Cloud Computing Challenges and Opportunities: A Survey

[1]Faraz Fatemi Moghaddam
Faculty of Computer Science and IT
U.P.M. University
Selangor, Malaysia
f.fatemi@ieee.org

[2]Mohammad Ahmadi
Faculty of Computing
Asia Pacific University of Technology and Innovation
Kuala Lumpur, Malaysia
ahmadi@apu.edu.my

[3]Samira Sarvari
Faculty of Computer Science and IT
U.P.M. University
Selangor, Malaysia
samirasarvari82@yahoo.com

[4]Mohammad Eslami
Dept. of Electrical & Computer,
College of Engineering, Islamic
Azad University, Zahedan, Iran
mohammad.eslami@chmail.ir

[5]Ali Golkar
Dept. of Computer Science & IT
Shiraz University
Shiraz, Iran
agolkar@cse.shirazu.ac.ir

***Abstract*—The approval rating of cloud computing as an emerging technology has been enhanced significantly and these days, there are many cloud storage and computing providers who offer their services regarding IaaS, PaaS, and SaaS. Despite these considerable benefits, there are serious concerns and challenges about this new technology. The most important issue is related to security and privacy subjects in cloud-based environments. Furthermore, resource allocation, load balancing, data management, data availability, scalability, compatibility and interoperability are the other challenges in cloud-based environments that decrease efficiency and reliability of this technology. In this paper, challenges and concerns related to cloud-based environments have been identified and most appropriate current solutions for each challenge have been described.**

Keywords—cloud computing; security; resource allocation; load balancing; data management; availability.

I. INTRODUCTION

Cloud Computing is an emerging technology with a rapid growth in IT industry by taking advantages of various concepts and sub-technologies such as virtualization, processing power, storage, sharing, distributed networks and connectivity. These days, cloud-based services have become as a leading on-demand services and unlimited storages for users and enterprises [1].

Despite the considerable benefits of cloud computing environments, there are some serious concerns that have affected the reliability and efficiency of this modern and ongoing concept [2]. Numerous challenging issues face cloud computing and have attracted the attention of many researchers and service providers. Hence, the most important challenges and issues in cloud computing environments have been classified and reviewed in this paper to provide a better understanding of the vision of cloud computing and the challenges ahead for further researches.

II. CLOUD COMPUTING CHALLENGES

In overall, cloud-based challenges have been classified to six main categories: Data Management and Resource Allocation, Security and Privacy, Load Balancing, Scalability and Availability, Migration to Clouds and Compatibility, Interoperability and Communication between Clouds. Each of these issues has affected the reliability and efficiency of cloud-based environments regarding to their concepts.

A. Data Management and Resource Allocation

One of the most important concepts in cloud computing data centers is resource allocation. The importance of this concept is specified according to the large number of resources in cloud-based environments. Accordingly, resource allocation process should fulfill network quality of service requirements, eliminate performance hiccups without significant enhancement of service provider cost, and manage energy consumption [3]. Challenges in resource allocations are classified into three major parts: data center network resources, data center processing resources, and energy efficient data center resource allocation.

The first part focuses on scheduling resources according to user requests. Therefore, the computational throughput in cloud-based data center should be as minimized as possible [4] and the distance between the locations of virtual machines in cloud data center should be decreased by distributing virtual machines [5].

The most challenging issue in data center network resources is optimizing virtual network provisioning while maximizing the revenue. Moreover, the process of selecting the

978-1-4799-7315-6/15/$31.00 ©2015 IEEE

most appropriate virtual network with IP over the cloud-based network is a considerable issue that has been considered regarding to propagation, delay, and flow conversion constant [6].

Furthermore, establishment of an energy efficient data center resource allocation is an exigent requirement in cloud-based data centers for minimizing power consumption in cloud data centers or reducing number of used servers by amalgamating tasks and virtual machines on the minimized number of servers.

B. Security and Privacy

One of the most challenging issues that decrease the rate of reliability and efficiency in cloud computing environments is ensuring the security and privacy of stored resources. Cloud computing security has become an important topic in industry and academic research and has become the leading cause of impeding its development [7].

Typically, security issues in cloud-based environments have been divided [2] to three main parts: vulnerability to attack, standard security practices and being subject to state or national data-storage laws related to privacy or record keeping. These issues led to the appearance of considerable concerns in various levels (*i.e.* service providers, infrastructure, and end-users). Hence, more detailed security concerns have been shown in table 1 with reference to defined levels.

According to table 1, user authentication and access control processes are the most challenging issues in cloud-based environments, which cause the appearance of notable concerns in both levels of service providers and end-users. Therefore, many studies and researches have been performed to improve the efficiency and reliability of managing accesses and authentications, but there has been a slight increase in the results of these researches in comparison with the rapid growth of cloud computing communications.

Applying agent-based authentication system [8] and multi-factor authentication process [9] are two common solutions for increasing the reliability of authentication process. Furthermore, ensuring the security of resources during transmission processes between providers and tenants is a challenging issue that is considered by researchers and service providers.

The other important security issue is the ability of resistance in cloud-based servers against possible attacks or unpredictable events. A significant resistance may improve the process of data protection by service providers and enhance the rate of reliability in cloud computing environments. Applying real-time cryptography algorithms may be the most appropriate solution for enhancement of resistance against possible attacks or unpredictable events [10].

C. Load Balancing

Load balancing is an important issue in cloud computing environment that is related to storage utilization and download performance. The main objective in this case is to establish an algorithm for assigning tasks to the cloud nodes effectively according to exiting limitations [11] (*e.g.* high communication delays and heterogeneity). Typically, challenges and issues related to load balancing in cloud computing environments are classified to four major parts [12]: spatial distribution of the cloud nodes, data replication, performance, and point of failure.

The first issue is the way of managing load balancing process between all spatially distributed cloud nodes. This management should consider several delays that can possibly happen because of the distance between clients and task processing nodes, speed of the network links between cloud nodes, and the distance between service nodes [13].

The second problem is the way the data is replicated (*i.e.* full and partial) in several nodes according to complexity of load balancing algorithm in various clouds nodes during partial replication and the additional requirement for more storage in full replication. The next issue is related to the performance of load balancing algorithm.

Due to the nature of load balancing processes, the complexity of the designed algorithm should be as less as possible to avoid faults and delays during intricate processes. The last challenge is the ability of resistance in load balancing algorithms against usual or unpredictable failures. Using a controller function to manage and minimize failures due to the increasing complexity of load balancing algorithms is a challenging issue in cloud based environments that attracts researchers' considerations.

TABLE I: SECURITY CONCERNS IN CLOUD COMPUTING ENVIRONMENTS

Service Provider	***Infrastructure***	***End-User***
Identity and User Authentication	Data-Storages	Data Protection
Privacy and Access Control	Network Hardware	Loss of Governance
Data Transmission	Other Hardware	Managing Accesses
Possible Attacks		Reliable Authentication
Unpredictable Events		Browser Security
Supporting Processes		Visibility of Data
Third-Party Applications		

978-1-4799-7315-6/15/$31.00 ©2015 IEEE

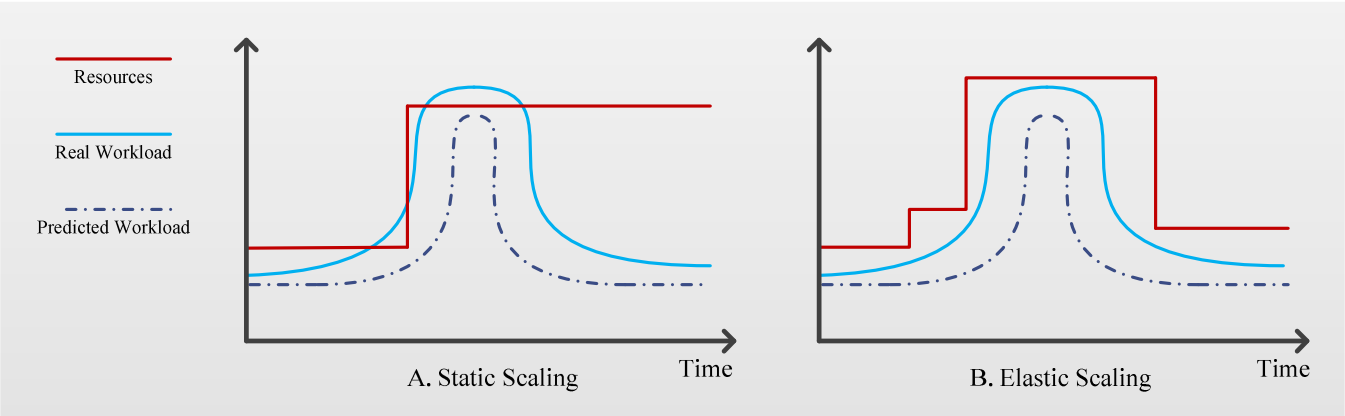

Fig. 1. Static and Elastic Resource Scaling in Cloud Computing

D. Scalability and Availability

The ability of adapting cloud capacity to on-demand services during occurrence of various workloads (*i.e.* static, periodic, once-in-a-lifetime, unpredictable, and continuously changing workloads) is a challenging issue in offering cloud-based services. The lack of this ability may cause performance degradation (in the peak of workload) or even oversizing (in the bottom of workload) during the provision of on-demand services [14].

Elastic resource scaling is the most appropriate solution against possible workloads in cloud computing environment to make resource provisioning more flexible in comparison with static scaling and decrease the dependency of this process on workload predictions [15].

The difference between elastic and static resource scaling has been shown in Fig. 1. According to Fig. 1A, resource provisioning has started according to the predicted workload and finished after reaching the predicted workload peak. This type of scaling has a significant weakness against once-in-a-lifetime and unpredictable workloads and may cause system failure during resource provisioning when experienced workload is more than predicted workload.

Apart from that, in elastic resource scaling (Fig. 1B) new resources are provisioned in small intervals when the increase is detected. Moreover, the process of provisioning is terminated regardless of reaching the predicted workload peak. Elastic resource scaling can handle all types of workload. However, the complexity of elastic resource scaling algorithms may decrease the performance of resource provisioning processes especially during unpredicted workloads.

E. Migration to Clouds and Compatibility

The rapid growth and the popularity of cloud computing between users and enterprises has encouraged traditional IT providers to move and adapt their products (*e.g.* traditional applications, operating systems, middleware, etc.) to cloud-based environments.

However, the possibility of success in this migration process is a challenging issue due to the existing limitations in traditional IT products. Typically, an IT product should contain five main specifications to enhance the rate of success in migration process [16]:

- **Modularity**: The ability of scaling up or scaling down the computing capability according to similar and replicated components on numerous virtual or physical nodes.
- **Portability**: The ability of implementing components in various IaaS infrastructure.
- **Changeability**: The ability to adapt the IT product with appropriate changes in cloud-based environments.
- **Scalability**: The ability to manage numerous workloads.
- **Backward Compatibility**: The ability for constant interaction of components when they are exposed in form of representational state transfer URLs.

The process of migration involves four main steps. In the first step components should be evaluated regarding scalability requirements and the rate of their dependency on persistence related functionality. Furthermore, the orchestration of system should be investigated in the second steps to meet appropriate interaction requirements between various components in virtual or physical nodes. In the third step, components that could not meet sufficient scalability requirements should be identified and be re-factored by decomposing to smaller components. Finally, the compatibility and adaptation of components against various cloud computing environments (*i.e.* public, private, and hybrid) should be considered and appropriate changes should be concluded.

Migration to cloud-based services has considerable benefit for users and enterprises such as enhancing the rate of satisfaction in work, providing proper opportunities for

978-1-4799-7315-6/15/$31.00 ©2015 IEEE

organizational growth and new skills development, and declining monotonous work.

Nevertheless, there are some important concerns during the process of migration such as increasing dependency of enterprises on external third party, departmental downsizing, lack of understanding about cloud features and structure, lack of supporting resources, and uncertainty in new technology [17]. Hence, all possibilities and highlights should be considered by enterprises and providers during the process of migration and after that.

F. Interoperability and Communication Between Clouds

The lack of interoperability between various cloud vendors is an important issue that has happened due to disparate approaches and structure between them.

This deficiency may happen in different level of cloud-based environments (*e.g.* when an infrastructure-as-a-service environment cannot be moved to any platform-as-a-service provider effortlessly) or in a same level between providers (*e.g.* when an Amazon tenant cannot simply move his resources to Force.com) [18].

Establishment of interoperability between clouds are classified to four main levels [19]:

- **Agreement**: Using specified data format or communication protocols or using common information exchange reference model.
- **Adaptation**: Using same standards to make products that decrease competitive strength between service providers.
- **Deployment**: Establishing horizontal or vertical interoperability for providing services at the same or different deployment level.
- **Interaction**: Using synchronous patterns when the response time is critical in real-time services or using asynchronous patterns by decreasing the interaction between services.

Each level of interoperability has some drawbacks that makes the process of achieving interoperable cloud computing environments more challengeable.

III. CONCLUSION

In recent years, the approval rating of cloud computing as an emerging technology has been enhanced significantly and these days, there are many cloud storage and computing providers who offer their services regarding IaaS, PaaS, and SaaS.

Despite these considerable benefits, there are serious concerns and challenges about this new technology. The most important issue is related to security and privacy subjects in cloud-based environments. Furthermore, resource allocation, load balancing, data management, data availability, scalability, compatibility and interoperability are the other challenges in cloud-based environments that decrease efficiency and reliability of this technology.

Accordingly, challenges and concerns related to cloud-based environments have been identified and most appropriate current solutions for each challenge have been described.

In Overall, it is anticipated that cloud computing will be the most important and challenging issue in IT industry. As regarding the importance of this technology, we hope this paper will contribute to a better understanding of the vision of cloud computing and the challenges ahead for further researches.

ACKNOWLEDGMENT

This work is supported by Nero Cloud Research & Development Institute (NCRD-14-AC-3650 Grant Scheme) and Department of Computer Science and Information Technology, University Putra Malaysia (UPM). We gratefully acknowledge the generous assistance provided by Dr. Pardis Najafi, Ms. Fatemeh Afsahi, and the bright memory of Dr. Enayat Fatemi Moghaddam.

REFERENCES

[1] M. Ahmadi, N. Khanezaei, S. Manavi, F. Fatemi Moghaddam, and T. Khodadadi, "A Comparative Study of Time Management and Energy Consumption in Mobile Cloud Computing," in *Proc. of IEEE 5th Control and System Graduate Research Colloquium (ICSGRC)*, Shah Alam, 2014, pp. 199–203.

[2] X. Tan, "The Issues of Cloud Computing Security in High- Speed Railway," in Proc. of International Conference on Electronic and Mechanical Engineering and Information Technology (EMEIT), 2011, pp. 4358-4363.

[3] M.A. Sharkh, M. Jammal, A. Shami, and A. Ouda, "Resource Allocation in a Network-Based Cloud Computing Environment: Design Challenges," *IEEE Communications Magazine*, vol. 51, no. 11, pp. 46-52, November 2013.

[4] S. Maguluri, R. Srikant, and L. Ying, "Stochastic Models of Load Balancing and Scheduling in Cloud Computing Clusters," in *Proc. of IEEE INFOCOM*, Orlando, FL, 2012, pp. 702–710.

[5] M. Alicherry and T.V. Lakshman, "Network Aware Resource Allocation in Distributed Clouds," in *Proc. of IEEE INFOCOM*, Orlando, FL, 2012, pp. 963–971.

[6] T.D. Wallace, A. Shami, and C. Assi, "Scheduling Advance Reservation Requests for Wavelength Division Multiplexed Networks with Static Traffic Demands," *IET Communications*, vol. 2, no. 8, pp. 1023-1033, September 2008.

[7] S. Chaves, C. Westphall, C. Westphall, and G. Geronimo, "Customer Security Concerns in Cloud Computing," in *Proc. of 10th International Conference on Networks (ICN)*, The Netherlands, 2011, pp. 7-11.

[8] M. Hajivali, F. Fatemi Moghaddam, M.T. Alrashdan, and A.Z.M. Alothmani, "Applying an Agent-Based User Authentication and Access Control Model for Cloud Servers," in *Proc. of International Conference*

on ICT Convergence (ICTC), Jeju Island, South Korea, 2013, pp. 807–812.

[9] R.K. Banyal, P. Jain, and V.K. Jain, "Multi-factor Authentication Framework for Cloud Computing," in *Proc. of Fifth International Conference on Computational Intelligence, Modeling and Simulation (CIMSim)*, Seoul, South Korea, 2013, pp. 105-110

[10] F. Fatemi Moghaddam, O. Karimi, and M.T. Alrashdan, "A Comparative Study of Applying Real-Time Encryption in Cloud Computing Environments," in *Proc. of 2nd International Conference on Cloud Networking (CloudNet)*, San Francisco, CA, 2013, pp. 185–189.

[11] M. Randles, D. Lamb and A. Taleb-Bendiab, "A Comparative Study into Distributed Load Balancing Algorithms for Cloud Computing," in *proc. of 24th International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, Perth, Australia, 2010, pp. 551-556.

[12] K. A. Nuaimi, N. Mohamed, M.A. Nuaimi, and J. Al-Jaroodi, "A Survey of Load Balancing in Cloud Computing: Challenges and Algorithms," in *proc. of Second Symposium on Network Cloud Computing and Applications (NCCA)*, London, UK, 2012, pp. 137-142.

[13] R. Buyya, R. Ranjan and R.N. Calheiros, "InterCloud: Utility-Oriented Federation of Cloud Computing Environments for Scaling of Application Services, Algorithms and Architectures for Parallel Processing," *Lecture Notes in Computer Science*, vol. 6081, pp. 13-31, 2010.

[14] R. Moreno-Vozmediano, R.S. Montero, and I.M. Llorente, "Key Challenges in Cloud Computing: Enabling the Future Internet of Services," *IEEE Internet Computing*, vol. 17, no. 4, pp. 18-25, July 2013.

[15] C. Fehling, F. Leymann, R. Retter, W. Schupeck, and P. Arbitter, "Chapter 2: Cloud Computing Fundamentals," Cloud Computing Patterns: Fundamentals to Design, Build, and Manage Cloud Applications, *Springer Press*, pp. 21-75, 2014.

[16] M.A. Chauhan, and M.A. Babar, "Migrating Service-Oriented System to Cloud Computing: An Experience Report," in *proc. of IEEE International Conference on Cloud Computing (CLOUD)*, Washington, DC, 2011, pp. 404-411.

[17] A. Khajeh-Hosseini, D. Greenwood, and I. Sommerville, "Cloud Migration: A Case Study of Migrating an Enterprise IT System to IaaS," in *proc. of IEEE 3rd International Conference on Cloud Computing (CLOUD)*, Miami, FL, 2010, pp.450-457.

[18] L. Schubert, and Expert Group Report, "The Future of Cloud Computing: Opportunities for European Cloud Computing Beyond, European Commission," *Information Society and Media*, 2010.

[19] D. Petcu, "Portability and Interoperability between Clouds: Challenges and Case Study, Towards a Service-Based Internet," *Lecture Notes in Computer Science*, vol. 6994, pp. 62-74, 2011.