A New E-learning Model Based on Elastic Cloud Computing for Distance Education

Wei Zhang [1], Yanchun Zhu [2*]

[1] School of Information, Central University of Finance and Economics, Beijing 100081, CHINA
[2] Business School, Beijing Normal University, Beijing 100875, CHINA

ABSTRACT

The development of information technology has a positive influence over distance education. For instance, it gives birth to E-learning, a novel and convenient mode of online learning. However, the existing E-learning platforms are often in need of heavy infrastructural investment. To solve the problem, this paper introduces the architecture of elastic cloud computing to E-learning, and proposes a new load balancing algorithm based on elastic cloud computing. Then, it establishes a five-layer elastic cloud-based E-learning model for distance education. The subsequent evaluation demonstrates that the model can effectively respond to the dynamic load imposed on by numerous users, and maximize the revenue of the E-learning platform.

Keywords: e-learning, elastic cloud computing, distance education, load balancing

INTRODUCTION

E-learning is an effective mode of multimedia learning based on electronic education technology (Moreno & Mayer, 1999; Greer et al., 2013). Featuring flexibility and diversity, the mode has been widely adopted by educational institutions to teach learners in faraway places. The main advantage of e-learning lies in timeliness. E-learning saves learners from traveling to and from classrooms in traditional training, imparts the latest policies, ideas and concepts in real time, and keeps the coursework refreshed and updated as required. Nevertheless, the development of E-learning platform is very expensive, due to the procurement of servers, storage devices, and other network hardware (Dong, 2009). Besides the excessive cost, the lack of platform design and management personnel is another serious problem.

The cloud computing technology offers a good solution to these problems. As an information technology, cloud computing helps share pools of configurable resources, such as servers, apps and services (Wang et al., 2010). The basic architecture of cloud computing is shown in **Figure 1**.

Cloud computing uses a service-oriented architecture (Kulshrestha & Verma, 2013) and treats "everything as a service" (Duan et al., 2015). There are three standard models for cloud computing, namely Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) (Mell & Grance, 2009).

The PaaS means the user can deploy self-developed apps on the cloud infrastructure, but at the same time are obliged to manage the deployed apps and configure the setting environment. In the SaaS, the user is allowed to utilize the apps on the cloud infrastructure, but does not have to manage the cloud infrastructure (e.g. network, servers, operating systems, storages, apps). The IaaS aims at providing the user with fundamental computing resources (e.g. processor, storage, network) for running operating systems and apps. The three models are arranged as layers in a stack frame (**Figure 2**).

The greatest strength of cloud computing is cost efficiency, so using this technology can help improve the cost-effectiveness of E-learning. In this way (**Figure 3**), educational institutions can utilize the E-learning platforms created by providers of cloud computing services.

In the conventional E-learning mode, the platform is designed, implemented and maintained by educational institutions. By contrast, cloud-based E-learning greatly eases the burden on educational institutions. For example, the new E-learning mode, featured by high scalability, allows cloud service providers to offer infrastructure layers
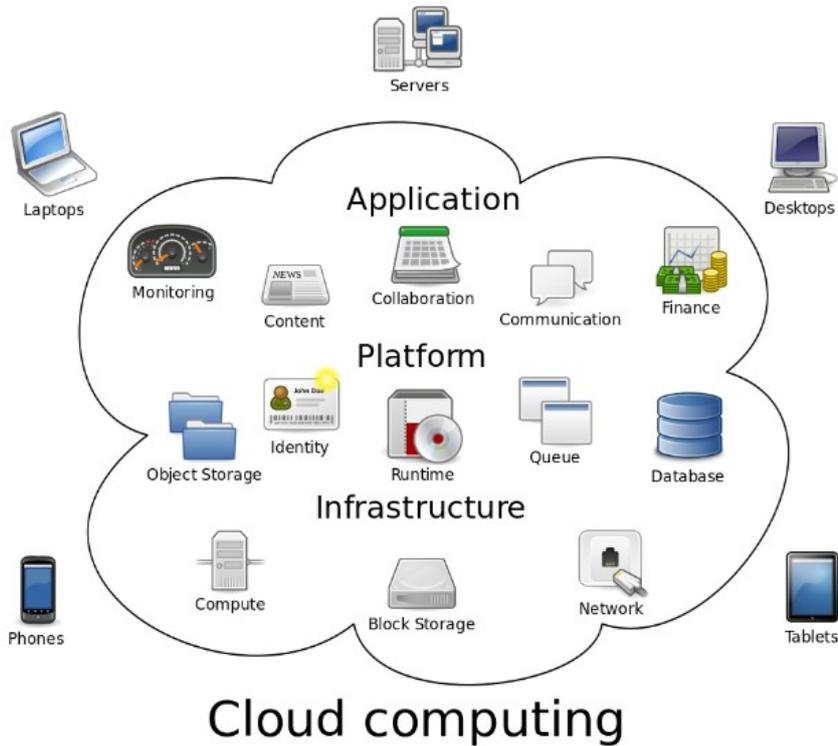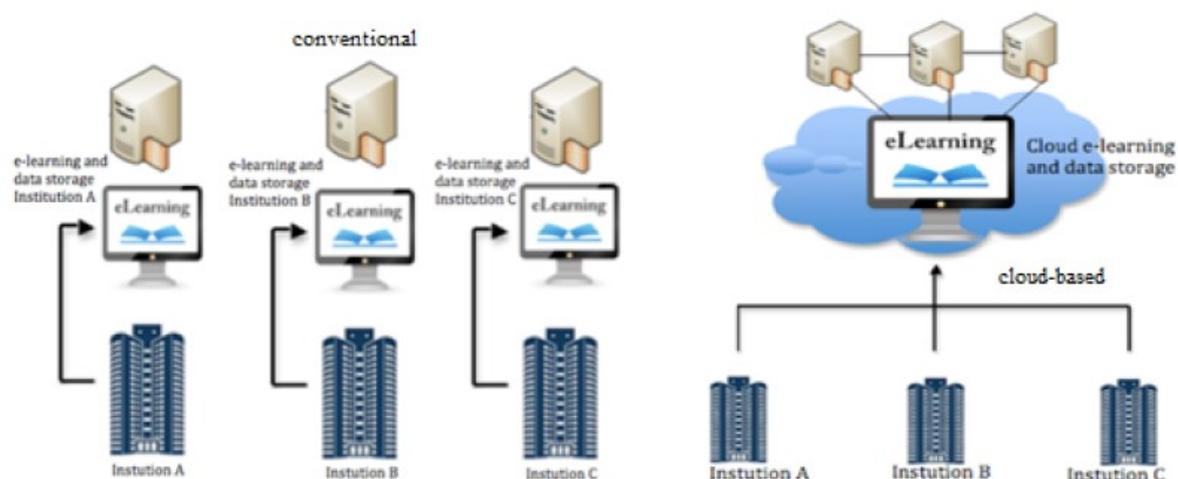
Figure 1. The basic architecture of cloud computing
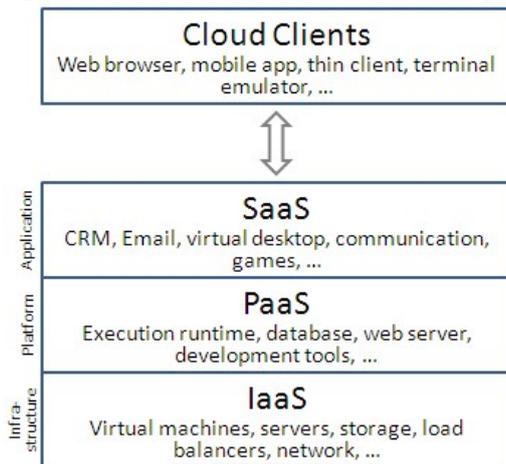
Figure 2. Three cloud computing service models

through virtualization, which makes up for the lack of hardware resources, and also achieves automated management of these resources.

Figure 3. Conventional e-learning vs. cloud-based e-learning

CLOUD-BASED E-LEARNING ARCHITECTURE

E-Learning is a web-based learning method with the aim to provide conventional learning services through the Internet (Alavi et al., 2013). Common e-learning platforms are built on the client-server infrastructure (Phankokkruad & Woraratpanya, 2008). Due to the lack of interoperability and accessibility, the infrastructure cannot give full play to the functions of e-learning platforms. Some scholars (Xu et al., 2003; Bhatt et al., 2009; Shanmugapriya & Tamilarasi, 2012) have solved the interoperability problem by implementing web services to e-learning. However, the accessibility problem still exists when there are too many online users.

Over the years, a lot of research has been done on E-learning based on the cloud architecture. Praveena & Betsy (2009) described the application of cloud-based E-learning in universities. Shri suggested implementing E-learning system based on private cloud (Shri & Subha, 2013). Masud & Huang (2012) designed a new E-learning architecture for educational institutions based on cloud computing. Arpaci (2017) assessed the management of educational knowledge, and incorporated the knowledge into an efficient infrastructure based on cloud computing. However, there are still lots of difficulties in the implementation of cloud-based E-learning. The most prominent problem lies in the high cost of the cloud architecture for E-learning.

Existing Architecture

This section will discuss the existing cloud-based architectures previously proposed by other researchers and also the contribution from the proposed architecture.

The existing cloud-based E-learning architecture typically consists of three layers (Thyagharajan & Nayak, 2007): the infrastructure layer, the platform layer, and the application layer. The infrastructure layer, also known as the hardware layer, provides the computing and storage capacity for the entire system. Consisting of a content module and a course module, the platform layer is a middleware that supplies learning resources as web services, and enables users to select and access suitable learning courses and content from the database. The application layer is responsible for providing interactive interfaces for distance education learners.

Chuang et al. (2011) improved the three-layer architecture into a five-layer one. The first layer is still the infrastructure layer; the second is the software layer consisting of the operating system and middleware; the third is the resource management layer, which allocates different hardware resources to software and implements the cloud computing scheduling strategy; the fourth is the service layer, involving such three services as SaaS, PaaS and IaaS; and the fifth is the application layer of integrated learning sources, which include teaching resources, interactive courses and so on.

As stated above, the infrastructure layer is the most fundamental layer of the existing cloud-based E-learning architecture. With the expansion of infrastructure scale, the use of the infrastructure now faces problems like high energy consumption, waste of hardware resources, and low efficiency. This calls for urgent enhancement of resource utilization. Although the proposed model is also multi-layered like those existing architectures, this paper introduces the elastic cloud computing into the infrastructure layer, which saves a great portion of hardware
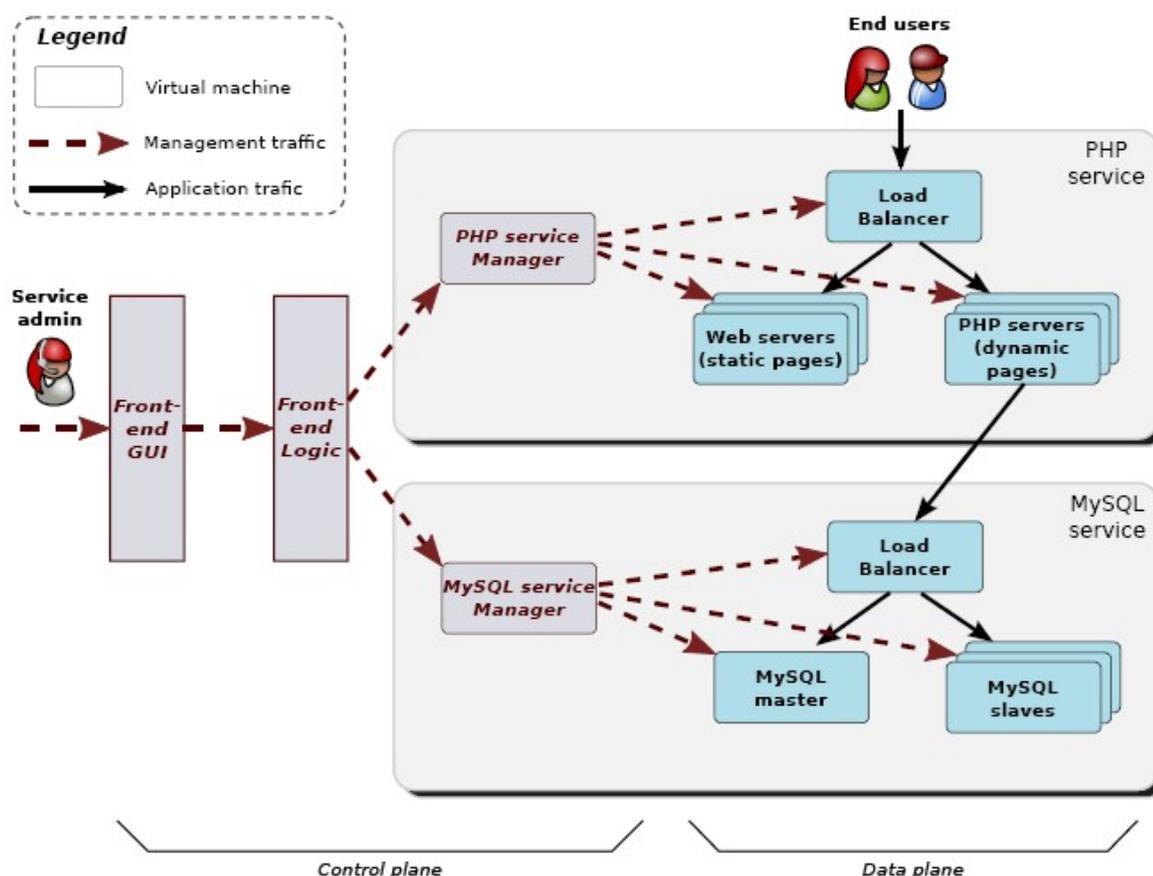
Figure 4. Elastic cloud computing architecture

investment, and maximizes the performance of the entire platform. The main advantage of the proposed model is cost efficiency, which is an interesting point and can be adopted to develop e-learning.

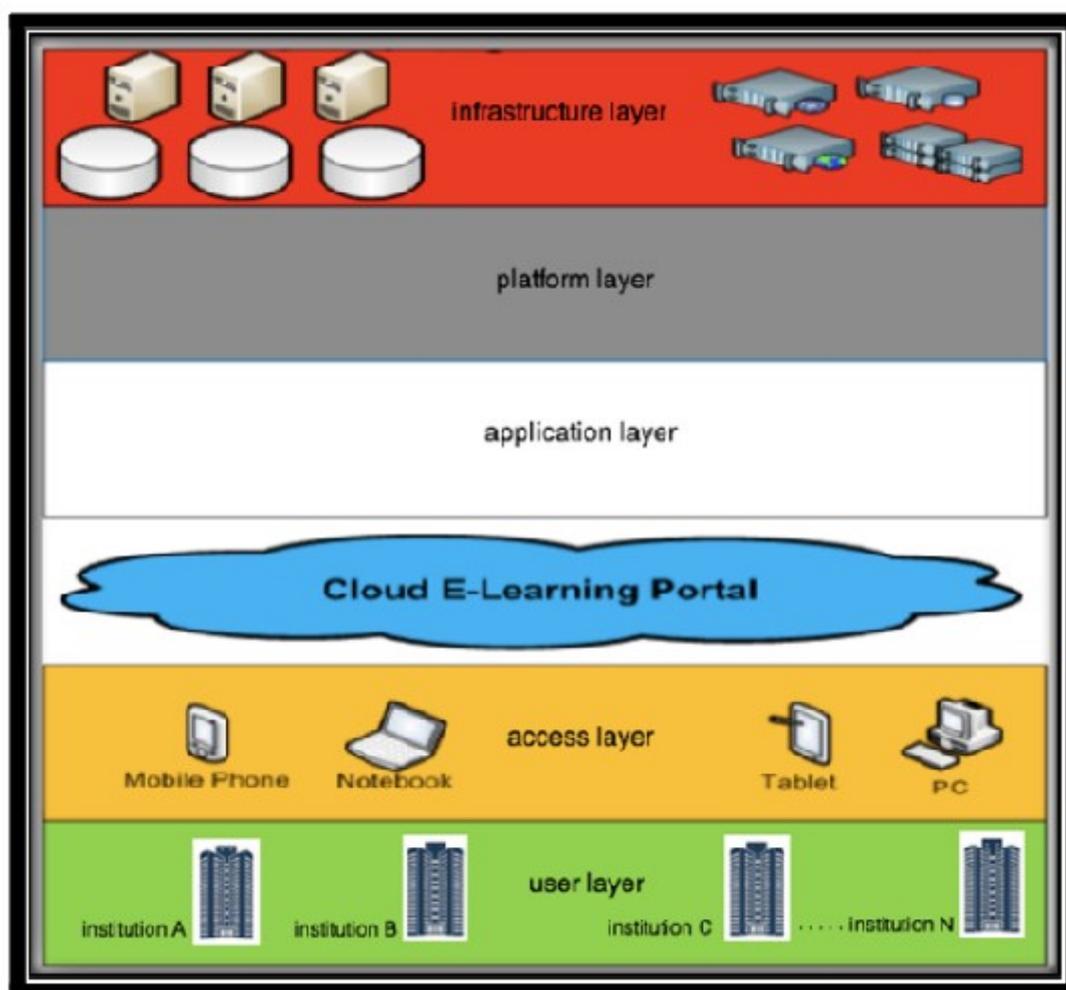## E-learning Architecture Based on Elastic Cloud Computing

In cloud computing, elasticity is defined as "the degree to which a system is able to adapt to workload changes through provisioning and de-provisioning of resources in an autonomic manner, so that at each time point the available resources match the current demand as closely as possible" (Al-Dhuraibi et al., 2017). The dynamic adaptation of capacity, e.g., by altering the use of computing resources, to meet a varying workload is called "elastic computing" (Peng et al., 2008).

Elasticity aims at matching the amount of resources allocated to a service with the amount of resources it actually requires, avoiding over- or under-provisioning, that is, allocating more or fewer resources than required. Otherwise, the service provider may not provide its users with good services.

An example of elastic cloud computing is given to explain its operating mechanism. **Figure 4** displays an app composed of two services - PHP and MySQL - on an elastic cloud computing platform.

In the example, the end user generates all the traffics by accessing the app. The requests of the end user are handled by load balancers based on an elastic cloud computing architecture. The PHP service processes HTTP traffic, while the MySQL service handles database queries. Each service is controlled by a manager. The ultimate goal of elastic cloud computing is to automate the traffic control. To this end, the system must react quickly through dynamic adjustment of its capacity to address the overload. To lower the cloud computing cost, the system should also deprovision resources at the decrease of the traffics.

As shown in **Figure 4**, the elastic cloud computing system is made up of load balancers, PHP servers and MySQL servers. The system capacity must be elevated whenever the load increases. Besides, it is also necessary to identify the performance bottleneck of the app and decide the exact resources to be utilized.

Figure 5. Structure of the proposed model

E-LEARNING MODEL BASED ON ELASTIC CLOUD COMPUTING

In this paper, a new E-learning model is proposed based on elastic cloud computing. The proposed model has five layers: the infrastructure layer, the platform layer, the application layer, the access layer and the user layer. The structure of the proposed model is detailed in **Figure 5**.

The infrastructure layer offers an elastic cloud computing platform, which contains all kinds of hardware resources. The sharing of these resources is made possible by elastic cloud computing. The platform layer provides the space for the running of E-learning apps. The application layer is the home to specific learning resources. The access layer is unique part of the model, the purpose of which is to access the services in the cloud platform. To do so, the multi-channel access technology is employed to enable various services through various end-devices. The user layer consists of various educational institutions.

Based on the elastic cloud computing architecture, the infrastructure layer acts as a service provider with unlimited amount of resources, which is implemented through the procedure illustrated in **Figure 6**.

In **Figure 6**, the proxy represents the front-end server that receives user requests; the web server monitor (WSM) monitors the response time of all the deployed servers; the scheduler adapts the entire architecture to the needs of all users; the web provider responds to the web requests from distance users through various servers; and the cloud service provider allocates unlimited amount of external resources, aiming to support the running of any number of virtual machines (VMs).
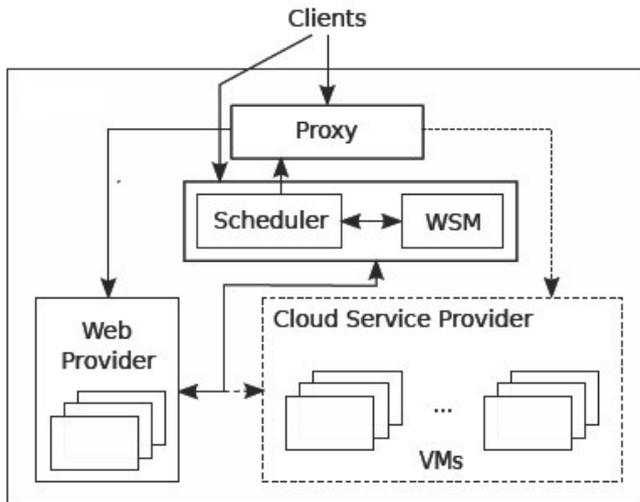
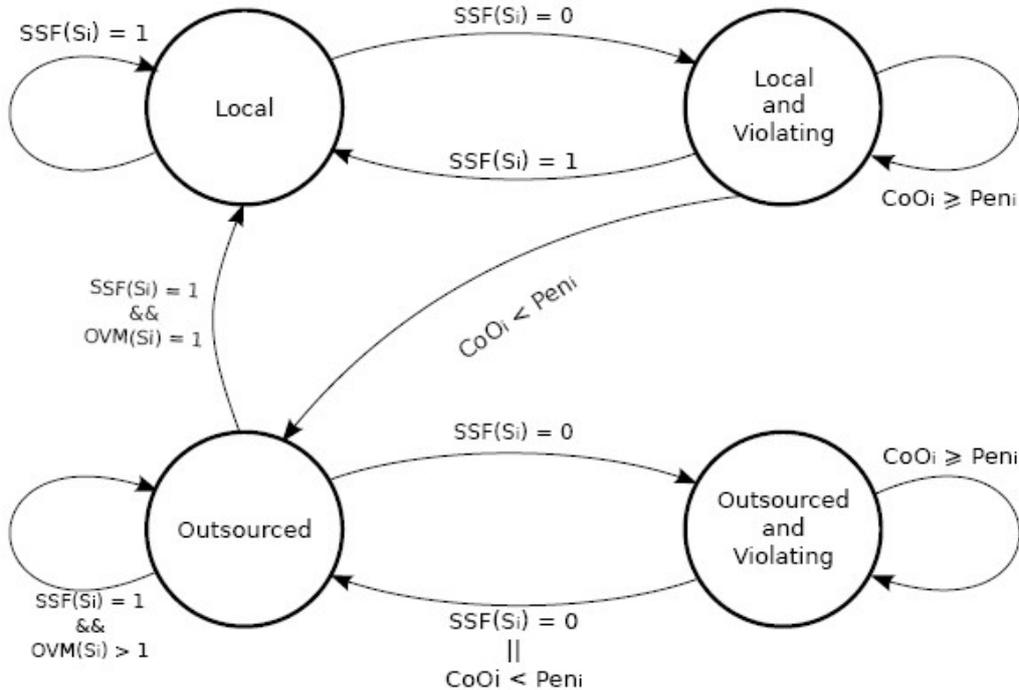**Figure 6.** The implementation process of infrastructure layer

Figure 7. State variation of web apps

Moreover, the scheduler is responsible for deploying new web apps. The deployment starts with the creation of a local server, and ends with the running of web apps on the server. **Figure 7** explains the four models deployed with web apps. The state of a web app will change if the penalty surpasses or falls below the cost of outsourcing.

The performance of elastic cloud computing is critical to the effectiveness of the proposed model. One of the representative performance metrics is the response time. The metric is defined as the time it takes for the server to respond to a user request. The response time will remain constant if the server is not overloaded; otherwise, it will be increased fast and substantially.

The exponentially weighted moving average (EWMA) (Lucas et al., 1990) is adopted to capture the variation pattern of the response time of a server in a short time period. In this method, the weight factor increases exponentially. The degree of increase can be measured by the factor α ($0 \leq \alpha \leq 1$). The current observed value is determined based on the last observed value. The observed value at the time period t can be calculated as follows:

$$S_t = \alpha \cdot Y_{t-1} + (1 - \alpha) \cdot S_{t-1} \tag{1}$$

The response time of the server can be expressed as:

$$R_t = (e^{-(1/T)}) \cdot R + (1 - e^{-(1/T)}) \cdot R_{t-1} \tag{2}$$

where R_t and R_{t-1} are the degrees of the current and last increases, respectively; T is the total monitoring time over the server; and R is the last response time.

LOAD BALANCING METHOD BASED ON ELASTIC CLOUD COMPUTING

The growing number of distance education users has exerted a tremendous pressure on elastic cloud computing. To cope with the pressure, a load balancing method is proposed based on elastic cloud computing.

The proposed load balancing framework encompasses a load balancer, a cloud resource management platform, a VM mirror library, a VM cluster and a resource pool. The VM cluster has several VMs, each of which carries several apps and agents. The apps are responsible for receiving and executing user tasks, while the agents are responsible for collecting VM load data and task execution data and sending them to the load balancer.

Load Balancing Model Based on Elastic Cloud

The load data are measured by the usage of the CPU, memory, hard drive and bandwidth. It is assumed that the load data index is p, that the VM cluster has a total of m VMs, and that the data are collected at a total of n time points.

Definition 1 l_{ij}^k represents the load data of the load index i collected at time point j in the VM k. The VM's data record set L^k is introduced to express the load data of the VM k. Hence, the VM cluster load data are recorded as a set of load data L for all VMs within the cluster.

$$L^k = \{l_{ij}^k | 1 \leq i \leq p, 1 \leq j \leq n\} \tag{3}$$

$$L = \{L^k | 1 \leq k \leq m\} \tag{4}$$

The proposed model assumes that the task type remains the same for a given cloud-based E-learning system. The weight factor set W, which represents the weight of the VM load index, is used to evaluate the task requirement on each resource.

Definition 2 w_i represents the weight of load index i.

$$W = \{w_i | 1 \leq i \leq p\} \tag{5}$$

Definition 3 f^k represents the load of VM k in the evaluation of the VM load condition.

The load serves as the primary reference for the task scheduling algorithm on the VM's real-time load status. Thus, the load data collected at the nearest time point n are taken for analysis. Then, f^k is calculated by the following formula:

$$f^k = \sum_{i=1}^{p} w_i \times l_{in}^k \tag{6}$$

Owing to the heterogeneity of the VMs in the cluster, the resource utilization of the VM is evaluated against both the load and ability of the VM.

Definition 4 c_i^k represents the capability of VM k on load index i, and c^k represents the comprehensive ability of VM k.

$$c^k = \sum_{i=1}^{p} w_i \times c_i^k \tag{7}$$

Definition 5 s_i represents the resource utilization of the VM cluster for load index i.

The resource utilization demonstrates the efficiency of the VM cluster over a period of time. In theory, the load data should be continuously changing. In fact, the actual load data is a discrete random variable because the data is collected at various time points for different VMs. The time interval for load data acquisition is set to Δt. Then, the total amount of resources R_i^k consumed by VM m on load index i in periods 1~n can be defined as:

$$R_i^k = \sum_{j=1}^{n} (l_{ij}^k \times c_i^k \times \Delta t) \tag{8}$$

For the cluster in the periods *1~n*, the total consumption of resources R_i for load index i can be defined as:

$$R_i = \sum_{k=1}^{m} R^k \tag{9}$$

Then, S_i can be obtained by:

$$S_i = \frac{R_i}{\sum_{k=1}^{m} c_i^k \times n \times \Delta t} \tag{10}$$

According to Eq. (7) ~ (10), S_i can be expressed as follows:

$$S_i = \frac{\sum_{k=1}^{m} \sum_{j=1}^{n} (l_{ij}^k \times c_i^k)}{\sum_{k=1}^{m} c_i^k \times n} \tag{11}$$

Using the weight factor W for S_i, the virtual machine cluster resource utilization rate S can be obtained.

$$S = \sum_{i=1}^{p} w_i \times S_i \tag{12}$$

Load Balancing Algorithm Based on Elastic Cloud

Definition 6 q_i^k represents the VM k in state i, q_0^k represents the VM in the initial state, and Q_i represents the set of VMs in state i.

$$Q_i = \{q_i^k | 1 \leq k \leq m\} \tag{13}$$

The basic idea of the task scheduling algorithm is to sort the VMs by their real-time load conditions in the VM cluster, and assign the task to the VM with the smallest load. The algorithm is implemented in the following steps:

Step 1: The task scheduling module receives a task;

Step 2: Extract the VM load of the VM cluster $Q_{cluster}$ from the data analysis module;

Step 3: Calculate f^{min} based on the VM load: $f^{min} = \{f^k | q^k \in Q_{cluster}\}$, and find the VM with the smallest load q^{min};

Step 4: Assign a task to the VM with q^{min}.

The elastic algorithm consists of two parts: the VM recovery and the VM cluster expansion. If the VM cluster resource utilization is below the minimum threshold δ_{min}, the VM cluster has relatively abundant computing resources. In this case, some of the VMs can be recycled to improve resource utilization. If the VM cluster resource utilization is above the maximum threshold δ_{max}, the VM cluster is under excessively heavy loads. In this case, more computing nodes should be introduced to share tasks and shorten response time. The algorithm is implemented in the following steps:

Step 1: Extract resource utilization ratio S of VM cluster $Q_{cluster}$ from the data analysis module, and compare S with δ_{min} and δ_{max}: if $S < \delta_{min}$, go to Step 2; if $S > \delta_{max}$, go to Step 3; if $\delta_{min} \leq S \leq \delta_{max}$, go to Step 7;

Step 2: Compute $f^{collect} = \min\{f^k | q^k \in Q_{cluster}\}$, select $q^{collect}$ from $Q_{cluster}$, turn it into the recycling state $q_5^{collect}$, and go to Step 7;

Step 3: Compute $f^{collect} = \min\{f^k | q^k \in Q_5\}$; if $Q_5 = \emptyset$, go to Step 4; if $Q_5 \neq \emptyset$, select $q^{extend1}$ from Q_5, turn it into the steady state $q_3^{extend1}$, and go to Step 7;

Step 4: Compute $c^{extend2} = \max\{c^k | q^k \in Q_6\}$; if $Q_6 = \emptyset$, go to Step 5; if $Q_6 \neq \emptyset$, select $q^{extend2}$ from Q_6, turn it into the steady state $q_3^{extend2}$, and go to Step 7;

Step 5: Compute $c^{extend3} = \max\{c^k | q^k \in Q_7\}$' if $Q_7 = \emptyset$, go to Step 6; if $Q_7 \neq \emptyset$, select $q^{extend3}$ from Q_7, turn it into the open state $q_2^{extend3}$, and go to Step 7;

Step 6: Create a VM $q^{extend4}$, and execute the open command;

Step 7: Delay for T_{delay} time, and return to Step 1.

EXPERIMENTS AND ANALYSIS

To verify the proposed method, several experiments were performed to simulate the E-learning platform based on elastic cloud computing.

Figure 8 shows the creation of a necessary VM, and the scheduling actions to balance the load. From top to bottom, the figure displays the input load and response time of the servers, the number of back-end servers, and the load of the cloud service machine. It can be seen that the proposed algorithm has balanced the load between servers.
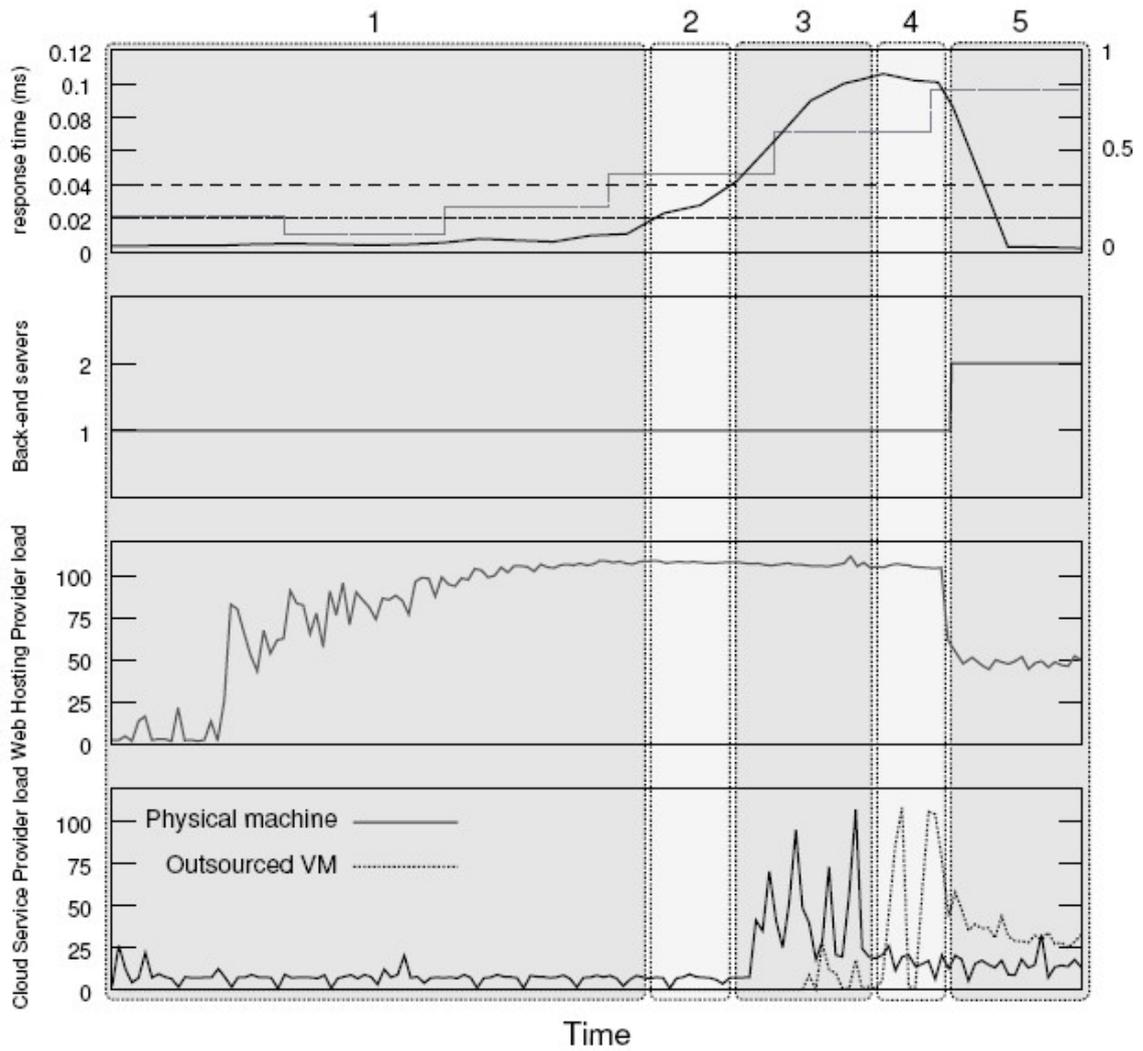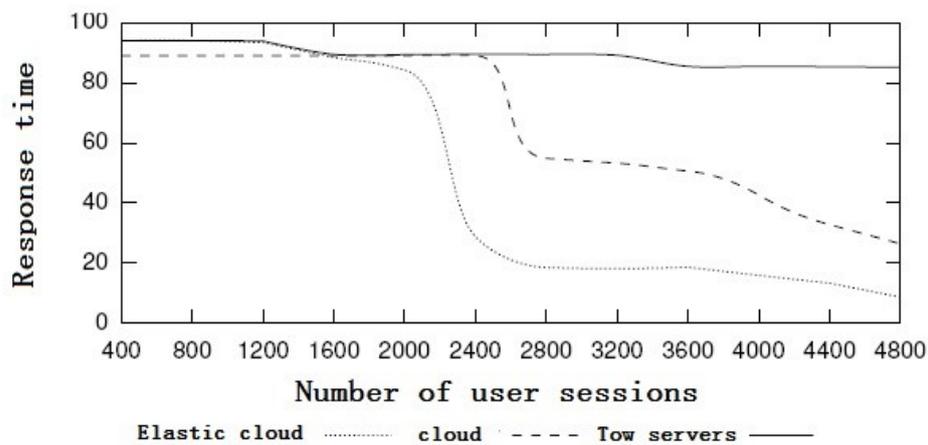
Figure 8. Load balancing demand

Figure 9. Response time vs. the number of user sessions

Figure 9 depicts the response time to end-user requests on three different E-learning platforms. It is clear that the proposed model boasts the best response time in all the cases. The greater the number of users, the more obvious the advantage of the proposed platform.

CONCLUSIONS

There have been several cloud-based e-learning models proposed by previous researchers. Most of these models are multi-layered. This paper proposes a five-layer e-learning model based on elastic cloud computing. Compared with those existing researches, the introduction of elastic cloud computing saves a great portion of hardware investment, and maximizes the performance of the entire platform. Currently, the e-learning platforms are usually operated independently by each educational institution. The independent operation is very costly because of the provision, development and maintenance of the infrastructure. The architecture of elastic cloud computing is mainly adopted in the first layer, namely, the infrastructure layer, which is the greatest advantage of the proposed model. In order to maximize the effectiveness of the proposed model, the load balancing algorithm based on elastic cloud is introduced to save hardware resources and shorten response time. The experimental results show that the proposed model can greatly improve the resource utilization with the aid of the load balancing algorithm. The prototype of the proposed model has been developed and piloted in two institutions of higher education. Further research will continue to focus on the load balancing algorithm for improving the performance of the whole e-learning platform so that it can be applied to more higher education institutions.

ACKNOWLEDGEMENT

REFERENCES

Alavi, E., & Mohan, M. C. (2013). An E-Learning System Architecture Based on New Business Paradigm Using Cloud Computing. *International Journal of Engineering Sciences & Research Technology*, 2, 17-28.

Al-Dhuraibi, Y., Paraiso, F., & Djarallah, N. (2017). Elasticity in Cloud Computing: State of the Art and Research Challenges. *IEEE Transactions on Services Computing*, 99, 1-9.

Arpaci, I. (2017). Antecedents and consequences of cloud computing adoption in education to achieve knowledge management. *Computers in Human Behavior*, 70, 127-132.

Bhatt, M., Rahayu, W., & Soni, S. P. (2009). Ontology driven semantic profiling and retrieval in medical information systems. *Web Semantics Science Services & Agents on the World Wide Web*, 7(4), 317-331.

Chuang, S. M., Chang, K. E., & Sung, Y. T. (2011). The cost effective structure for designing hybrid cloud based enterprise E-learning platform. *IEEE International Conference on Cloud Computing and Intelligence Systems*, IEEE, 523-525.

Dong, B. (2009). BlueSky Cloud Framework: An E-learning Framework Embracing Cloud Computing. *Cloud Computing Proceedings*, 5931, 577-582.

Duan, Y., Fu, G., & Zhou, N. (2015). Everything as a Service (XaaS) on the Cloud: Origins, Current and Future Trends. *International Conference on Cloud Computing*, IEEE, 621-628.

Greer, D. L., Crutchfield, S. A., & Woods, K. L. (2013). Cognitive Theory of Multimedia Learning, Instructional Design Principles, and Students with Learning Disabilities in Computer-Based and Online Learning Environments. *Journal of Education*, 193, 41-51.

Kulshrestha, V., & Verma, S. (2013). Analyze the Future: Cloud Computing, a New Phase in Information Technology Infrastructure Management. *International Journal of Computers & Technology*, 11(6), 43-49.

Lucas, J. M., Saccucci, M. S., & Woodall, W. H. (1990). Exponentially weighted moving average control schemes: properties and enhancements. *Technometrics*, 32(1), 1-29.

Masud, M. A. H, & Huang, X. (2012). An E-learning System Architecture based on Cloud Computing. *System*, 6, 65-74.

Mell, P., & Grance, T. (2009). The NIST definition of cloud computing. *Communications of the Acm*, 53(6), 50-50.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology*, 91(2), 358-368.

Peng, S., Zhang, M., & Niu, X. (2008). Magnetically responsive elastic microspheres. *Applied Physics Letters*, 92(1), 5492-5498.

Phankokkruad, M., & Woraratpanya, K. (2008). Web Service Architecture for Computer-Adaptive Testing on e-Learning. *Proceedings of World Academy of Science Engineering & Technology*, 46, 347-358.

Praveena, K., & Betsy, T. (2009). Application of Cloud Computing in Academia. *Social Science Electronic Publishing*, *25*(2), 101-104.

Shanmugapriya, M., & Tamilarasi, A. (2012). Developing a Mobile Adaptive Test (MAT) in an M-Learning Environment for Android Based 3G Mobile Devices. *International Journal on Computer Science & Engineering*, *4*(2), 153-161.

Shri, M. L., & Subha, S. (2013). An implementation of E-learning system in private cloud. *International Journal of Engineering & Technology*, *5*(3), 3036-3040.

Thyagharajan, K. K., & Nayak, R. (2007). Adaptive content creation for personalized E-learning using web services. *Journal of Applied Sciences Research*, *3*(9), 828-836.

Wang, L., Laszewski, G. V., & Younge, A. (2010). Cloud Computing: a Perspective Study. *New Generation Computing*, *28*(2), 137-146.

Xu, Z., Yin, Z., & Saddik, A. E. (2003). A Web services oriented framework for dynamic e-learning systems. *Electrical and Computer Engineering*, *2*, 943-946.

http://www.ejmste.com