ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

Force-Sentient Consignment Matching and Appliance Scaling For the Cloud Bionetwork

K.Saranya[1], C.Anitha[2], T.R.Vithya [3], K.K.Kavitha [4]

Research Scholar, Dept. of Computer Science, Selvamm Arts & Science College, Tamilnadu, India [1]

Asst. Professor, Dept. of Computer Science, Selvamm Arts & Science College (Autonomous), Tamilnadu, India [2]

Asst. Professor, Dept. of Computer Science, Selvamm Arts & Science College (Autonomous), Tamilnadu, India [3]

HOD&Vice Principal, Department of Computer Science, Selvamm Arts & Science College (Autonomous), Tamilnadu, India [4]

ABSTRACT: A new model of cloud servers that is based on different operating regimes with various degrees of energy efficiency" (processing power versus energy consumption);a novel algorithm that performs load balancing and application scaling to maximize the number of servers operating in the energy-optimal regime; and analysis and comparison of techniques for load balancing and application scaling using three differently-sized clusters and two different average load profiles. The objective of the algorithms is to ensure that the largest possible number of active servers operate within the boundaries of their respective optimal operating regime. The actions implementing this policy are: (a) migrate VMs from a server operating in the undesirable-low regime and then switch the server to a sleep state; (b) switch an idle server to a sleep state and reactivate servers in a sleep state when the cluster load increases; (c) migrate the VMs from an overloaded server, a server operating in the undesirable-high regime with applications predicted to increase their demands for computing in the next reallocation cycles.

KEYWORDS: load balancing, application scaling, idle servers, server consolidation, energy proportional systems.

I. INTRODUCTION

Warehouse-scale computers (WSCs) are the building blocks of a cloud infrastructure. A hierarchy of networks connects 50; 000 to 100; 000 servers in a WSC. The servers are housed in racks; typically, the 48 servers in a rack are connected by a 48-port Gigabit Ethernet switch. The switch has two to eight up-links which go to the higher level switches in the network hierarchy. Cloud elasticity, the ability to use as many resources as needed at any given time, and low cost, a user is charged only for the resources it consumes, represents solid incentives for many organizations to migrate their computational activities to a public cloud. The number of CSPs, the spectrum of services offered by the CSPs, and the number of cloud users have increased dramatically during the last few years. For example, in 2007 the EC2 (Elastic Computing Cloud) was the first service provided by AWS (Amazon Web Services); five years later, in 2012, AWS was used by businesses in 200 countries. Amazon's S3 (Simple Storage Service) has surpassed two trillion objects and routinely runs more than 1.1 million peak requests per second. Elastic Map Reduce has launched 5:5 million clusters since May 2010 when the service started . The rapid expansion of the cloud computing has a significant impact on the energy consumption in US and the world. The costs for energy and for cooling large-scale data centers are significant and are expected to increase in the future. In 2006, the 6 000 data centers in the U.S. reportedly consumed 61 _ 109 kWh of energy, 1:5% of all electricity consumption in the country, at a cost of $4:5 billion [34]. The energy consumption of data centers and of the network infrastructure is predicted to reach 10; 300 TWh/year (1 TWh = 109 kWh) in 2030, based on 2010 efficiency levels [28]. These increases are expected in spite of the extraordinary reduction in energy requirements for computing activities.

Idle and under-utilized servers contribute significantly to wasted energy, see Section 2. A 2010 survey reports that idle servers contribute 11 million tons of unnecessary CO2 emissions each year and that the total yearly costs for idle servers is $19 billion. Recently, Gartner Research reported that the average server utilization in large data-centers is 18%, while the utilization of x86 servers is even lower, 12%. These results confirm earlier estimations that the average server utilization is in the 30% range. The concept of \load balancing" dates back to the time when the first

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

distributed computing systems were implemented. It means exactly what the name implies, to evenly distribute the workload to a set of servers to maximize the throughput, minimize the response time, and increase the system resilience to faults by avoiding overloading the systems. An important strategy for energy reduction is concentrating the load on a subset of servers and, whenever possible, switching the rest of them to a state with a low energy consumption. This observation implies that the traditional concept of load balancing in a large-scale system could be reformulated as follows: distribute evenly the workload to the smallest set of servers operating at optimal or near-optimal energy levels, while observing the Service Level Agreement (SLA) between the CSP and a cloud user. An optimal energy level is one when the performance per Watt of power is maximized.

II. EXISTING SYSTEM

An important strategy for energy reduction is concentrating the load on a subset of servers and, whenever possible, switching the rest of them to a state with low energy consumption. This observation implies that the traditional concept of load balancing in a large-scale system could be reformulated as follows: distribute evenly the workload to the smallest set of servers operating at optimal or near-optimal energy levels, while observing the Service Level Agreement (SLA) between the CSP and a cloud user. An optimal energy level is one when the performance per Watt of power is maximized.

In order to integrate business requirements and application level needs, in terms of Quality of Service (QoS), cloud service provisioning is regulated by Service Level Agreements (SLAs): contracts between clients and providers that express the price for a service, the QoS levels required during the service provisioning, and the penalties associated with the SLA violations. In such a context, performance evaluation plays a key role allowing system managers to evaluate the effects of different resource management strategies on the data center functioning and to predict the corresponding costs/benefits.

Drawbacks of Existing System

- On-the-field experiments are mainly focused on the offered QoS, they are based on a black box approach that makes difficult to correlate obtained data to the internal resource management strategies implemented by the system provider.
- Simulation does not allow to conduct comprehensive analyses of the system performance due to the great number of parameters that have to be investigated.

III. PROPOSED SYSTEM

There are three primary contributions of this paper:

A new model of cloud servers that is based on different operating regimes with various degrees of \energy efficiency" (processing power versus energy consumption);

A novel algorithm that performs load balancing and application scaling to maximize the number of servers operating in the energy-optimal regime; and analysis and comparison of techniques for load balancing and application scaling using three differently-sized clusters and two different average load profiles.

The objective of the algorithms is to ensure that the largest possible number of active servers operate within the boundaries of their respective optimal operating regime. The actions implementing this policy are: (a) migrate VMs from a server operating in the undesirable-low regime and then switch the server to a sleep state; (b) switch an idle server to a sleep state and reactivate servers in a sleep state when the cluster load increases; (c) migrate the VMs from an overloaded server, a server operating in the undesirable-high regime with applications predicted to increase their demands for computing in the next reallocation cycles.

Advantages of Proposed System

- After load balancing, the number of servers in the optimal regime increases from 0 to about 60% and a fair number of servers are switched to the sleep state.
- There is a balance between computational efficiency and SLA violations; the algorithm can be tuned to maximize computational efficiency or to minimize SLA violations according to the type of workload and the system management policies.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

System Architecture

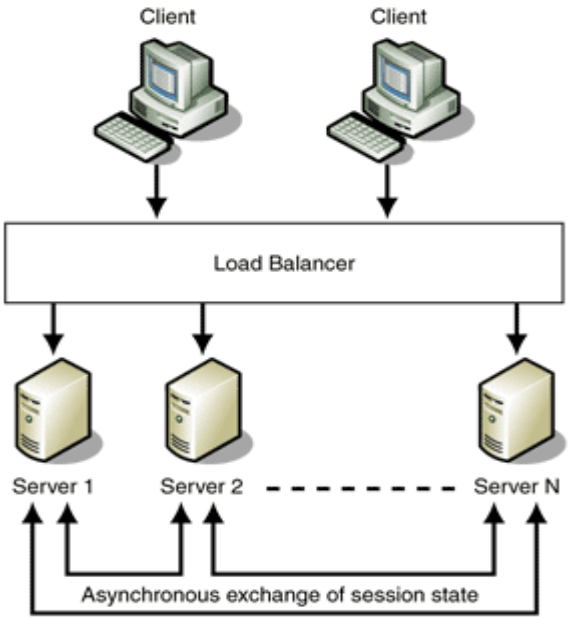

Fig 3.1: Architecture Diagram

IV. IMPLEMENTATION

System Model

In this module, we design the system, such that client makes request to server. Usually, a it is designed with adequate resources in order to satisfy the traffic volume generated by end-users. In general, a wise provisioning of resources can ensure that the input rate is always lower than the service rate. In such a case, the system will be capable to efficiently serve all users' requests. Applications for one instance family have similar profiles, e.g., are CPU-, memory-, or I/O-intensive and run on clusters optimized for that profile; thus, the application interference with one another is minimized. The normalized system performance and the normalized power consumption are different from server to server; yet, warehouse scale computers supporting an instance family use the same processor or family of processors and this reduces the effort to determine the parameters required by our model. In our model the migration decisions are based solely on the CPU units demanded by an application and the available capacity of the host and of the other servers in the cluster. The model could be extended to take into account not only the processing power, but also the dominant resource for a particular instance family, e.g., memory for R3, storage for I2, GPU for G2 when deciding to migrate a VM. This extension would complicate the model and add additional overhead for monitoring the application behavior.

Server

The term server consolidation is used to describe: switching idle and lightly loaded systems to a sleep state; (2) workload migration to prevent overloading of systems; or (3) any optimization of cloud performance and energy efficiency by redistributing the workload. In this module we design the Server System, where the server processes the client request. Cloud is a large distributed system of servers deployed in multiple data centers across the Internet. The goal of a cloud is to serve content to end-users with high availability and high performance. Cloud serves a large fraction of the Internet content today, including web objects (text, graphics and scripts), downloadable objects (media files, software, documents), applications (e-commerce, portals), live streaming media, on-demand streaming media, and social networks. Besides better performance and availability, cloud also offload the traffic served directly from the content provider's origin infrastructure, resulting in cost savings for the content provider.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

Creating Load

In this module, we create the load to the server. Though, in this paper we focus exclusively on critical conditions where the global resources of the network are close to saturation. This is a realistic assumption since an unusual traffic condition characterized by a high volume of requests, i.e., a flash crowd, can always overfills the available system capacity. In such a situation, the servers are not all overloaded. Rather, we typically have local instability conditions where the input rate is greater than the service rate. In this case, the balancing algorithm helps prevent a local instability condition by redistributing the excess load to less loaded servers.

Energy Aware Load balance

The objective of the algorithms is to ensure that the largest possible number of active servers operate within the boundaries of their respective optimal operating regime. The actions implementing this policy are: (a) migrate VMs from a server operating in the undesirable-low regime and then switch the server to a sleep state; (b) switch an idle server to a sleep state and reactivate servers in a sleep state when the cluster load increases; (c) migrate the VMs from an overloaded server, a server operating in the undesirable-high regime with applications predicted to increase their demands for computing in the next reallocation cycles. We present a new mechanism for redirecting incoming client requests to the most appropriate server, thus balancing the overall system requests load. Our mechanism leverages local balancing in order to achieve global balancing. This is carried out through a periodic interaction among the system nodes. Depending on the network layers and mechanisms involved in the process, generally request routing techniques can be classified in cloud request routing, transport-layer request routing, and application-layer request routing.

VI. CONCLUSION

The lazy approach would eliminate this effect. Even the definition of an ideal case when a clairvoyant resource manager makes optimal decisions based not only on the past history, but also on the knowledge of the future can be controversial. For example, we choose as the ideal case the one when all servers operate at the upper boundary of the optimal regime; other choices for the ideal case and for the bounds of the have regimes could be considered in case of fast varying, or unpredictable workloads. The have-regime model introduced in this paper reflects the need for a balanced strategy allowing a server to operate in an optimal or near-optimal regime for the longest period of time feasible. A server operating in the optimal regime is unlikely to request a VM migration in the immediate future and to cause an SLA violation, one in a sub-optimal regime is more likely to request a VM migration, while one in the undesirable high regime is very likely to require VM migration. Servers in the undesirable-low regime should be switched to a sleep state as soon as feasible. The model is designed for clusters built with the same type of processors and similar configurations; the few parameters of the model are then the same for all the servers in the cluster. The clustered organization allows an effective management of servers in the sleep state as they should be switched proactively to a running state to avoid SLA violations. It also supports effective admission control, capacity allocation, and load balancing mechanisms as the cluster leader has relatively accurate information about the available capacity of individual servers in the cluster. Typically, we see a transient period when most scaling decisions require VM migration, but in a steady-state, local decisions become dominant.

REFERENCES

[1] D. Ardagna, B. Panicucci, M. Trubian, and L. Zhang. \Energy-aware autonomic resource allocation in multitier virtualized environments." IEEE Trans. on Services Computing, 5(1):2{19, 2012.
[2] J. Baliga, R.W.A. Ayre, K. Hinton, and R.S. Tucker \Green cloud computing: balancing energy in processing, storage, and transport." Proc. IEEE, 99(1):149-167, 2011.
[3] L. A. Barroso and U. H ozle. \The case for energyproportional computing." IEEE Computer, 40(12):33{ 37, 2007.
[4] L. A. Barossso, J. Clidaras, and U.H ozle. The Data- center as a Computer; an Introduction to the Design of Warehouse-Scale Machines. (Second Edition). Morgan & Claypool, 2013.
[5] A.Beloglazov,R.Buyya\Energye cient resource management in virtualized cloud data centers." Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Comp., 2010.
[6] A. Beloglazov, J. Abawajy, R. Buyya. \Energy-aware resource allocation heuristicsfore cient management of data centers for Cloud computing." Future Generation Computer Systems, 28(5):755-768, 2012.
[7] A. Beloglazov and R. Buyya. \Managing overloaded hosts for dynamic consolidation on virtual machines in cloud centers under quality of service constraints." IEEE Trans. on Parallel and Distributed Systems, 24(7):1366- 1379, 2013.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

[8] M. Blackburn and A. Hawkins. \Unused server survey results analysis." www.thegreengrid.org/media/White Papers/Unused%20Server%20StudWP 101910 v1. ashx?lang=en (Accessed on December 6, 2013).
[9] M. Elhawary and Z. J. Haas. \Energy-efficient protocol for cooperative networks." IEEE/ACM Trans. on Net- working, 19(2):561{574, 2011.
[10] A. Gandhi, M. Harchol-Balter, R. Raghunathan, and M.Kozuch. \AutoScale: dynamic, robust capacity management for multi-tier data centers." ACM Trans. On Computer Systems, 30(4):1{26, 2012.
[11] A. Gandhi, M. Harchol-Balter, R. Raghunathan, and M.Kozuch. Are sleep states ective in data centers?" Proc. Int. Conf. on Green Comp., pp. 1{10, 2012.
[12] D. Gmach, J. Rolia, L. Cherkasova, G. Belrose, T. Tucricchi, and A. Kemper. \An integrated approach to resource pool management: policies, e_ciency, and quality metrics." Proc. Int. Conf. on Dependable Systems and Networks, pp. 326{335, 2008
[13] Google. \Google's green computing: efficiency at scale."http://static.googleusercontent.com/external content/untrusted dlcp/www.google.com/en/us/green/pdfs/google-green-computing.pdf (Accessed on August 29, 2013).
[14] V. Gupta and M. Harchol-Balter. \Self-adaptive admission control policies for resource-sharing systems." Proc. 11th Int. Joint Conf. Measurement and Modeling Computer Systems (SIGMETRICS'09), pp. 311{322, 2009.
[15] K. Hasebe, T. Niwa, A. Sugiki, and K. Kato. \Powersaving in large-scale storage systems with data migration." Proc IEEE 2nd Int. Conf. on Cloud Comp. Technology and Science, pp. 266{273, 2010.