Resource Allocation in a Network-Based Cloud Computing Environment: Design Challenges

Mohamed Abu Sharkh, Manar Jammal, Abdallah Shami, and Abdelkader Ouda, Western University

ABSTRACT

Cloud computing is a utility computing paradigm that has become a solid base for a wide array of enterprise and end-user applications. Providers offer varying service portfolios that differ in resource configurations and provided services. A comprehensive solution for resource allocation is fundamental to any cloud computing service provider. Any resource allocation model has to consider computational resources as well as network resources to accurately reflect practical demands. Another aspect that should be considered while provisioning resources is energy consumption. This aspect is getting more attention from industrial and government parties. Calls for the support of green clouds are gaining momentum. With that in mind, resource allocation algorithms aim to accomplish the task of scheduling virtual machines on the servers residing in data centers and consequently scheduling network resources while complying with the problem constraints. Several external and internal factors that affect the performance of resource allocation models are introduced in this article. These factors are discussed in detail, and research gaps are pointed out. Design challenges are discussed with the aim of providing a reference to be used when designing a comprehensive energy-aware resource allocation model for cloud computing data centers.

INTRODUCTION

Cloud computing is an increasingly popular computing paradigm, now proving a necessity for utility computing services. Several providers have cloud computing (CC) solutions available, where a pool of virtualized and dynamically scalable computing power, storage, platforms, and services are delivered on demand to clients over the Internet on a pay as you go basis. This is implemented using virtualization technology where clients are just a credit card payment away from scaling their rented virtual machines (VMs) dynamically to include as many machines as they need. The physical location of client data is typically a large data center (DC) accommodating thousands of servers. Clients still have the choice between using private clouds, which are DCs specialized for the internal needs of a certain business organization, and public clouds, which are open for the public to use over the Internet. Services are offered under several deployment models: infrastructure as a service (IaaS), platform as a service (PaaS), software as a service (SaaS), and network as a service (NaaS). Providers offer varying service portfolios that differ in service specification. This includes computational resource configuration of the VMs, the programmer's degree of control, network service configuration, the nature of hardware/software security services, portability guarantees, storage scalability, and so on. To move to the cloud, clients demand guarantees with regard to achieving the required improvements in scale, cost control, and reliability of operations. In spite of its importance, providing computation power alone is not sufficient as a competitive advantage. Other factors have gained more weight recently, such as networking solution offerings. Network performance and resource availability can be the tightest bottleneck for any cloud. This is seen as an opportunity for network service providers who are building their own clouds using distributed cloud architecture.

Here, we see the need for a comprehensive resource allocation (RA) and scheduling system for CC data center networks (DCNs). This system would handle all the resources in the cloud providers' DCN and manage client requests, dictate RA, ensure satisfaction of the network quality of service (QoS) conditions, and eliminate performance hiccups while minimizing service provider cost and controlling the level of energy consumed.

The resource management of the DCs' servers and network resources, while managing new client requests for VMs that are located on these servers, is a crucial success factor. Excess resources can be promoted and sold for additional revenue. Also, with the impact it has on performance, a detailed RA strategy is a key factor that can draw potential clients to the cloud (or a provider) or cause more client reluctance to fully move to the cloud.

Previous RA models can be classified into three categories.

EFFORTS WITH A FOCUS ON DC PROCESSING RESOURCES

Several solutions can be seen in the literature, where resources are scheduled based on user requests. In [1], a queuing model is implemented where a client requests VMs for a fixed time interval. No data communication is assumed between jobs or VMs. The objective is to maximize the computational throughput for a DC. In [2], the objective is to distribute VMs to minimize the distance between their locations in a DC grid. No network constraints are imposed except Euclidean distance between DCs. No specific connection requests or user differentiation is used. Also, a scheduling technique is used to schedule VMs on processors, blades, and racks in a DC, where the objective is to minimize the total communication cost.

EFFORTS WITH A FOCUS ON DC NETWORK RESOURCES

The authors of [3] tackle the scenario where a client has more than one job being processed simultaneously but not sharing the same server. Connection demands are represented as a virtual network (VN) where nodes are VMs and edges are physical network paths. The optimization problem of VN provisioning is solved with the objective of maximizing revenue. No reservation start time or duration was introduced. The case where a user wants to request more connectivity for an already reserved VM is not considered. In [4], the problem of proposing the best virtual network with IP over a wavelength-division multiplexing (WDM) network is considered. The authors generated their constraints according to propagation delay, capacity, and flow conversion constant.

EFFORTS WITH A FOCUS ON ENERGY-EFFICIENT DC RA

Multiple solutions were proposed with the aim of reaching an energy-efficient RA scheme. A common concept is the idea used in [5], which is to consolidate tasks or VMs on the smallest number of servers and then switch the unused servers off or make them idle. The problem is modeled as a bin-packing problem with the assumption that servers are the bins and are full when their resources reach a predefined optimal utilization level. Power consumption by network components is not considered.

Other works took a hardware planning approach to the problem. Instead of targeting the highest performance possible, they aim at executing a certain workload with as little energy as possible. This would not suit cloud clients' needs as this architecture does not support applications with high computational demands.

An economic approach to control shared resources in order to minimize consumed power in hosting centers is proposed in [6]. A solution is presented that dynamically resizes the active servers and responds to the thermal/power supply events by downgrading the service based on the service level agreement (SLA).With the scheduling component already allocating the requests at the lower limit of the SLAs to have enough resources, it will not be easy to find requests that can tolerate downgrades.

NETWORK-AWARE RA: DESIGN CHALLENGES

A COMPREHENSIVE SOLUTION FOR NETWORK AND PROCESSING RA

A comprehensive solution for RA is fundamental to any cloud computing service provider. Any resource allocation model has to consider computational resources as well as network resources to accurately reflect practical demands. Cloud DCs are marketed mainly as ways of outsourcing computational tasks. A successful DC RA model has to provide answers to questions like: What policies are used to allocate VMs? How are processing resources modeled? What resource portfolio is being promoted? How are the servers distributed physically? The other side of the coin is networking resources. When clients execute tasks on the VMs, they need networking service with adequate QoS standards to ensure the successful delivery of their application data.

As reported in [7], only 54 percent of the IT professionals surveyed about their use of cloud services indicated that they involve network operations personnel, down from 62 percent in 2009. This directly affects the use of network best practices and the attention to the health of overall traffic delivery. Also in [7], 28 percent of survey respondents believed that monitoring and troubleshooting packet traces between VMs is required. In addition, 32 percent believed that monitoring and troubleshooting traffic data from virtual switches is required.

Bandwidth costs deeply affect cloud clients' financials. Microsoft Azure, for example, charges clients for downloading based on the exact amount downloaded. Downloading around 950 Gbytes/mo costs the client $113/mo. In comparison, Comcast — the largest Internet provider in the United States — offers a plan with a bandwidth that can download the same amount at $40/month. Azure offers free upload and free data exchange between VMs that are located in the same DC. However, the price difference is an issue clients will consider. Therefore, optimizing the bandwidth cost represents an opportunity of profit for providers and an opportunity of saving for clients.

The network resources weight in the cloud market has alerted network service providers to build their own distributed DCs with a vision to enter the CC market. They envision replacing a large DC with multiple smaller DCs to be closer to the clients. This setup turns the network infrastructure into a distributed cloud. That, in turn, helps in controlling costs and increasing service differentiation.

A cloud service provider caters network services to clients to support one of three functions [8]:

- Connecting the clients' private cloud (or headquarters) to VMs the client reserved in the Dcs using Internet or VPNs, as shown in Fig. 1

A comprehensive solution for RA is fundamental to any cloud computing service provider. Any RA model has to consider computational resources as well as network resources to accurately reflect practical demands.

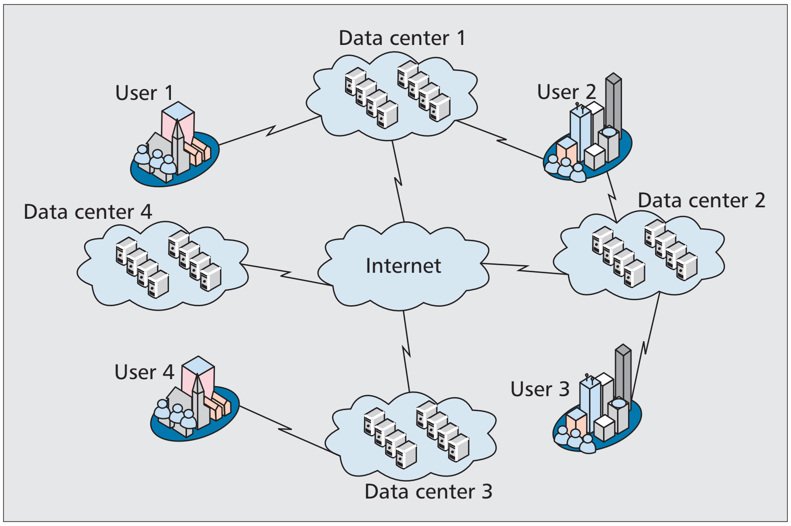

Figure 1. *A sample network of private and public clouds connected through the Internet or virtual private networks (VPNs).*

- Connecting the VMs on different public clouds to facilitate data exchange between two VMs reserved by the same client
- Connecting VMs to each other on the same public cloud

From the client perspective, there is no benefit if results are produced according to the time and quality constraints when these results cannot be delivered in time using a secure and stable network. In [8], when discussing the main obstacles for cloud client base growth, data communication bottlenecks arise as a major challenge. It is shown that when moving large amounts of data in a distributed cloud model, the network service performance is a critical point for the whole process.

Main Design Challenges

Targeting a network-aware RA system brings to the forefront multiple challenges that face the CC community. Addressing those issues is of utmost importance to form a complete solution. These design challenges can be classified into external challenges, which are enforced by factors outside the RA process (illustrated in Fig. 2), and internal challenges, which are related to the RA algorithm (shown in Fig. 3).

External Challenges

Regulative and Geographical Challenges — In the virtualization model used in cloud offerings, the client does not manage the physical location of data. Also, there is no guarantee given by the provider as for the data physical location at a certain moment [9]. In fact, assigning client data to multiple geographically distant DCs is a common practice. Splitting the data will enhance fault tolerance, but it presents regulative and security challenges. An example would be the regulative obligation of complying with the U.S. Health Information Portability and Accountability Act (HIPAA) (the Health Information Protection Act [HIPA] in Canada). Although HIPAA does not apply directly to third-party service providers, it is imperative that health care organizations require third-party providers to sign contracts that require them to handle all patient data in adherence with HIPAA standards. This raises some constraints on handling and storing data:

- Geographical constraints: HIPAA requires that patient data does not leave U.S. soil. This constraint limits the choice of DCs to which to allocate a VM and limits data movement maneuvers while trying to optimize performance. Additionally, when data is stored in the cloud, it is necessary to know the physical location of the data, the number of data copies, data modification details, and data deletion details.
- Client actions: To get more assurance about data security, clients may require guarantees like instant data wiping (writing over byte by byte) instead of deletion. They might also require storing encrypted data on the cloud. This would pose extra pressure on the performance and make it harder to comply with QoS requirements.
- Under HIPAA, patients have the right to access any information stored about them. A careful study of the locations of the patients and the usage distribution of these patients is crucial for the RA system. Considering this factor when placing the data would minimize the distance patient data can travel in the network. Making a decision on where the data is located has a direct effect on minimizing the cost.

Charging Model Issues — The resources management system should incorporate the client's charging model. For example, when using Amazon EC2, a client can pay for the instances completely on demand, reserve an instance for a term contract, or choose spot instances that enable her to bid for unused Amazon EC2 capacity. Issues to be considered here include:

- Finding the service portfolio offering that maximizes the revenue weight of excess resources in the DC. Examining the options available in the market, it is clear that cost is not calculated based on static consumptions.
- Finding the best way to integrate virtual network usage into the cost analysis. Challenges would arise because a virtual link's length/distance (and in turn cost) varies from link to link. A virtual link could even change to use another physical path on the substrate network based on the methodology used.

Internal Challenges

Data Locality: Combining Compute and Data Management — There is a need for systems to implement data locality features "the right way." This involves how to combine the management of compute (processing) and data (network) resources using data locality features to minimize the load of migrated data and in turn enhance the performance/scalability of the application while meeting end users security concerns. Also, managing computational load near the data and recognizing the expenditure of workload migration minimize the data migration bottlenecks.

To have a full view of how to use data locality, these issues need to be considered:

- A data-aware scheduler enhances the scalability and performance of applications. A more specific perspective needs to be reached. This includes answering questions like: How much would the scheduler know at a certain moment? What are the policies and decision criteria for moving data? What data integration policies should be enforced?
- Analyzing the behavior of data-intensive applications is a good starting point to understand data locality and data movement patterns.
- Also, an idea to be evaluated is moving the application itself to servers in the DC where the needed data is. This raises questions about the availability of servers in the other DC, policy/algorithm specifications regarding when to move the application considering that future demand might need data sets that are stored in the original location, and decision criteria regarding whether to migrate the whole VM or just move the concerned application.

Reliability of Network Resources Inside a DC — The DC internal network affects the performance deeply. The DC internal network design decisions affect performance and reliability of the DC resources. These decisions relate to factors like network topology, traffic routing, flow optimization, bandwidth allocation policies, and network virtualization options.

SDN Design Challenges Inside the DCs — Software defined networking (SDN) is a networking paradigm that decouples the forwarding plane from software control. This paradigm can enable many advantages if it is coupled with an efficient RA model. SDN supports agile applications deployment, enhances network and services adaption, and improves their performance. The central network controller aggregates the client's requests in order to manage the RA task in the DCN. The SDN controller executes the RA algorithms, then sends the allocation commands across the network. Figure 4 shows a view of the SDN architecture.

Since it is a relatively new paradigm, the community still has to deeply tackle these issues regarding SDN:

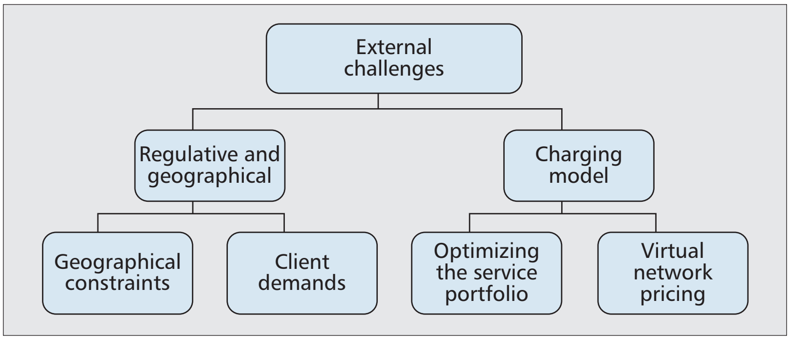

Figure 2. *External challenges.*

- Reliability: Using a centralized SDN controller affects reliability. Although solutions like standby controllers or using multiple controllers for the network are suggested, practical investigation is needed to reveal the problems and analyze the trade-offs of using such solutions.
- Scalability: As the number of switches and end hosts increase in the network, the SDN controller becomes a key bottleneck. For example, [10] estimates that a DC accommodating 2 million VMs may provoke 20 million flows/s. Recently, controllers can support about 10^5 flows/s in the optimal case [12]. Extensive scalability results in losing visibility of the network traffic, making troubleshooting nearly impossible.
- Visibility: Previously, a network team could easily capture and solve, for instance, the reason behind a slow network. However, SDN allows only visibility of a tunnel source and an endpoint with UDP traffic, and hides the user identity. There is no way of determining the origin of the problem. Since the actual user is obscured by the UDP tunnels, determining the problem location becomes impossible for slow networks, and users complain. Losing visibility prohibits troubleshooting, decreases scalability, and increases the resolution delay, which might adversely affect business.
- The controller placement problem affects the performance of the control plane, its fault tolerance, and the state management of the distributed SDN system. This prob-

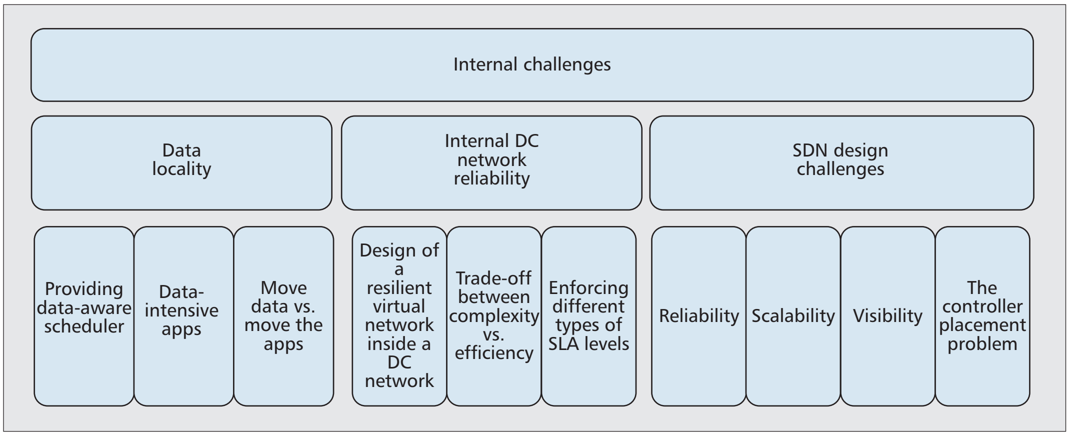

Figure 3. *Internal challenges.*

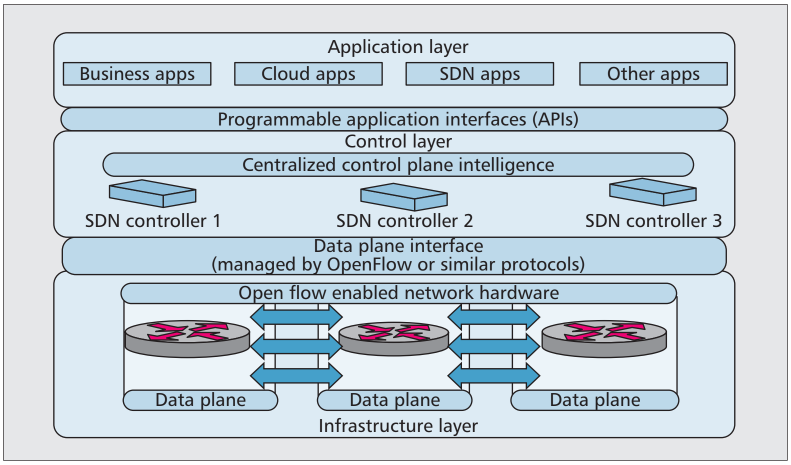

Figure 4. *SDN architecture [10].*

lem includes placement of controllers with respect to the available topology in the network and deciding the number of needed controllers.

The placement is related to certain metrics defined by clients, such as latency and increasing number of nodes. According to [13], random controller placement using small *k* values of the *k*-median problem generates an average latency between 1.4× and 1.7× larger than that of the optimal solution.

Thus, the network service specification affects the clients' movement to the cloud or the selection of their cloud provider. Any inclusive model should acknowledge the effect of bandwidth, port speed, number of IP addresses, load balancing, and VPN availability.

Fault Tolerance vs. Performance — Despite its several applications and wide ranging acceptance, the current CC technology is still prone to hardware, VM, and application failures. Therefore, a stable and efficient fault tolerance (FT) strategy is a crucial requirement to achieve availability, security, and reliability of CC services and real-time applications as well as ensuring seamless task execution.

Due to the complexity and interdependability of FT, implementing it in CC requires delicate analysis and consideration. CC requires autonomic FT policies for instances of VM applications. These techniques must integrate with workflow scheduling algorithms and synchronize among different clouds. Furthermore, CC requires either reactive FT or proactive FT based on the application type and level [14]. Reactive FT techniques, such as restart, replay, and retry, reduce the faults' effect on the application execution. On the other hand, the proactive FT techniques, such as software rejuvenation and preemptive migration, predict faults and errors, and get rid of paralyzed components. Hence, in the context of FT , CC providers aim to implement failure recovery, and cost-aware and performance-effective FT policies.

FT strategy affects how VMs are distributed across fault domains. This distribution often contradicts performance. The challenge here is to find the fault domain definitions and VM distribution that complies with fault tolerance constraints without compromising the performance.

Portability and Vendor Lock-In — This issue is a concern for cloud clients. Clients require guarantees of the applications being portable and easily movable to other cloud providers. This affects VM deployment design and raises a concern for cloud providers regarding the optimal procedure when a certain client leaves. Which RA adjustments are made and how? Here, designing an efficient procedure is a big performance booster. Figure 5 shows where the RA controller lies in the cloud computing architecture. The figure summarizes the main RA functionalities in a CC DC that are performed by the multiple modules of the RA controller.

Energy-Efficient Network-Based RA

As DC number and average size expand, so does the energy consumption. Electricity used by servers doubled between 2000 and 2005, from 12 to 23 billion kWh [15]. This is not only due to the increasing amount of servers per DC; the individual server consumption of energy has increased too. The increase in energy consumption is of major concern to DC owners because of its effect on operational costs. It is also a major concern of governments because of the increase in DCs' carbon footprint. The cloud client base is expanding by the day. This demand will lead to building new DCs, and developing the current ones to include more servers and upgrade the existing servers to have more functionality and use more power. Power-related costs are estimated to represent approximately 50 percent of the DC operational cost, and they are growing faster than other hardware costs [16]. Thus, energy consumption is a major obstacle that would limit the providers' ability to expand. Recently, the response to this fact is seen in the practical landscape as major players in the cloud market are taking more serious steps. Companies as large as Microsoft and Google are aiming to deploy new DCs near cheap power sources to mitigate energy costs [16]. Recently, leading computing service providers have formed a global consortium, the Green Grid, which aims at tackling this challenge by advancing energy efficiency in DCs. This is also pushed by governments in an attempt to decrease the carbon footprints and the effect on climate change. For example, the Japan Data Center Council has been established to mitigate the high energy consumption of DCs in Japan.

A Comprehensive Solution for Energy Efficient Network-based RA

Any model that aims at allocating resources while minimizing energy consumption in a distributed cloud should consider all sources of energy consumption. It should include analysis of power used by CPUs, memory, hard disks, and the power supply unit in a server. An illustration of the power consumption of the possible server components is shown in Fig. 6.

Also, the model should investigate power consumed by network components to transmit data inside and outside the DC. Although the power consumed by a cable or router, for example, is a small percentage of the power consumed by a server rack, the large number of devices that constitute local and global networks consume significant amounts of power. Hardware design optimization is a direction researchers focus on when trying to minimize power consumption. However, the most rewarding concept to save power is to optimize network performance. Moving data using shorter paths and flow optimization causes significant savings. An efficient VM placement technique directly affects the number of network components used per connection. An efficient data-aware scheduler can be the difference between moving data within the same rack, using the local network within a DC, or sending the data to another one across the ocean. Any energy gain from any of these methods is an important achievement since one DC's operational cost impact on the environment is high.

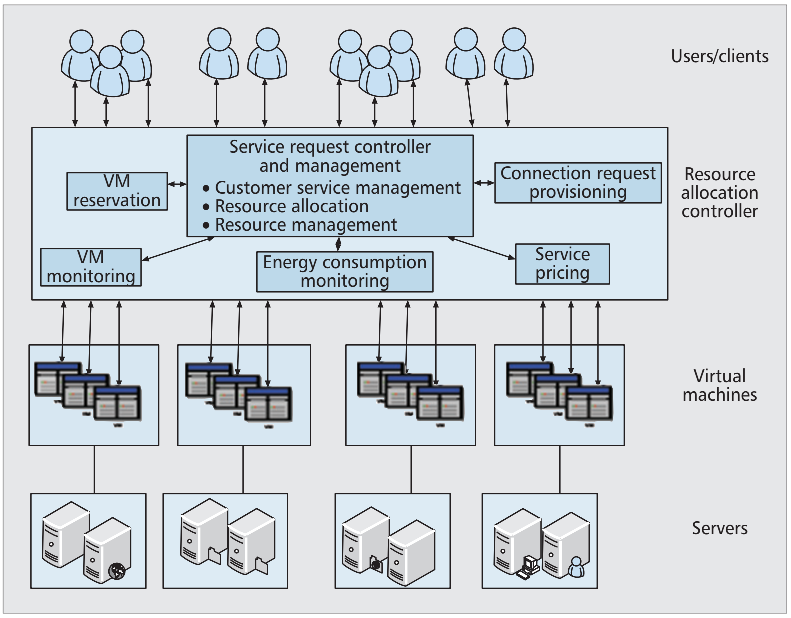

Figure 5. *Cloud computing architecture.*

Common Solutions and Common Trade-offs

- A solution with multiple variations in the literature is consolidation of applications on fewer servers. This concept, despite its positive effect on power consumption, affects performance negatively. There are three main issues here:
  - –Consolidation could quickly cause I/O bottlenecks. Concentration of VMs increases the competition for physical server resources, which threatens the performance as it has a high probability of having I/O bottlenecks. In addition to that, it can cause more power consumption because of the latency in task completion.
  - –Network bottlenecks: Connection blocking would increase visibly as connections from and to all the consolidated VMs compete for the links available to the physical node holding the server. For applications with heavy data transactions, a higher blocking percentage would be found around the servers carrying the consolidated VMs. This would cause even more latency and would consume more network related power.
  - –A method used to hibernate or shut down unused servers should be considered. There is latency and power consumption caused by system hibernation and waking up. If used, consolidation should be part of a more global solution that takes into consideration those issues along with client priorities. In [17], the authors explain the energy waste that happens because of idle servers. "Even at a very low load, the power consumed is over 50 percent of the peak power." This is more apparent when there is a bottleneck since all the other idle resources are wasting power.
- VM migration is the core of the consolidation process. The methodology might differ based on the VM size and configuration variations. Nevertheless, trade-offs have to be considered between the power gained by moving the VM and hibernating the machine it is on and the total losses caused by this migration. These losses include:
  - –Time lost moving the VM through the network
  - –Power consumed by network components during the move
  - –Latency of the task completion caused by the changed node on the network and the need to provision new network resources

Energy Consumption vs. Optimal Performance: Hardware Contradictions

The way processors work currently, higher performance is achieved by maximizing the use of the processor cache memory and minimizing the use of the main memory and disks. In addition, using mechanisms like out-of-order execution, high-speed buses, and support for a large number of pending memory requests increases the transistor counts leads to more wasted power. Thus, the question of the optimal point between performance and power consumption arises.

Cooling Challenges

A considerable amount of the electrical energy utilized by the computing tasks and network resources is converted into thermal energy. This thermal energy reduces the DCs' devices lifetime and affects the system availability negatively. Therefore, dissipation of such energy is a crucial requirement in any cloud infrastructure in order to protect devices from failure and crashes, and maintain them at a safe operating point. As reported in [18], the initial cost of buying/setting up the infrastructure of a DC consisting of 1000 racks is between \$2–\$5 million. However, the cooling system costs annually around \$4–\$8 million. For this reason, software-side optimization might be a promising solution to mitigate the cooling system problem.

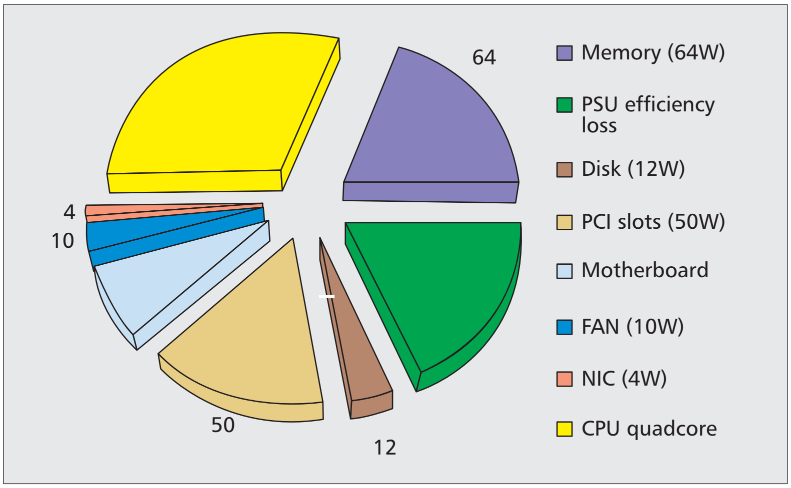

Figure 6. *Server power consumption [15].*

CONCLUSION

Nowadays, CC shows its paramount importance for computing services. To reach a complete RA solution for managing CC DCs, optimizing computational resources, network resources, and energy consumption are the main sides. This article introduces some internal and external factors that affect the design of DC RA models. External challenges are mostly caused by regulative, geographical, and charging model factors. Internal challenges include maximizing the benefits from data locality features. They also include designing a reliable internal DC network. Other internal factors are related to SDN, fault tolerance, and portability. Designing an energy-aware RA model faces performance challenges that are mostly caused by consolidation, VM migration, and server idle state configuration. These design challenges are discussed with the aim of providing a reference to be used when designing a comprehensive energy-aware resource allocation model for CC data centers.

REFERENCES

[1] S. Maguluri, R. Srikant, and L. Ying, "Stochastic Models of Load Balancing and Scheduling in Cloud Computing Clusters," *Proc. IEEE INFOCOM '12*, 25–30 Mar. 2012, pp. 702–10.
[2] M. Alicherry and T.V. Lakshman, "Network Aware Resource Allocation in Distributed Clouds," *Proc. IEEE INFOCOM '12*, 25–30 Mar, 2012, pp. 963–71.
[3] G. Sun *et al.*, "Optimal Provisioning for Elastic Service Oriented Virtual Network Request in Cloud Computing," *IEEE GLOBECOM 2012*, 3–7 Dec. 2012, pp. 2541–46.
[4] B. Kantarci and H.T. Mouftah, "Scheduling Advance Reservation Requests for Wavelength Division Multiplexed Networks with Static Traffic Demands," *IEEE Symp. Computers and Commun.*, 1–4 July 2012, pp. 806–11.
[5] S. Srikantaiah, A. Kansal, and F. Zhao, "Energy Aware Consolidation for Cloud Computing," *Cluster Computing*, vol. 12, 2009, pp. 1–15.
[6] J. S. Chase *et al.*, "Managing Energy and Server Resources in Hosting Centers," *18th ACM Symp. Op. Sys. Principles*, 21 Oct. 2001.
[7] J. Frey, "Network Management and the Responsible, Virtualized Cloud," research rep., Feb. 2011; http://www.enterprisemanagement.com/research/asset.php/1927/Network-Management-and-the-Responsible,-Virtualized-Cloud.
[8] M. Abu Sharkh, A. Ouda, and A. Shami, "A Resource Scheduling Model for Cloud Computing Data Centers," *9th Int'l. Wireless Commun. & Mobile Computing Conf.*, 1–5 July 2013, Cagliari, Sardinia, Italy.
[9] M. Armbrust *et al.*, "Above the Clouds: A Berkeley View of Cloud Computing," tech. rep. UCB/EECS-2009-28, EECS Dept., UC Berkeley, Feb 2009.
[10] M. Jammal *et al.*, "Software Defined Networking: A Survey," submitted to *IEEE Commun. Surveys and Tutorials*, July 2013.
[11] A. Tavakoli *et al.*, "Applying NOX to the Datacenter," *Proc. HotNets*, Oct 2009.
[12] Pluribus Networks, "Of Controllers and Why Nicira Had to Do a Deal, Part III: SDN and Openflow Enabling Network Virtualization in the Cloud)," Aug. 2012; http://pluribusnetworks.com/blog/.
[13] B. Heller, R. Sherwood and .,N. McKeown, "The Controller Placement Problem," *Proc. 1st Wksp. Hot Topics in Software-Defined Networks*, Aug. 2012, pp. 7–12.
[14] A. Tchana, L. Broto, and D. Hagimont,, "Approaches to Cloud Computing Fault Tolerance," *2012 Int'l. Conf. Computer, Info. and Telecommun. Sys.*, 14–16 May 2012, pp. 1, 6.
[15] L. Minas and B. Ellison, "The Problem of Power Consumption in Servers," prepared in Intel Lab, *Dr. Dobbs J.*, Mar 2009.
[16] B.G. Chun *et al.*, "An Energy Case for Hybrid Datacenters," *ACM SIGOPS Op. Sys. Rev.*, vol. 44, no. 1, Jan. 2010.
[17] G Chen *et al.*, "Energy-Aware Server Provisioning and Load Dispatching for Connection-Intensive Internet Services," *Proc. 5th USENIX Symp. Networked Sys. Design and Implementation*, 16–18 Apr. 2008, pp. 337–50.
[18] C. D. Patel, C. E. Bash, and A.H. Beitelmal, "Smart Cooling of Data Centers," Hewlett-Packard Development Company L.P. Patents, 2003

BIOGRAPHIES

MOHAMED ABU SHARKH received his B.Sc. degree in computer science from the Faculty of Science, Kuwait University, in 2005, and his M.Sc. eegree in computer engineering from the Faculty of Engineering and Petroleum, Kuwait University, in 2009. He has five years of professional experience as a software engineer and business analyst, then as an ERP consultant. Since January 2012, he has been with Western University, London, Ontario, Canada, where he is currently a research assistant and a Ph.D. candidate in the Department of Electrical and Computer Engineering. His current research interests include cloud computing data center management and software engineering.

MANAR JAMMAL received her M.S. degree in electrical and electronics engineering from the Lebanese University, Beirut, in collaboration with the University of Technology of Compiègne in 2012. She is currently working toward her Ph.D. degree in cloud computing and virtualization technology at Western University. Her research interests include cloud computing, virtualization, software defined networking, and virtual machine migrations

ABDALLAH SHAMI [SM] received his B.E. degree in electrical and computer engineering from the Lebanese University in 1997 and his Ph.D. degree in electrical engineering from the Graduate School and University Center, City University of New York, in September 2002. In September 2002, he joined the Department of Electrical Engineering at Lakehead University, Thunder Bay, Ontario, Canada as an assistant professor. Since July 2004, he has been with Western University, where he is currently an associate professor in the Department of Electrical and Computer Engineering. His current research interests are in the areas of network optimization, cloud computing, and wireless networks. He is an Editor for *IEEE Communications Tutorials and Surveys* and has served on the Editorial Board of *IEEE Communications Letters* (2008–2013). He has chaired key symposia for IEEE GLOBECOM, IEEE ICC, IEEE ICNC, and ICCIT.

ABDELKADER OUDA is an assistant professor in the Department of Electrical and Computer Engineering at Western University. He received his B.Sc. (Hons) and M.Sc. degrees in computer science and pure mathematics from Ain Shams University, Cairo, Egypt, in 1986 and 1993, respectively, and his Ph.D. degree in computer science from Western University in 2005. He has been studying, researching, and teaching software engineering at Western University for 12 years. His current research is mainly in information and networking security. He worked for over 14 years in the software development industry, involving project management, system analysis/consulting, and database design.