become an opportunity

The dynamic migration of workload among data centers has become an opportunity to improve several aspects: better resiliency and failover management, improved load balancing, exploitation of the “follow the moon” paradigm (i.e., move the workload where the energy is cheaper/cleaner and/or cooling costs are lower), Host Thin Provisioning™ in managed services scenarios. Inter-site migration is enabled by the availability of a much higher network capacity and guaranteed latency, thanks to both physical improvements (e.g., through techniques such as wavelength division multiplexing) and logical/functional enhancements (e.g., the adoption of Software Defined Networks). Reliable and low-latency

__EcoInterCloud__ extends the __Eco4Cloud__ consolidation algorithm to the Inter-Cloud scenario.

connections can be used to shift significant amount of workload from one site to another through dedicated networks or even via regular Internet connections.

The existing approaches aim to solve the optimization problem as a whole, in a centralized fashion, undergoing the risk of originating two main issues: (i) algorithms of this kind may be poorly scalable, both for the number of parameters that they must consider and for the huge size of the problem, as it may involve tens of thousands of servers; (ii) they generally assume that all sites share the same strategy and algorithms, which may hamper their autonomy. The need for autonomous management is self-explanatory in multi-owned data centers, and is crucial even within a single-owner infrastructure, for example in the case that one or several sites are the former asset of an acquired company, or are hosted by co-located multi-tenant facilities.

2. Architecture for Inter-Cloud Workload Distribution

Beyond decentralization, a key feature of EcoInterCloud is its modularity

This section describes the hierarchical architecture of **ECOINTERCLOUD** for the efficient and autonomic management of the workload in a multi-site scenario. The architecture is composed of two layers: (i) the **upper layer** is used to exchange information among the different sites and drive the distribution of Virtual Machines among the data centers and (ii) the **lower layer** is used to allocate the workload within single data centers.

ECOINTERCLOUD extends the **ECO4CLOUD** consolidation algorithm[1] to the Inter-Cloud scenario. **ECO4CLOUD** dynamically consolidates Virtual Machines (VMs) to the minimum number of servers, and allows the remaining servers to enter low consuming sleep modes. With **ECO4CLOUD** key decisions regarding the local data center are delegated to single servers, which autonomously decide whether or not to accommodate a VM or trigger a VM migration. The data center manager only maintains a coordination role. In a similar fashion, at the upper level of the multi-site **ECOINTERCLOUD** architecture, most of the intelligence is left to individual data centers which, for example, decide which information is relevant and should be delivered to other data centers, which portion of the local workload should be migrated elsewhere, etc.

Coordinating decisions, for example about the necessity of migrating an amount of workload from one site to another, are taken combining the information related to individual data centers. Beyond decentralization, a key feature of **ECOINTERCLOUD** is its modularity: provided that the interface between the lower and the upper layer is preserved, each layer is free to modify the respective algorithms and their implementation. At the lower layer, each data center is fully autonomous, and can manage the internal workload using either **ECO4CLOUD** or any other consolidation algorithm. So different data centers can adopt different internal algorithms and pursue different optimization strategies. On the other hand, *Individual data centers can adopt different internal algorithms and pursue different optimization strategies*

the upper layer algorithms may be tuned or modified without causing any impact on the operation of individual sites.

The reference scenario is reflected in Figure 1, which shows the upper and lower layer for two interconnected data centers, as well as the main involved components. At each data center a data center manager (**DCM**) runs the algorithms of the upper layer, while the local manager (**LM**) performs the functionalities of the lower layer. In the most typical case, both the DCM and LM may be deployed on the same host as the manager of the local virtualization infrastructure, e.g., the vCenter in the case of VMware. The DCM integrates the information coming from the lower layer and uses it to implement the functionalities of the upper layer. The DCM is required to: (i) communicate with the local LM in order to acquire detailed knowledge about the current state of the local data center, for example regarding the usage of host resources and the state of running VMs; (ii) extract relevant high level information about the state of the data center; (iii) transmit/receive such high level information to/from all the other DCMs; (iv) execute the algorithms of the upper layer to combine the collected information and take decisions about the distribution of the workload among the data centers. For example, the assignment algorithm is used to decide to which data center a new VM should be assigned. Once the VM is delivered to the target site, the local LM runs the lower layer algorithms to assign the VM to a specific host.

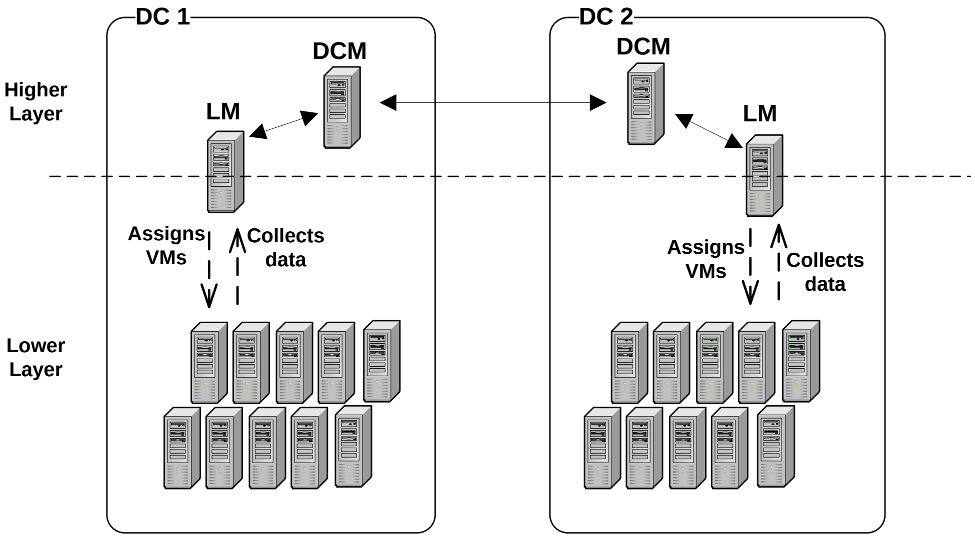


Figure 1 – **ECOINTERCLOUD** scenario: upper and lower layer

As shown in Figure 2, the framework is designed so that all the DCMs are able to execute the upper layer algorithms and, for example, choose the target DC for a VM originated locally. This requires an all-to-all data transmission among the DCMs, but this is not an issue due to the relatively low number of interconnected sites and the tiny amount of

This strategy avoids the choice of a single coordination point that in a multi-site scenario may be inappropriate for administrative reasons

transmitted data. On the other hand, this strategy avoids the choice of a single coordination point that in a multi-site scenario may be inappropriate for administrative reasons. If the number of interconnected data centers becomes relatively large, e.g., more than a few dozens, the DCMs may be organized in a hierarchical or peer-to-peer architecture.

Since the single data centers are autonomous regarding the choice of the internal algorithms for workload management, this paper focuses on the algorithms of the upper layer. Three basic algorithms are executed at each DCM: (i) the assignment algorithm that determines the appropriate target data center for each new VM; (ii) the redistribution algorithm that periodically evaluates whether the current load balance is appropriate and, if necessary, decides whether an amount of workload should be migrated to/from another site; (iii) the migration algorithm that determines from which source site and to which target site the workload should be migrated.

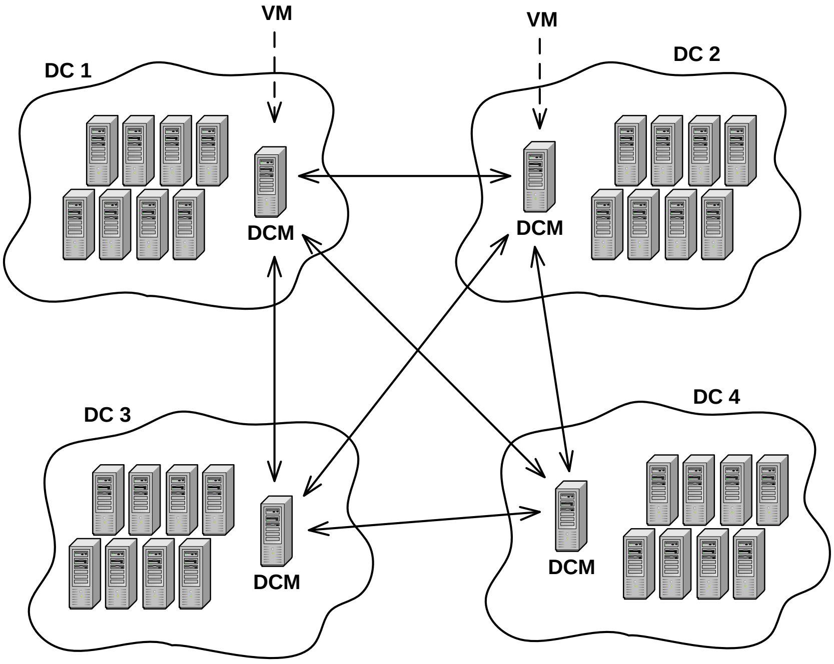


Figure 2 – **EcoInterCloud** scenario: the DCMs of four data centers exchange high level information about the state of local data centers

3. Business Goals Related to Inter-Cloud Optimization

As mentioned in the previous section, a key responsibility of the DCM is to analyze detailed data about the local data center and summarize relevant information that is then transmitted to remote DCMs and used for the assignment and redistribution of workload. The nature of

Some important benefits of Inter-Cloud workload optimization are in terms of cost savings, energy efficiency, CO_2 emissions, Host Thin Provisioning^{TM}, Quality of Service, load balancing, reduced inter-DC communications

the high level information depends on the objectives that must be achieved. Some important goals are:

- **Cost Savings**. The cost associated to the execution of a given workload depends on many factors, among which the cost of power needed for computation, for cooling and for power distribution, the costs related to staff, servers maintenance, etc. An important element to consider is that the cost of electricity is generally different from site to site and also varies with time during the day, even on a hourly basis, therefore the overall cost may be reduced by shifting portions of the workload to more convenient or cheaper sites;
- **Energy Efficiency**. The amount of consumed energy is generally easier to evaluate than the costs, as moderns data centers are equipped with instrumentation to monitor the power usage in computational resources. The total power may be obtained by multiplying the power consumed for computation by the PUE (Power Usage Efficiency) index;
- **Carbon Emissions**. Companies are today strongly encouraged to reduce the amount of carbon emissions, not only to compel to laws and rules, but also to advertise their green effort and attract customers that are increasingly careful about sustainability issues;
- **Host Thin Provisioning^{TM}**. Autonomically provisioning only the number of physical servers and cores that applications need, rather than the number provisioned. Host Thin Provisioning^{TM} works by consolidating workloads and switching on/off physical cores as workload varies in real-time;
- **Quality of Service**. The workload must be distributed without overloading any single site, as this may affect the quality of the service perceived by users. Moreover, quality of service may be improved by properly combining and assigning applications having different characteristics, for example, CPU-bound and RAM-bound applications;
- **Load Balancing**. In a multi-DC environment, especially if managed by the same organization, it may be important to balance the load distributed to the different sites. Among the relevant rationales are: a better balance may help improve the responsiveness of the sites, decrease the impact on physical infrastructure – e.g., in terms of cooling and power distribution – help prevent overload situations.
- **Reduced Inter-DC Communications**. In some cases it is more efficient to assign a VM to the local data center, instead of delivering it to a more convenient remote data center, depending on many factors, among which the amount of data used by the VM, the available inter-DC bandwidth and the type of applications hosted by the VMs. For example, choosing a local data center is more convenient in the case that

the VM hosts a database server, much less if it runs a Web service, especially in the frequent case that Web services are replicated on several data centers.

All the above mentioned goals are important, yet different data centers may focus on different aspects, depending on the specific operating conditions and on the priorities prescribed by the management. It is up to the company's management to specify the objectives and their relative weighs, and they are combined in order to assign the workload so as to match the objectives. For example, let us assume that the primary objectives are the reduction of overall carbon emissions, the load balancing and the reduction of costs. These goals are representative of opposite needs, the need for optimizing the overall efficiency (in terms of costs and carbon emissions) and the need for guaranteeing the fairness among data centers.

4. Algorithms for VM Assignment and Migration

The optimal distribution of the workload among the data centers is driven by the values of a purposely defined *assignment function*, which balances and weighs the chosen business goals. The strategy is to assign a VM to the data center with the lowest value of the function. For example, if the objectives are the balance of load, the minimization of carbon emissions and the minimization of costs related to energy, the assignment function **Ai**, for each data center **i**, is defined as follows:

(1)

$$A_i = \alpha \cdot \frac{C_i}{C_{max}} + \beta \cdot \frac{U_i}{U_{max}} + \gamma \cdot \frac{E_i}{E_{max}}$$

The three parameters **Ci**, **Ui** and **Ei**, are related, respectively, to carbon emissions, overall utilization and energy costs. The parameters are normalized with respect to the maximum values communicated by data centers, and weighed through the coefficients **α**, **β** and **γ**, whose sum equals to 1.

To compute the function, it is required that the DCM of each data center transmits to the others some very simple pieces of data. In the examined case, relevant information is: (i) the price of energy, **Pi**, which is assumed to be the same on all the servers of a data center, (ii) the utilization of the bottleneck resource, denoted as **Ui**, (iii) the best available *carbon footprint rate* of a local server, **cs**, and the data center PUE. The carbon parameter **Ci** of a data center **i**, measured in Tons/MWh, defines the amount of carbon emitted per consumed energy [2] and can be expressed as:

The use of EcoInterCloud assignment and migration procedures allows to efficiently optimize the workload and achieve the specified objectives

$$C_i = PUE_i \cdot min\{c_s | \ server \ s \ is \ available\}$$

Analogously, the parameter related to the cost of energy, **Ei**, is defined as:

$$E_i = PUE_i \cdot P_i$$

Indeed, the overall cost of energy is obtained by multiplying the energy consumed by the IT component of the data center first by the PUE – which gives the total amount of consumed energy, including power distribution and cooling – and then by the price of energy.

After computing the values of **Ai** for each data center, the VM is assigned to the data center having the lowest value. Once consigned to the target data center, the VM will be allocated to a physical host using the local assignment algorithm, for example **ECO4CLOUD**.

The assignment algorithm is used to route a new VM to the best target data center. However, the workload distribution may become inefficient when the conditions change, e.g., the overall load or the price of energy may vary in one or more data centers. In such cases inter-DC VM migrations are performed to redistribute the workload considering the new conditions.

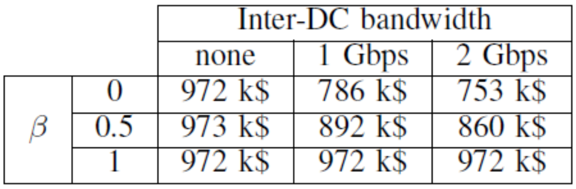
The migration algorithm is triggered when the values of the assignment functions of two data centers differ by more than a predetermined threshold, for example 3%. The frequency at which this condition is evaluated depends on the dynamic nature of the specific scenario, for example on the frequency at which the price of energy varies. When such an imbalance is detected, VMs are migrated from the data center having the highest value of the function **A** to the data center with the minimum value, until all the values reenter within the tolerance range. The frequency of migrations is limited by the bandwidth between the source and target data centers. This bandwidth may correspond to the physical bandwidth of inter-DC connections or may be a portion of the physical bandwidth reserved by data center administrators for this purpose.

EcoInterCloud is a viable option today

4.1 Cost Savings Example

The use of **EcoInterCloud** assignment and migration procedures allows to efficiently optimize the workload and achieve the specified objectives. Simulations were performed for a sample scenario with four data center located in California, Ontario, UK and Germany[1], with PUE of 1.56, 1.7, 1.9 and 2.1 respectively. Furthermore, it is assumed that the business goals are the reduction of overall energy cost and the fair distribution of the workload among the data centers. This corresponds to setting the parameter α to 0 in expression (1), while the value of the parameter β can be tuned to give the desired weighs to the two objectives, considering that $\gamma=1-\beta$.

Table 1 reports the values of daily energy cost for different values of the parameter β and of the available inter-DC bandwidth. Costs are notably reduced thanks to the inter-DC migration process. With $\beta=0.5$, the daily cost is equal to about 860 thousand dollars when the bandwidth is 2 Gbps, while it is about 973 thousand dollars without migrations: a cost saving of about 113 k$ is then achieved through inter-DC migrations. Cost savings are even higher with $\beta=0$, since the load balancing is not taken into account as an objective. In this case, the daily saving is equal to about 284 k$.



		Inter-DC bandwidth		
		none	1 Gbps	2 Gbps
β	0	972 k$	786 k$	753 k$
	0.5	973 k$	892 k$	860 k$
	1	972 k$	972 k$	972 k$

Table 1. Daily cost of energy for the considered Inter-Cloud environment

---

[1] Energy prices of the four data centers are taken from the following web sites:

- California: www.pge.com/tariffs/IndustrialCurrent.xls
- Ontario: www.hydroone.com/RegulatoryAffairs/RatesPrices/Pages
- UK: http://en.wikipedia.org/wiki/Electricity_billing_in_the_UK
- Germany: www.iwr-institut.de/en/press/background-informations

5. Inter-DC VM Migrations with Eco4Cloud and VMware

EcoInterCloud is a viable option today. As an example, in the context of the VMware environment, starting with vSphere 6.0, the vMotion tool for VM migration is expanded to include migrations across vCenters, across virtual switches and across long distances.

The availability of vMotion across vCenters will simultaneously change compute, storage, networks, and management. This will leverage migrations even with unshared storage and will support local-, metro- and cross-continental distances. **EcoInterCloud** exploits such features to implement its policies for inter-DC workload distribution, as reflected in Figure 3.

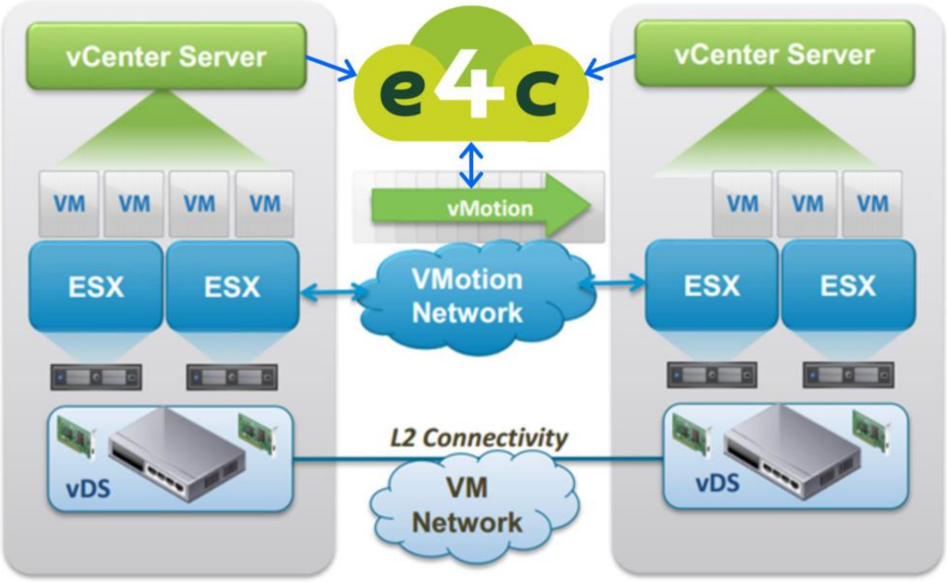


Figure 3 – **EcoInterCloud** in a VMware vSphere 6.0 environment, exploiting long-distance VM migrations through vMotion

Previously, vMotion could only occur within a network managed by a single virtual switch, either a Virtual Standard Switch (VSS) or Virtual Distributed Switch (VDS). vMotion across vCenters now allows VMs to migrate to a target network managed by a different virtual switch effectively switching the networks in a totally seamless way. Long-distance vMotion will support cross-continental distances with up to 100+ms RTTs, while still maintaining standard vMotion guarantees. Some use cases are:

- Disaster avoidance
- SRM and disaster avoidance testing
- Multi-site load balancing and better capacity exploitation (Host Thin Provisioning^{TM})
- Follow-the-moon scenarios

References

[1] C. Mastroianni, M. Meo, G. Papuzzo, "Probabilistic Consolidation of Virtual Machines in Self-Organizing Cloud Data Centers". IEEE Transactions on Cloud Computing, vol. 1, n. 2, pp. 215-228, 2013.

[2] A. Khosravi, S. Garg, and R. Buyya, “Energy and carbon-efficient placement of virtual machines in distributed cloud data centers,” in Euro-Par 2013 Parallel Processing, ser. Lecture Notes in Computer Science, F. Wolf, B. Mohr, and D. Mey, Eds., 2013, vol. 8097.

For more information

Eco4Cloud: www.eco4cloud.com