ENHANCED COST EFFECTIVE RESOURCE ALLOCATION SCHEME FOR STREAMING SERVICE IN HETEROGENEOUS CLOUD

C. Venish Raja, Research Scholar, Department of Computer Science, Bishop Heber College (Autonomous), Affiliated to Bharathidasan University, Trichy, Tamilnadu, India.
Dr. L. Jayasimman, Assistant Professor, Department of Computer Science, Bishop Heber College (Autonomous), Affiliated to Bharathidasan University, Trichy, Tamilnadu, India.

Abstract
An effective resource allocation strategy is required in cloud paradigms to achieve customer satisfaction and to optimize the income for cloud service providers. Cloud computing is rapidly being used in businesses and business markets. With the rise of dynamic and personalized content in media distribution systems, not all application-specific requirements can be met by a cloud-based solution or a content delivery network (CDN)-based solution. Therefore Cloud integration with CDNs and private data center can be beneficial. This work proposed a new architectural model for minimizing costs and providing services. An algorithm is designed to capture the event that the user raises when using the service. Resources are allocated on the basis of activities, user request number and task or request size. The only characteristic of this model is to provide the three media services such as video sharing, video streaming and live streaming with cost efficient flexible schemes. The aim of the research is to allocate resources based on customer requests and events. The experimental findings show that the new system is stronger with respect to the expense and time of response. Furthermore, the distribution of resources based on incidents often reduces the availability of services.
Keywords- Streaming video, Event Capturing, CDN, Video Streaming

Introduction
The theme of the cloud is "The transition we're leading." In simple terms, cloud computing refers to the sharing of resources. Resources can be either software or hardware that can be provided as a service. Deployment models such as private, public, hybrid and community cloud are used to provide a scalable and reliable service cloud. These models are based on four key characteristics, such as broad network access, rapid elasticity, pooling of resources and measured service. As mentioned at the outset, it is believed that users are moving towards a participatory era from the ice age, the iron or metal age, the industrial age, the information age or the digital age. It's called so because internet users are involved, not only looking for or collecting information from the internet. Some instances are that users can create blogs without understanding the technological aspects and take pictures publishing around the world etc., etc.

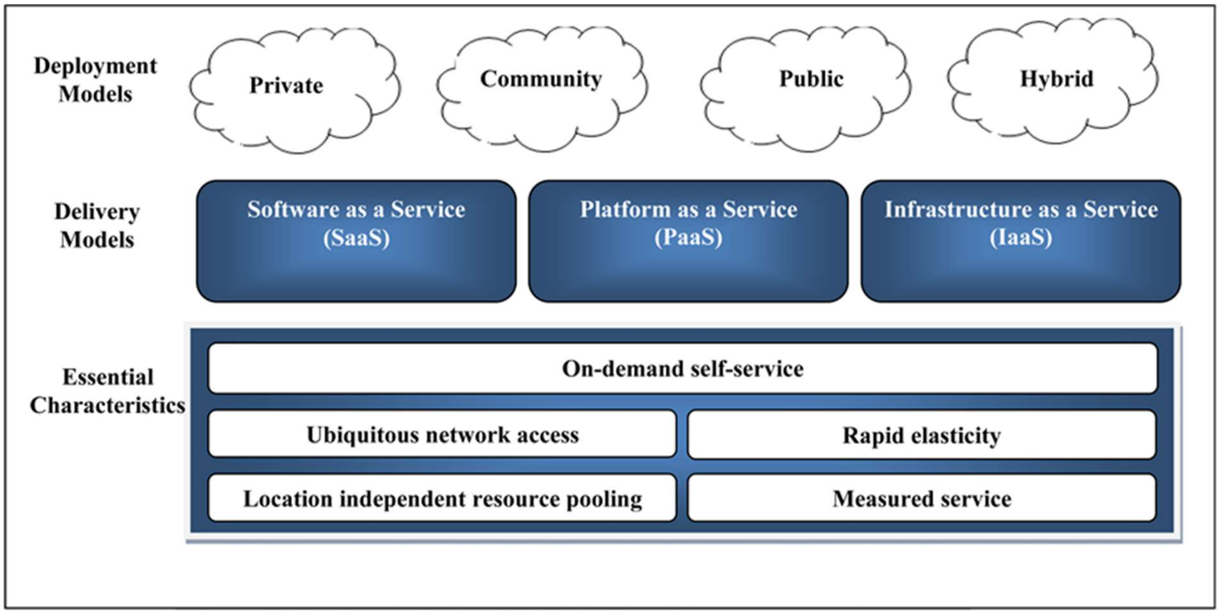

Figure-1 Architecture of Cloud Computing

Figure-1 shows the simplified architecture of cloud with its layers and components. Cloud scalability is one of the key features that attracts many customers to cloud services. In the sense of cloud computing scalability, the ability to manage increasing or rising resources in a capable way can be described as meeting the business demands. The cloud has three scalability forms, [2]. They are of horizontal,vertical and diagonal form. Horizontal scaling is the process of adding more machines to the resource pool, whereas vertical scaling is the process of adding more computing power to a machine, such as RAM, Storage, bandwidth etc . Scaling diagonal means the combination of the vertical servers and the horizontal servers, which updates, incorporates components and, in its current configuration, replicates the server. This provides the most powerful scaling method for both price and efficiency. Growth is the ultimate goal for most companies. Businesses need storage solutions to meet fluctuating needs to do so. Cloud computing provides the ability to be efficiently flexible when data and infrastructure requirements are increased and reduced. Cloud computing is scalable and cost-effective for companies serving millions of customers every day. Data from different devices can be accessed in today's world and a company needs to determine the type and quantity of resources required to meet its customers ' requirements with minimal costs and time. Video use is not a novel phenomenon in a number of applications for communication and learning. For decades, video has become a central medium of communication for a wide range of customers. As a medium of corporate communication, video is now creating much more excitement due to the availability of tools that allow organizations or individuals to cost-effectively create digital video assets and deliver them in real time or on demand. Video content from live and stage events or from smartphone or video cameras or video sharing sites can be developed. There are two forms of data accessed by customers. It is dynamic and static data. The weather forecast, stock market and traffic analysis are few examples of dynamic data while media streaming, e-commerce and social networking are static data services. Current research focuses on media streaming data service evaluating the possibility to provide a better response to the needs of the user. Cloud-based streaming services[3] rank in three types. It's video sharing, streaming content and watching online. Video sharing is a sharing or upload process within a free video sharing service. Example: youtube, vimeo, regular motion etc. Video sharing sites are websites that host authenticated user pre-commanded videos. Example: Netflix, Amazon prime, and Youtube Red, etc., Live streaming refers to viewing content in real time, much like watching live news or TV sports. Example: live on Twitter, live on Facebook, hotstar. Such subscription services use vast quantities of capital to satisfy their consumer needs, resulting in resource wastage and rising expense and energy costs.

The rest of the paper is created accordingly. Section 2 describes a summary of related field work. Section-3 describes the components proposed for the design and the framework. The description of architecture and its entities is defined in Section 4. Section-5 explains implementation and experimental set-up and results. Section-6 concludes future work and reporting.

Related Works

Some of the key aspects of resource management policy is to reduce costs while accessing media content. Most cloud-based streaming service providers have tools to satisfy cloud users ' expectations for quality of service (QoS). This leads to resource wastage, even though the cloud user uses different options such as browsing, trending, updates, etc., rather than using media player to view the videos on websites. A real-time QoE monitoring system for streaming services with Adaptive Media Playout was designed and implemented to adjust the play rate of videos according to the buffer fullness[4]. Some codecs like H.264 / AVC and H.265 / HEVC aren't supported in this approach. In addition it was not tested in the cloud environment. The parameters such as response time and cost were not tailored to fulfill user requirements. [5]'s work proposed the scheduling algorithm which can be implemented in the cloud setting to reduce the transcoder's delay time. This work lacks the amount of resources to be allocated for each request, because the algorithm for scheduling is focused on file size. However, with any proposal the amount of services added or decreased has not been taken into account and thus results in high costs. For efficient streaming of live videos using cloud content delivery network and priority-based round robin scheduling algorithm, a two-tier architecture has been proposed in [6]. This approach lacks the cost to allocate resources and time to process work to allocate those resources. This approach is also focused only on the live streaming videos. [7] 's work implemented a pre-transcoding method partly pre-transcoding to minimize the cost of storage and assess the efficacy of that strategy. This method does not focus on the live streaming videos as well as the video quality. The key drawback of transcoding in this method video was transcoded twice is quality loss. In [8] feedback control loop was designed to ensure maximum player and video quality use. The key downside to regulation of the feedback loop or the closed loop system is lack of reliability, costlier, dynamic and oscillating response. While live streaming videos were not analyzed using this method. New live video streaming has been proposed to reduce the cost and improve the efficiency of the streaming video[9]. The key downside of this strategy is that it lacks transcoder and it also fails to measure the cost when the stored video in the cloud lasts for several weeks. Cloud-based Multiple view crowd source streaming is proposed in[10] for the efficient sharing of stage event videos using the cloud environment. This approach is intended only for small videos that can be uploaded from mobile devices, and can only be viewed from mobile devices. Scalable and maintenable architecture has been designed for streaming data analytics in the work of [11]. This approach focuses more on the application of banking and e-commerce, rather than streaming data. Interactive resource management strategy[12] is being introduced to boost response time efficiency based on streaming video variations. This approach is not based on the expense, number of users and streaming video form. However, the emphasis on the geographical locations and application server can further enhance this approach, but this extension is not being promoted and the architekture has not been used for live and high definition videshooting services like Netflix, Amazon prime etc. [14] 's research is somewhat similar to the previous approach[13] based on the need for prediction a new dimension called demand forecasting is introduced in architecture. The operating process for allocating the resources is focused on the time frame. The allocation of resources is based on constant time , based on the aspect of demand forecasting. Nevertheless, this method does not benefit users from various geographical locations. In [15] crowdsourced Multi-View Live Streaming System Architecture powered by quality of experience (QoE) has been proposed with the Greedy Minimal Cost Algorithm to minimize bandwidth, memory and so on costs and other resources. The services according to the job specifications are not offered by this method. The updated analytical hierarchy method was introduced in [16] to enhance the task planning and resource assignment, but this methodology does not focus on the response time that is necessary for estimating the costs of used resources. In order to tackle the resource planning problem in hybrid architecture includes the content distribution network and private and cloud data centers, a generalized problem is created [15] which is converted into a Nash negotiation problem and a geometrically efficient solution is achieved. Most of the current approaches analyze cost and efficiency based on the demand forecast model and time slot scheduling approach. The services are often scheduled or reserved for several hours in advance. The research is focused on events or actions created by the user. The activities are listed as searching, streaming, reading commentary and scrolling, with limited resources. notification. The greatest challenge is to develop a flexible architecture that incorporates all content streaming media for different devices.

3. Proposed Design

Figure-2 presents the definition of the cost efficient allocation of resources (CERA), while Section-4 provides information. The purpose of the work proposed is to minimize costs by providing services based on the customer 's events. In addition, it means that even though the minimum resources are used for such activities such as searches, scrollings, and so on, output or response time is the same. This research focuses on: memory, bandwidth, number of users, data centers, number of virtual machines. etc. The proposed Cost Efficient Resource Allocation Scheme for Media Sharing (CERA) is demanded by cloud users on different devices. The CESRA model has various components which allocate resources on the basis of events. The following are the functions of every component.

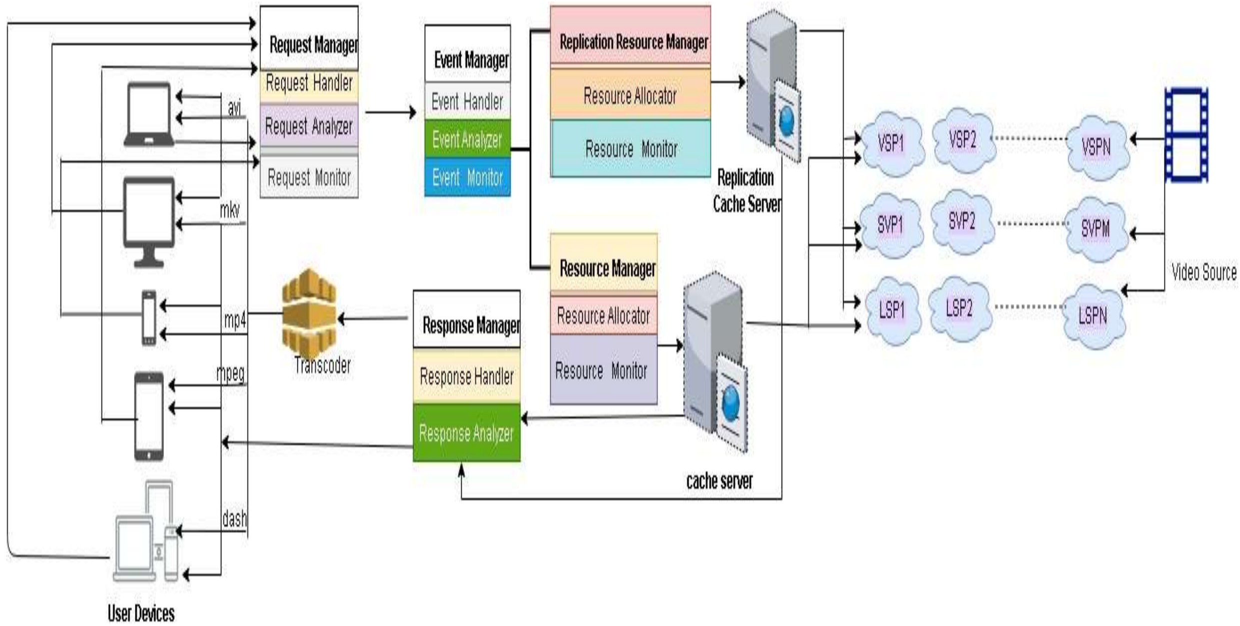

Figure-3 Architecture of CERA

Request Manager: It includes request handler, request analyser and request monitor. It is used for the administration of all user requests.

a) Request Handler-it is used to process and forward the incoming request to the request analyzer

b) Request Analyzer-The service needed is evaluated and checked by the Analyzer Analyzer and forwarded by the event planner**.**

c) Monitor Request – The request queue is monitored.

Event Handler: includes case manager, case analyzer and event controller. event manager. It is used to monitor and interpret the events that the application analyzer receives.

Event Manager: It comprises of event handler, event analyzer and event monitor. It is used to maintain and analyze the events received from the request analyzer.

a) Event Handler- This is used for handling the analyser 's request and passing it to the event analyzer.

b) Event Analyzer- The function is to define the type of event and to pass it to the resource manager or manager for replication. For example, when the event analyzed is browsing, trending or display unless the event is played the resource manager will be transferred.

c) Event Monitor- It is used to monitor the event queue.

Resource Manager: It consists of resource allocator and manager of resources. It keeps a list of all the tools available for potential use.

a) Resource Allocator- It is used to allocate the minimum set of resources required to complete the task and transfers it to the cache server to provide the service based on the allocated resources. ***b) Resource Monitor-*** It is used to manage all the resources available, and to keep track of the resources used.

Replication Resource Manager: It is used for monitoring and tracking all available resources.

a) Replication Resource Allocator: This is used to allocate and pass the required collection of maximum resources to the replication cache server to provide the service based upon the allocated resources.

b) Replication Resource Monitor- It is used to monitor all resources available and to keep tracks of the resources used.

Response Manager: This is used for providing service response to a request that the user sends. It consists of an answer hander and an analyzer of the response.

a) Response Handler: It is used to handle the response received from the server cache or replication cache and transfer it to the response analyzer.

b) Response Analyzer: It is used to assess response type. The category here refers to an incident or a media related incident. If the response is media-based, it is passed to the transcoder otherwise guided to the user devices by the response analyser

4. Architecture Overview

Media streaming is the transmission of audio and video files from a server to a device over the Internet or a cellular data network. The proposed architecture is comprised of on demand and event driven based service request. It consists of user devices, cloud servers, cache server and transcoder.

User Devices: The free as well as subscribed users submit their request using various devices to the proposed model. It uses request manager to handle all the request.

Cloud servers: It is also called as cloud data centers which is used as a video repository or video source. In this proposed model video source is to be assumed from three type of streaming providers. They are Video Sharing Provider (VSP), Streaming Video Provider (SVP) and Live Streaming Video provider. As mentioned in section 1 video sharing provider is free user such as you tube , daily motion etc., while other two are subscribed or paid users.

Cache Server: A Cache server is a dedicated network server or server which saves web page or other web content locally. A cache server also accelerates access to data and reduces demand for the bandwidths of an organization by putting requested information in temporary or cached storage**.** The proposed cache server has been built in the mode Content Delivery Network (CDN) edge server that stores cache in strategic or different locations in order to take the load off from cloud servers. A content delivery network (CDN) refers to a geographically distributed group of servers which work together to provide fast delivery of Internet content. In the proposed model adaptive bit rate with progressive streaming is added in content delivery network. The adaptive bit rate process is to transform a stream of videos to fragments or chunks, often 2-10 seconds in length. ABR generates separate streams at different rates and uses input from the video player of the internet user to determine the optimal video clip network speed dynamically. In case if there is change of network conditions the stream with the most appropriate bitrate is served for that give chunk of time. It also benefits the CDN by limiting the number of specialized streaming servers (resources) needed for on-demand video delivery. It also acts as a firewall to protect the cloud data centers. Very few request will sent to the cloud datacenters.

Transcoder: It is a dedicated cloud server that converts a video file from one format to another, to make videos viewable across different platforms and devices. In the proposed model it receives the response from the response manager and provide its service to the various devices.

Algorithm-1 CESSMS

Input : Request to watch or browse streaming videos
Output: Evaluate the cost and other resources consumed
Step 1:
User submits the request such as browsing, scrolling, comment, notifications and playing from various devices.
Step 2:
Request analyze and confirms the request as a legal one.
Step 3:
Event analyzer identifies the event type and sets the flag =1 for play and 0 for other events.
Step 4:
If flag==1
Configure the maximum resources for each event submitted
Goto replication cache server
Else
Configure the minimum resources for each event submitted
Goto cache server
Step 5:
Both servers sent their service task to the response manager.
Step 6:
Response Handler receives the request and transfers to the response analyzer.
Step 7:
Response analyzer analyze the task or event
If event is a media file response go to the transponder and it converts the media send to the appropriate devices.
Else
Response manager will sent the response to the appropriate devices.
Step 8:
Obtain the details of the resource allocated which is in the resource managers and compare with other approaches.
End

Algorithm-1 explains the user request from various devices are interpreted by the request analyzer and it transfers to the event manager. The event analyzer identifies type of event raised by the user and transfers the control to the resource managers. Resource allocation provisions are based on the type of event so that for browsing, trending, scrolling, commenting or reading notifications events limited resources can be allocated with minimum cost. For streaming media maximum resources can be allocated. In the extreme cases, service can be obtained from the cloud datacenters. At that time only user has to wait for few seconds more than the average response time. In normal cases, cache server gives the response very quickly for any number of users with minimum cost. It uses Content delivery network edge server along with adaptive bit rate technique to transfer the video file efficiently. With compatible devices the transcoder is used to translate the video into different formats.

5. Implementation

The Enhanced Cost Effective Scalable model for media streaming service is proposed to provide following benefits:

- Allocate resources based on the demand and the events generated by the users
- Increase interdependency among the cloud service providers.
- Provides support for video sharing, video streaming and live video streaming using single architecture.
- Reduced cost and over provisioning of the resources
- Highly elastic in nature
- Better Quality of services in terms of response time.

A. Experimental Set up

The proposed model is evaluated using cloud analyst. This software is implemented on a standalone system that uses the following configuration: Windows 7 Ultimate 32-bit, Intel Core (TM) 2 Duo CPU T6500 running at 2.10 GHz with 4 GB of DDR3 RAM. Cloud analyst is configured for minimum and maximum resource usage. Since our proposed model is based on event type, the resource replication manager increases the amount of required resources based on certain events, such as play so the data center be considered as a cache server, the number of tasks as an incoming request, the task size as a web page size (home page).Therefore, two sets of experiments must be carried out to calculate the total cost. For the minimum resource utilization cloud analyst is configured as follows: Data Center used-1, task size as size of video streaming web page size is 3 MB per request, 4 virtual machines of 256 MB RAM, bandwidth is 0.3Mbps, data transfer cost is 0.05$ per hour, cost per VM per hour 0.1$, memory cost 0.05$ per hour and storage cost is 0.1$ per hour for events such as browsing or searching the videos, scrolling, reading or writing the comment, trending and notification so on except watching or playing the video.
For maximum resource utilization configuration of cloud analyst is Data Center used-2, task size as size of video streaming data is 400 MB per request, 5 virtual machines of 512 MB RAM, bandwidth is 100 Mbps, data transfer cost is 0.05$ per hour, cost per VM per hour 0.1$, memory cost 0.05$ per hour and storage cost is 0.1$ for play or watch event, 1 CPU and cost as specified in [2][19][20][21]. The web page size of the you tube or daily motion or any video streaming is obtained using google chrome features such as developer tools. The average size of any streaming video web page is 3 Mb. For loading 3Mb web page minimum resource configuration is specified above and it is obtained from [20].Round Robin Scheduling algorithm is used in the cloud analyst and cost is calculated for hourly basis. Table-1 shows the total cost for the minimum resource used by the similar type of cloud users that generates the similar events such as browsing or searching etc., using the proposed scheme. Most of the existing approaches while allocating the resources does not consider the number of users and the size of the task generated by the user. Table 1, 2 and 3 shows the comparison of proposed scheme with Efficient Multi-Resource Scheduling Algorithm (EMRSA) and Dynamic Pricing Scheme (DPS).

Table-1 Comparison of total cost and response time with other existing schemes using 900 tasks, 100 users with task size of 3MB

Features	EMRSA[17]	DPS[21]	CESSMS (Proposed)

No.of Tasks/hour	700	900	900
Task Size(Mb)	3	3	3
No.of users/hr	-	-	100
RAM per VM	1 GB	1 GB	256 Mb
Bandwidth	1000M bps	1000 Mbps	100Mbps
Response Time	0.2 sec	0.4sec	0.3 sec
Total Cost per Hour per user in $	2.59	1.89	0.156
Number of VMs	5	8	4

From table-1, it is observed that total cost is the cost generated for 900 request by the number of users. Table-2 shows the total cost for the maximum resources used by the similar type of cloud users that generates the similar events such as playing or watching the video using the proposed scheme. The total cost in the table-2 depends on the number of request generated by the users who receive their response in time. Total cost is based on the 900 request not for the single request made by the user for an hour. Total cost is calculated using the cost generated by the simulator for the number of request stimulated by the number of users.

Total cost per user=Total cost generated by the simulator/Number of users

Total cost per request=Total cost per user/number of request.

Table-2 Comparison of total cost and response time with other existing schemes using 900 tasks, 100 users with task size of 400MB

Features	***EMRSA [17]***	***DPS[21]***	***CESSMS (Proposed)***
No.of Tasks/hour	250	900	900
Task Size per request(Mb)	-	400	400
No.of users/hr	-	-	100
RAM per VM	1 GB	1 GB	512 Mb
Bandwidth	1000M bps	1000 Mbps	100Mbps
.Response Time	2.5 sec	2.0 sec	2.5 sec
Total Cost per Hour per user in $	1500.65	1653.56	408.17
Number of VMs	5	8	4

The stimulation is further extended by adding the number of users that generates similar request to test the proposed model in various circumstances. From table-3 it is observed that maximum number of task has been generated by the simulator in table -2 because the all the values remains the same in table-3 when users are further added in a random manner for the given configuration as mentioned in section 5.A. or below.

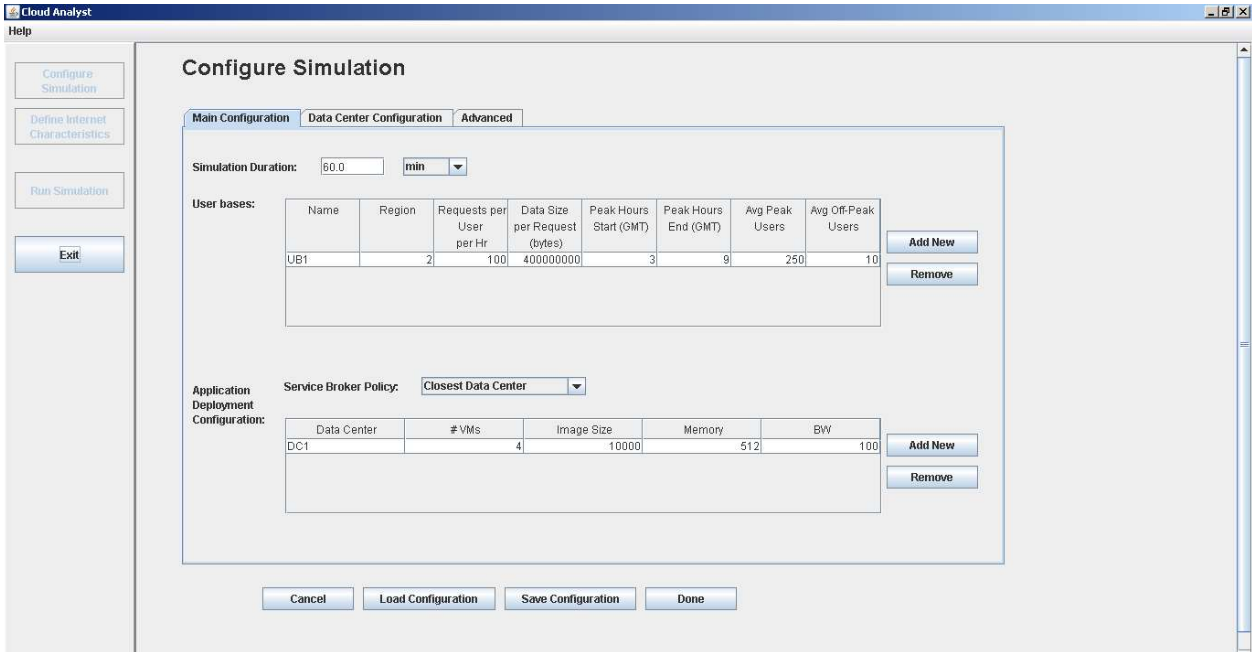

Figure-3 Main Configuration used in Table-3

Figure-3 shows the user base configuration or main configuration which is used to simulate the result for the proposed scheme. Similarly, figure-4 shows the datacenter configuration used for the proposed approach. The experiment has been extended by varying the number of the request in table 4. However the datacenter configuration and the number of users remains the same as in table-2 or 3.

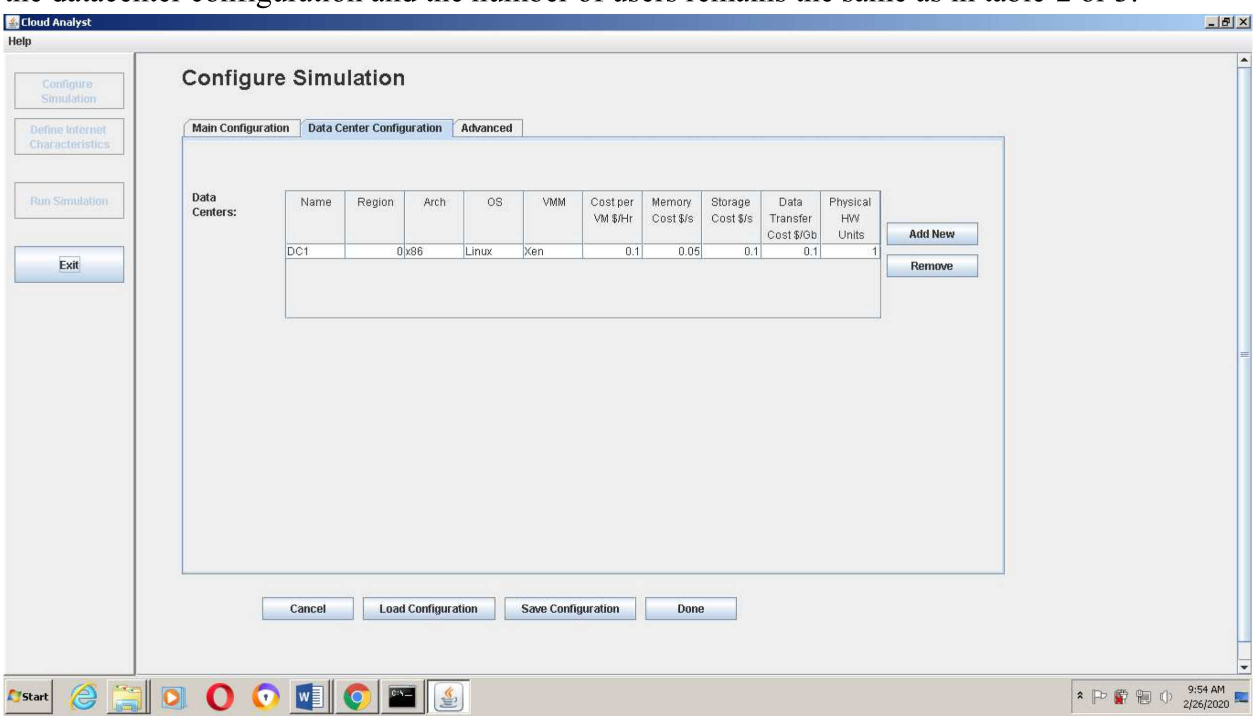

Figure-4 Data Center Configuration used in Table-3

Table-3 Comparison of total cost and response time with other existing schemes using 400 tasks, 250 users with task size of 400MB

Features	***EMRSA [17]***	***DPS[21]***	***CESSMS (Proposed)***
No.of Tasks/hour	250	900	900
Task Size per request(Mb)	-	400	400
No.of users/hr	-	-	250
RAM per VM	1 GB	1 GB	512 Mb
Bandwidth	1000M bps	1000 Mbps	100Mbps
Response Time	2.5 sec	2.0 sec	2.5 sec
Total Cost per Hour per user in $	1500.65	1653.56	408.17
Number of VMs	5	8	4

Acronyms used in Table 5

TS -Task Size
NOT -Number of Tasks
RT -Request Type
TCO/HR/USR/req -Total Cost for single user for single request per Hour
EMRSA - Efficient Multi-Resource Scheduling Algorithm
DPS -Dynamic Pricing Service
CESSMS -Cost Effective Scalable Scheme for Media Streaming

Table-5 Comparison of scalability features of the proposed model in terms of number of tasks, size of the task, total cost and response time with existing models

	Existing Cost Based Scheduling Schemes								**Proposed**			
S.NO	***EMRSA[7]***				***DPS[21]***				***CESSMS***			
	NOT	***TS (MB)***	***RT(S)***	***TCO/HR ($)***	***NOT***	***TS(MB)***	***RT (S)***	***TCO/HR ($)***	***NOT***	***TS MB)***	***.RT (S)***	***TCO/HR /USR/req ($)***
1	100	400	0.9	0.57	100	400	0.9	0.75	100	400	0.9	0.18
2	200	400	0.91	0.60	200	400	0.92	0.80	200	400	0.9	0.19
3	300	400	0.92	0.59	300	400	0.95	0.79	300	400	0.93	0.19
4.	400	400	0.9	0.59	400	400	0.93	0.77	400	400	1.0	0.18
5	500	400	1.0	0.58	500	400	0.99	0.79	500	400	1.5	0.17

The two other existing approaches in table-5 use the proposed input parameters, such as the number of tasks and the size of tasks, using their data center configuration to obtain the results.

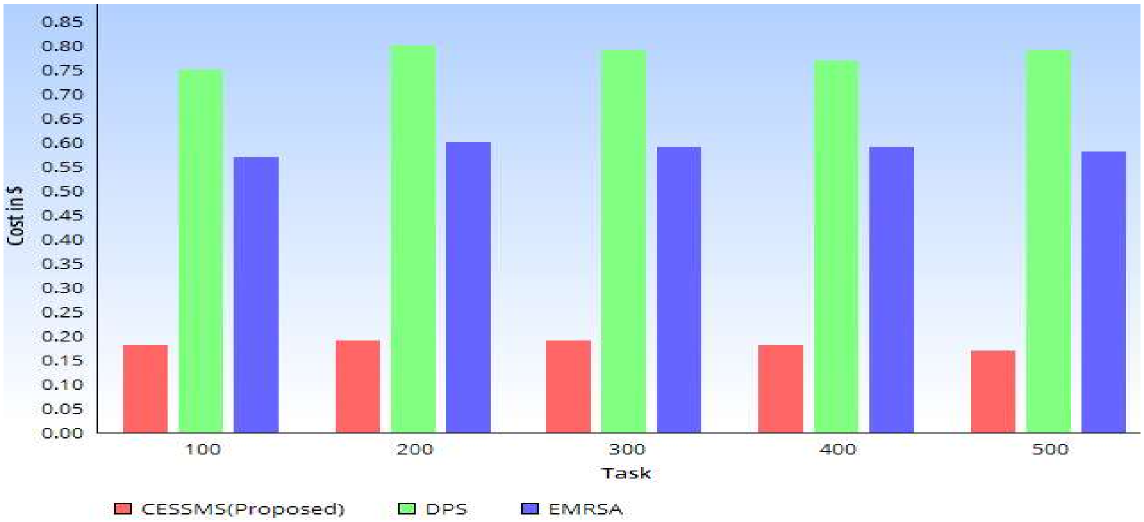

Figure-5 Comparison of Proposed CESSMS task vs cost with other models

Total Cost: It comprises of the cost of used resources, storage cost and data transfer from datacenter to receiver's machine or device.

Figure-5 shows a comparison between the current schemes and the Cost Efficient Scalable Media Streaming Scheme (CESSMS) proposed scheme. The tasks are carried out in the x-axis while the total cost (price) per hour is carried out along the y-axis. It is noted that the proposed scheme is performing better in terms of resources, response time and cost. It is because the resources are reserved for the existing approaches before they are allocated. In other words, the provision of resources before identifying the type of request used to provide their service. It is also to be noted that the total cost remains same because the total cost is calculated for single user per request for the duration of 1 hour. In table 1, the total cost is calculated for total number of request per hour per user. It is shown from Table 1 and Table 5 that when the size of a task increases, the total cost increases as well.

Response Time: It is the overall time taken to complete the total number of request from the user device to cache Content Delivery Network (CDN) server and vice versa.

6. Future Enhancement

Scalability can be a major barrier when it comes to delivering video assets, whether live or on-demand. To generate more revenue and high resource efficiency, service providers need to schedule resources and deploy different user applications, but existing cloud scheduling strategies are geared towards using a single objective, such as execution time, rather than several criteria being considered. Scheduling and deploying multiple applications present the cloud providers with a new set of challenges. Importantly, different changes in the current scheduling, monitoring and resource management algorithms should be applied to ensure the QoS requirements. For example, according to the agreed SLAs, deciding which jobs should be run in a specific context (e.g. peak time) would be a different issue.

Conclusion

A new event-driven architecture for cloud-based resource allocation for streaming media services is proposed here in this work. It uses the event analyzer and the size of the task to allocate resources along with the on-demand service. The proposed scheme supports various streaming services and also various devices. This work also ensures that data centers are protected by a cache server that acts as a proxy server. The experimental results indicate that the proposed model is a choice to provide a better solution for reducing costs and using resources more effectively. The empirical results demonstrate that the scheme proposed is scalable and effective when allocating resources to a streaming media service.

References:

[1] A.Anandhavalli, F.Leo John, "End User Issues in Cloud -A Survey", International Journal for Research and Development in Technology Vol.7, No.6, 2017, pp.354-357.

[2] C.Venish Raja, Dr.L.Jayasimman, "A Cost Effective Scalable Scheme for Dynamic Data Service in Heterogeneous Cloud Environment", International Journal of Advanced Science and Technology, Vol.28. No.20, 2019, pp.764-776.

[3] Media streaming types https://vidooly.com/blog/video-sites/

[4] Migfu Li, hen-Lin Yeh, Shao-Yu Lu, "Real Time QoE Monitoring System for Video Streaming Services with Adaptive Media Layout", International Journal of Digital Multimedia Broadcasting, Vol.1, 2018, pp.1-11.

[5] Madhumtha R Keerthana V Dhivaya V and ThamilzhSelvi S.P, "Cloud Scheduling in Video Transcoding," Proceedings of the 7th International Conference on Technical Advancements in Computers and Communications (ICTACC), Melmaruvathur, India, 2017, April 10-11.

[6] Rohit Kumar Gupta, Rishav Hada, Sohi Sudhir, "2-Tiered Cloud Based Content Delivery Network Architecture", Proceedings of the Proceedings of the International Conference on Signal Processing and Communication(ICSPC), Coimbatore, India, 2017 July 28-29.

[7].Mahmoud Darwich, Ege Beyazit, Mohsen Amini Salehi, and Magdy Bayoumi, "Cost Efficient Repository Management for Cloud Based On-Demand Video Streaming" Proceedings of the 5th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering, San Francisco, USA, 2017 April 6-8.

[8] Rahma Gharsallaoui, Mohamed Hamdi, Tai-Hoon Kim "Video Streaming Strategy Based on Feedback Control Loop Theory in Cloud Platform" Proceedings of the 9th International Conference on Ubiquitous and Future Networks, Milan Italy, 2017 July 4-7.

[9] Y. Tian, R. Babcock, C. Taylor and Y. Ji, "A new live video streaming approach based on Amazon S3 pricing model," Proceedings of IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, 2018 January 8-10.

[10] S. S. Panhalkar and H. P. Khandagale, "Video Streaming Using Cloud Computing Based on Android Application," Proceedings of 4th International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018 August 16-18.

[11] S. Hoque and A. Miranskyy, "Architecture for Analysis of Streaming Data," Proceedings of IEEE International Conference on Cloud Engineering (IC2E), Orlando, FL, 2018 April 17-20.

[12] Cheng, Yingchao & Hao, Zhifeng & Cai, Ruichu & Wen, Wen & Wang, Lijuan and Zhou, Zhongrun, "Resource-Aware Stream Processing in High Performance Cloud Environment", Proceedings of IEEE 15th International Conference on Ubiquitous Intelligence and Computing, Guangzhou, China,2018, October 8-12.

[13] Reddy, P.M., Ahmed, A., Manjula, S.H. Venu Gopal, K.R., “Resource allocation in the cloud for video-on-demand applications using multiple cloud service providers”, Cluster Computing Vol.22, 2018 ,pp 223–239.

[14] A. Alasaad, K. Shafiee, H. M. Behairy and V. C. M. Leung, "Innovative Schemes for Resource Allocation in the Cloud for Media Streaming Applications," in IEEE Transactions on Parallel and Distributed Systems, vol. 26, no. 4, 2015, pp. 1021-1033.

[15]. Bilal, K., Erbad, A., Hefeeda, M., QoE-aware distributed cloud-based live streaming of multisourced multiview videos, Journal of Network and Computer Applications 2018, pp.1-38.

[16] Gawali, M.B., Shinde, S.K., "Task scheduling and resource allocation in cloud computing using a heuristic approach", Journal of Cloud Computing Advances, Systems and Applications, Vol. 7, 2018,pp 1-16.

[17] Yang Liu, Wei Wei, Heyang Xu, "Efficient multi-resource scheduling algorithm for hybrid cloud-based large-scale media streaming", Computers and Electrical Engineering, Vol.75, 2019, pp 123-134.

[18].what is content delivery network and its benefits https://www.streamingmedia.com/Articles/Editorial/What-Is-.../What-Is-a-Content-Delivery-Network-(CDN)-74458.aspx

[19] "Pricing - Windows Virtual Machines | Microsoft Azure", Azure.microsoft.com, 2019. [Online]. Available:https://azure.microsoft.com/en-in/pricing/details/virtual machines/windows/. [Accessed: 27- Oct- 2019].

[20].Manish Ranjan, Abdul Kayum, M.suri, “Broadband definition and standardization of Measurement method in the context of developing countries like India having large rural population” Telecom Engineering Center, Department of Telecommunications, Government of India. Tec.gov.in, 2019. [Online]. Available:http://tec.gov.in/pdf/Studypaper/Study%20paper%20on%20BB%20Definition.pdf.

[21] "What is Bandwidth? - Definition and Details", Paessler.com,2019.[Online].Available: https://www.paessler.com/it-explained/bandwidth.

[22] N. binti Muhamad Shaari, T. Ang, L. Por and C. Liew, "Dynamic Pricing Scheme for Resource Allocation in Multi-Cloud Environment", Malaysian Journal of Computer Science, vol. 30, no. 1, 2017, pp. 1-17.

[23] How to create video streaming webiste and type of parameters to be focused https://codetiburon.com/create-video-streaming-website-like-netflix-amazon-hulu/