

An Extensive Survey over Traffic Management/Load Balance in Cloud Computing

Amith Shekhar C

Asst. Professor
Dept. of IS&E

GM Institute of Technology, Davangere,
Karnataka, India

Dr. Sharvani. G S

Associate Professor
Dept. of CS&E

RV College of Engineering, Bengaluru,
Karnataka, India

Abstract—Cloud Computing (CC) is all about carrying out processing in other's system. There are various vendors who provide CC services. The basic algorithm that should be met to access CC services is a need for steady internet connection. As everything is done online the traffic across the internet is to be managed efficiently so that the transmission delay can be minimized and better quality of service can be given to the customers. The network should not be too congested at any moment of time. Hence the traffic management becomes a crucial factor for the better performance of the CC network. This paper addressed the most valuable terms and topics concerning the load balance/traffic management in cloud computing. Also, the paper is meant to discuss the study analysis of the recent researches in load balance of CC. From the study analysis the current research gap is addressed with a future scope of research to overcome the research gap.

Keywords—Cloud Computing; Load Balance; Traffic Management

I. INTRODUCTION

The quick advancement of storage and processing technologies and the accomplishment of the Internet, computing resources have turned out to be less expensive, more intense and more pervasively accessible than any time in recent time. This technological pattern has empowered the acknowledgment of another computing model called Cloud Computing (CC), in which resources (e.g., storage and CPU) are given on rented use and discharged by clients through the Internet in an on-demand mold. In a CC environment, the conventional part of service supplier is partitioned into two: the infrastructure vendors who oversee cloud stages and rent resources as per a utilization based estimating the model, and service providers, who lease resources from one or numerous infrastructure suppliers to serve the end clients. The rise of CC has had a huge effect on the Information Technology (IT) industry. In recent years, where the organizations, for example, Amazon, Microsoft, and Google are using CC to give all the more powerful, stable and cost-proficient cloud stages, and business ventures try to reshape their business models to pick up advantage from this paradigm [1 2].

The load in a cloud is be founded on memory required, CPU (Virtual Machine) limit which will accommodate for finishing of client job. Cloud is an innovation given distributed environment so sharing of work among various resource supportive to enhance the use of comprehensive resources and

achieve great execution. Also, load balancing is a procedure of guaranteeing the consistent balance of workload on the pool framework hub or processor. Load balancing mechanism predominantly characterized in two types: Dynamic and Static balancing. In static balancing earlier availability of resources is required, so moving of load not relies on upon current condition of resources. This balancing helps in handling low variation load. In Dynamic balancing is done as per the load variation. For this ongoing communication with the network is required which can build the traffic all through a network. Thus, like static, element additionally check the present condition of resource it is possible that they are loaded [3].

The current traffic analysis and management techniques can't be effectively reached out to the data centers. The thickness of connections at the data centers is much higher than the thickness of connections at the undertaking network. This presents the issues for the current techniques to be connected to gauge traffic over the CC system. A significant portion of the current strategies is utilized for the hard level of traffic management. These techniques are fit for investigating traffic over the higher number of hosts. For a particular server may have a few thousand servers; this is the place the current techniques fails. Since these strategies are intended for the hard networks, that accepts the stream designs of sensible in the Internet and undertaking networks. Also, with regards to the cloud system, it is unrealistic to allow under traffic variation. The frameworks conveyed on cloud stage should be adaptable since the information traffic over the framework may change at any time [4 5]. The traffic management choices are frequently made in a unified way. This prompts to high multifaceted nature and poor scalability. This paper presents the essential aspects of the load balancing/traffic management in the cloud and also highlights survey of the existing researches in load balancing along with the research gap. Also, significant future line of research is addressed.

The section wise discussion of this paper is categorized as Section II - conceptual description of Cloud computing along with architecture, business model, and cloud types. In section III the existing methods for traffic management are mentioned, then recent research survey of load balance/traffic management are discussed in section IV. The research gap in the existing research is addressed in section V while the future study analysis is given in section VI. Finally, the conclusion is discussed in section VII.

II. BASIC CONCEPTS OF CC

The evolution at the end of the twentieth century to the present day facilitates of pervasive computing; the web has changed the computing. It has gone from the idea of parallel computing to distributed computing to matrix computing and as of late to CC. Also, the possibility of CC has been around since long ago; it is a developing field of software engineering. CC can be characterized as a computing environment where computing needs by one gathering can be outsourced to another gathering, and when need emerges to utilize the computing force or resources like database or messages through the web [6]. CC is a late pattern in IT that moves computing and information far from desktop and compact PCs into substantial data centers. The primarily favorable position of CC is that clients don't need to pay for infrastructure, its establishment, required labor to handle such infrastructure and support.

A. Cloud Computing

In recent years, CC accomplished milestones of turning point past the conviction. The existing researches will supports to increasing cloud issues. In this realized that cloud is an interconnected fast network which permits incredible facility such as elasticity, on-demand resource provisioning, usage of resource given ordering, pays for what you require. Achieving proper resource balance for fulfillment of client's requirement is main task in cloud services. In CC, clients get to the information, applications or whatever other services with the assistance of a browser to the devices utilized and the client's area. The Cost is lessened to an exceptional level as the infrastructure and need not be obtained for periodically escalated computing issues [7].

CC gives a few striking elements that are unique about conventional service computing, which is given below:

- *The Multi-Tenancy:* In a cloud environment, services possessed by numerous suppliers are co-situated in a data center. The execution and management issues of these services are shared among service suppliers and the infrastructure supplier. The layered engineering of CC gives a characteristic division of responsibilities: the proprietor of every layer just needs to concentrate on the particular functionalities connected with this layer.
- *Resources pooling:* - The infrastructure supplier offers a pool of computing resources that can be powerfully assigned out to numerous resource users. Such dynamic resource task ability gives much adaptability to infrastructure suppliers for dealing with their particular resource use and working costs. For example, an Infrastructure as a Service (IaaS) provider can influence VM relocation mechanism to accomplish a high level of server combination, henceforth expanding resource usage while minimizing cost, for example, cooling and power utilization.
- *Network Access and Geo-Distribution:* - Clouds are by and large open through the Internet and utilize the Internet as a service conveyance network. Consequently, any devices with Internet network like

cell phone, a PDA or a tablet can get to cloud services. Also, to accomplish high network execution and localization, a large number of today's clouds comprise of data centers situated at numerous areas around the world. A service supplier can without much of a stretch influence geo-distribution to accomplish greatest service utility.

- *Service Based:* - The CC receives an operation model of service. Subsequently, it puts a solid accentuation on service management. In a cloud, every Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) supplier offers its service as indicated by the Service Level Agreement (SLA) consulted with its customers. The SLA assures service with basic target of each supplier.
- *Provisioning of Dynamic Resources:* - This is the main point in CC is that computing resources can be acquired. In comparison with the conventional model that arrangement resource as per provision of dynamic resources, peak demands permits service providers to obtain resources given the present demand, which can extensively bring down the operating cost.
- *Utility-based Valuing:* - CC utilizes a usage based estimating model. The correct valuing plan may change from service to service. For instance, a SaaS supplier may lease a virtual machine from an IaaS provider on every hour premise. Then again, a SaaS provider that gives on-demand customer relationship management (CRM) may charge its customers given the quantity of customers it serves. Utility-based evaluating brings down service working cost as it charges customers on a for every utilization premise.
- *Self-Organization:* - The resources can be allocated and de-allocated based on its demand; the service suppliers are engaged in dealing with their resource utilization as per their needs. Besides, the automatic management of resources highlight yields high agility that empowers service provider to react rapidly to fast changes in service demand.

B. Architecture of CC

The architectural diagram of a CC environment can be separated into four layers: the equipment/server farm layer, the infrastructure layer, the stage layer and the application layer, as appeared in Fig. 1.

- *Infrastructure Layer (IL):* - This layer is also virtualization layer (VL), the IL makes a pool of storage and computing resources by dividing the physical resources utilizing virtualization technologies, for example, Xen and VMware. The IL is a basic segment of CC, since many key components, for example, dynamic resources, are just made accessible through virtualization technologies.
- *Platform Layer (PL):* - This is kept above the IL; the PL comprises of working frameworks and application systems. The reason for the PL is to minimize the burden of conveying applications specifically into VM

units. For instance, Google App Engine works at the PL to give API support to actualizing database, business and storage logic to execute web applications.

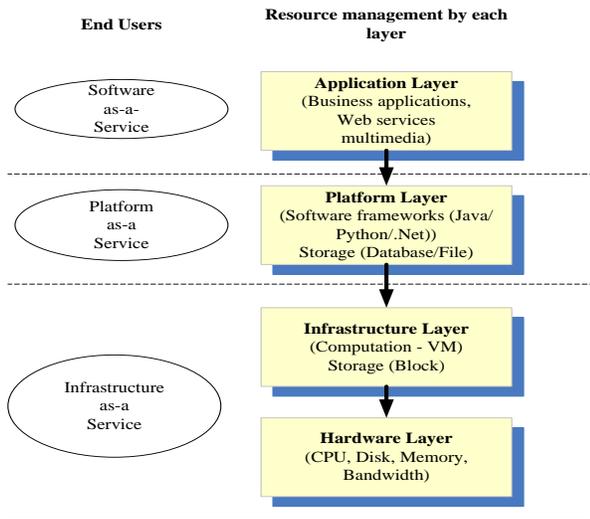


Fig. 1. CC Architecture

- **Hardware Layer (HL):-** This layer is used to manage the physical resources for the cloud, also includes routers, physical servers, cooling and power frameworks. Practically speaking, the HL is normally executed in data centers (DC). A DC more often than not contains a huge number of servers that are sorted out in racks and interconnected through routers, switches. The issues related with HL incorporate hardware setup, adaptation to non-critical failure (fault tolerance), cooling resource and traffic management.
- **Application Layer (AL):-** The significance of AL in a high-end hierarchy is that AL comprises of the real cloud applications. The cloud applications can influence the programmed scaling highlight to accomplish better execution, accessibility, and lower working expense.

In comparison with the conventional allocations, the cloud-based scaling can offer better performance and low operating cost. The design particularity permits CC to support an extensive variety of utilization necessities while decreasing maintenance and management overhead.

C. Business model of CC

The CC utilizes a service based business model or HL and PL resources are given as services on an on-request premise. The each layer of the design depicted in the above unit be executed as a support of the layer above. The client perspective of every layer is discussed as Software as service (SaaS), Infrastructure as a service (IaaS), and platform as service (PaaS).

- **Infrastructure as a Service (IaaS):-** This is an on-request provisioning of infrastructural resources like Virtual machines (VMs). The cloud service provider can offer IaaS. For example Amazon, Flexiscale and GoGrid.

- **Platform as a Service:-** This is a PL resource, composed of operating system support and programming improvement. Example: Google App Engine, Force.com, and Microsoft Windows Azure.
- **Software as a Service:-** This offers service to on-demand applications by using the Internet. Example: SAP Business ByDesign, Rackspace, and Salesforce.com.

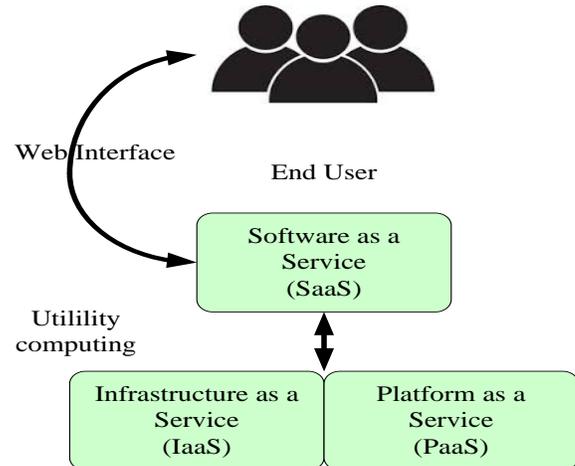


Fig. 2. Business Model of CC

The business model of CC is represented in Fig. 2. As indicated by the layered design of CC, it is totally conceivable that a PaaS provider runs its cloud over IaaS provider's cloud. Presently, the IaaS and PaaS providers are frequently parts of a similar association (Salesforce and Google). This is the reason PaaS and IaaS suppliers are regularly called the cloud or infrastructure suppliers.

D. Types of Cloud

In the cloud, there exist some issues moving an enterprise application to the cloud environment. Hence service providers are focusing on minimization of operation cost, while others may need high reliability and security. Thus, there are different types of clouds, each with its benefits and drawbacks:

- **Public Cloud (PC):-** A cloud in which specialist organizations offer their resources as the service platform to the world. The PC offers a few key advantages to providers, including no initial investment for infrastructure and moving of risks to the infrastructure provider. The control of data, system and security settings of PC can hamper their viability in numerous business situations.
- **Private Cloud (PrC):-** This is an internal cloud are intended for restrictive use within the single organization. A PrC might be constructed and managed with the by the external providers and cloud. A PrC offers the most astounding level of control over execution, security, and reliability. The frequently used

conventional server doesn't offer significances like low cost.

- *Hybrid Cloud (HC)*:- An HC is a combination of PC and PrC that tries to address the issues of every approach. In an HC, unit of infrastructure service keeps running in PrC while the rest of the part keeps running out on the PC. The HC offers more adaptability than both PC and PrC. HC give more control and security over application information than PC, while as yet encouraging on-request service contraction and expansion. The drawback is that designing of HC requires precisely deciding the best split amongst PC and PrC parts.
- *Virtual Private Cloud (VPrC)*:- This is a PrC which solves the limitations of PC and PrC. A VPrC is a platform running on top of PC. The principle change is that a VPrC influences virtual private system (VPN) mechanism that permits service providers to develop an own particular topology and security settings like firewall rules. VPrC is a more all holistic design since it virtualizes servers and applications, as well as the basic applications too. In most of the organizations, VPrC gives consistent service.

For most service-based organizations, selecting the correct cloud model is reliant on the business situation. In the scientific computation, applications are placed on PC for cost-effectiveness. Apparently, certain clouds will be more significant than others.

III. EXISTING METHODS FOR TRAFFIC MANAGEMENT IN CLOUD

- The data analysis in the data center is more needful for the current cloud environment. Many web applications depend on data traffic analysis to offer client data optimization [8]. The current network operators need to have an understanding of data traffic in a network so that proper data management can be followed to avoid traffic. Currently, there are a few difficulties for existing data traffic analysis and management strategies in Internet Service Providers (ISPs) systems and organization to reach out to web data centers.
- The links density is much higher than the density of ISPs or undertaking systems, which puts forth the most exceedingly bad situation for existing techniques.
- Most of the existing techniques can figure traffic between hundreds of end server hosts, yet even a flexible data centers can have a few thousand servers.
- The existing methods more often than not expect some flow pattern based designs that are sensible in the internet and organizational network systems, yet the applications conveyed on data centers, i.e., use of Map Reduce permanently change the traffic flow pattern.
- There is more tightly coupling in application's utilization of the network, storage resources, and computing than what is seen in different settings.

Right now, there is very little work on traffic analysis and management. Some of them are addressed below.

E. Virtual Private Network (VPN):

The simplest technique in which an organization can get to a CC application is through the Internet/VPN. It needs a little data synchronization between the cloud have and the undertaking data centers. The utilization of VPN puts just a little effect on the network enterprise, but it can bring about technical aspects that ought to be settled in both specialized as far as the agreement with the CC provider. The VPN is very complex to configure, and its maintenance is costly. The principle drawbacks of utilizing the VPN on the cloud for traffic service are in the cloud VPNs are difficult to scale, nearby VPN customers won't work the vast majority of the circumstances, records like to what extent the association was utilized, who got to your server are difficult to keep up. The VPN, however, is the simplest technique for an endeavor to get to the cloud organize.

F. Microsoft Azure Traffic Manager

This is an organizational Azure cloud service, or website might keep running on various data centers over the world. To control the traffic distribution to organizations predefined endpoints one can make utilization of the Traffic Manager. The Traffic Manager applies an insightful approach motor to DNS questions for the area name of your Internet resources. The arrangement of Traffic Manager enhances the accessibility of basic applications, responsiveness for performance applications. It permits overhauling and performing QoS maintenance without downtime. The execution of an extensive complex framework is upgraded by traffic distribution.

G. Cloud Network Management (CNM) Model

The CNM model demonstrates utilizes an arrangement of specialists sending redesigns to the administrators in the cloud about their execution. Every operator contains an arrangement of articles called Management Information Base (MIB) that stores the execution and other related data. To give better QoS of cloud services the system director ought to know about the present status of the chief in the group, their CPU, stockpiling and system usage, what numbers of examples of a virtual machine are assigned, etc. With a specific end goal to guarantee appropriate service every one of the messages in the CNM model is recognized. The CNM model is a half of unified and decentralized service. The CNM model is a half and half of brought together and decentralized approach. The provisos in the SNMP demonstrate prompt to the improvement of CNM model. It evacuates or minimizes a couple of pitfalls of the previous model. It gives upgraded security than the SNMP display [9].

CNM model display upgrades the system execution: utilization of less number of bundles decreases the jitter it which like this improves the system execution.

- The arrange traffic is diminished
- There are no surveying issues.
- Ensures the safe correspondence

- The concept of virtualization aides in the quicker recuperation in the setting of disappointments.
- It accommodates better security.

A system is always breaking down in light of the parameters, for example, execution, security, speed, adaptability. The execution of a cloud system ought to be broken down from both specialist co-ops' view furthermore from the client's viewpoint. Specialist organization's view point: The expert organization might be more worried about the foundation execution of the cloud arrange. Cloud expert organization needs to screen the capacity, VMs, and system traffic. Client's perspective: The clients have a tendency to choose the system execution on the premise of how the applications function; the speed of the system, availability of the remote information, and unwavering quality. In the wake of concentrate the above apparatuses, unmistakably dynamic traffic service is productive and a savvy decision for traffic service related issues over the distributed computing systems than the equipment arrangements.

H. VeriSign Traffic Management (VTM) Services

The Dynamic Traffic Management (DTM) benefit makes it conceivable to deal with the traffic in the system given continuous data. It permits for all intents and purposes boundless routes for manage based customization of association's traffic. VeriSign offers traffic service units, for example, failover, geo-allocation, weighted load adjusting, DTM. The Lua scripting dialect is solely accessible as a part of the VTM. Verisign gives dynamic traffic administration benefits over the cloud arrange. The worldwide associations can screen their system traffic effortlessly, the traffic examples can be checked, and generally in light of the fact that it is alert, the application downtime can stay away from by and large. This is a product apparatus to oversee traffic over the system.

- Low cost: When contrasted with equipment arrangements, it offers to bring down working expenses.
- Since it is conveyed all-inclusive, it permits simple versatility
- As it is a product bundle, it can be effortlessly transmitted
- It is reasonable for basic web-based services as a result of its upgraded accessibility and execution.
- Speed: the information can be conveyed to the goal much quicker as it offers the least dormancy or slack time.

The traffic over the cloud system is not the same at constantly, it differs. For instance, traffic over the VPN to get to the cloud may be diverse at various times. At the point when the traffic is less, it won't bring on any issues however if the traffic over the system is all the more; then, the system gets to be distinctly congested, and the odds of utilization downtime are high because of bottlenecks. In the distributed computing situation downtime is not acknowledged. The applications ought to run easily constantly [10]. Henceforth traffic management assumes an essential part. The system ought to

have the capacity to scale up and downsize contingent upon the necessities. The hardware arrangements get to be distinctly costly because when the traffic low.

IV. SURVEY OF EXISTING RESEARCHES

In this section, various existing researches towards the traffic management in CC and load balance.

The idea of Value of Service (VoS) based task scheduling for a CC system was represented in Tunc et al. [11]. The author presented another time-based value metric to empower scheduling algorithms considers the arrival task time and also task completion value and power utilization for a given period. Authors have analyzed the other existing scheduling algorithms of VoS. In this, an author has focused to implement proposed new time-of-use VoS metric, actualize it to work in a real-time framework, and build real VMs to execute benchmark applications. Authors have analyzed the proposed VoS system framework performance and out forms significant results.

Exploiting Geo-Distributed Clouds for an E-Health Monitoring System with Minimum Service Delay and Privacy Preservation is presented in Shen et al. [12]. In the framework, the schemes of resource allocation empower the distributed cloud servers to agreeably allocate the servers to the asked for clients under the load adjust condition. In this manner, the service delay for customers is minimized. The traffic shaping (proposed) algorithm changes over the client health and non-health data traffic that the ability of traffic analysis is to a great extent reduced. Authors demonstrate the proposed traffic shaping algorithms effectiveness as far as security maximization and minimization of service delay.

The work of Abranches and Solis [13] portrayed an algorithm based on response time and traffic demands to scale containers on a CC System. The presented algorithm depends on the web requests characterization and a PID (Proportional - Integral-Derivative) controller. The proposed mechanism was evaluated with a continuous arrangement got from an operational huge web framework in a controlled infrastructure. The outcomes demonstrate that the proposition accomplishes the normal response times allocating a lower container than other works.

The idea for versatile traffic management over the cloud data centers was presented in Assi et al. [14]. This novel innovation brings new difficulties, generally in the conventions that oversee its hidden challenges. The Traffic designing in cloud data centers is the major difficulties that have pulled in consideration from the research group, especially since the legacy protocols utilized in data centers offer constrained and un-versatile traffic management. Many supported for the utilization of VLANs as an approach to give adaptable traffic management; be that as it may, finding the ideal traffic split among the VLANs is the outstanding NP-Complete VLAN task issue. The extent of the hunting space of the VLAN assignment issue is tremendous, notwithstanding for little size systems.

In Brindha et al. [15] Agent-Based Bidirectional Bidding Mechanism for Efficient Scheduling of Real-Time Tasks in CC. In this a bidirectional declaration based bidding system to allow tasks and resources progressively. Likewise, it comprises

of three stages, i.e., basic matching, forward declaration bidding and in reverse declaration bidding stage. The scalability is fundamental when data scheduling is performed in progressively including VMs which gives flexibility. Authors have planned computation rules for both forward and in reverse declaration bidding system which helps for selecting contractors. The analysis is performed for both Google cloud and synthetic workload. The examination comes about demonstrate that agent-based scheduling that gives better execution contrasted with existing strategies.

In the work of Doyle et al. [16] Load Balancing the Cloud for Carbon Emissions Control (Stratus framework) is examined. The work analyzes the power cost, carbon discharges, and normal service request for time for an assortment of situations. The choice concerning how to adjust the different components will rely on upon SLAs, government legislation, and the cost of carbon on exchanging plans. Utilizing this data and the specifics of the cloud the administrator can run the cloud in the most attractive mold. The way of the service will figure out whether a cloud provider can execute this calculation while fitting in with service level understandings.

An enhanced Genetic algorithm utilizing populace decrease for load adjusting in CC is exhibited in Patel et al. [17]. Essentially cloud depends on utilize standard pay situation recognized by client's services. Be that as it may, for every single fulfilling, that services cloud needs some predefine necessity conditions to take after which influence distinctive parameters like reaction time, resource usage, adjusting load, ordering of resources and also employments and so on. Authors work concentrated on the use of resources and reaction time given the genetic algorithm, yet Authors changed that genetic algorithm with the assistance of fractional populace decrease technique that will fulfill the demand of client services.

A work of Ran et al. [18] has said the balancing backhaul stack in different cloud radio get to organize. This work proposed to adjust the information to be transmitted on backhauls of the remote radio heads (RRHs) to diminish the data transfer capacity of backhaul required by the RRHs or accomplish better execution with given backhaul limit uniquely in contrast to pressure strategies.

The work concentrating on work process scheduling for multi-inhabitant CC situations was introduced in Rimal et al. [19]. Multi-tenure is one of the key components of distributed computing, which gives adaptability and economic advantages to the end-clients and a group of service providers by having a similar cloud stage and its basic foundation with the disengagement of shared system and register resources. Be that as it may, resource service with regards to multi-inhabitant distributed computing is getting to be distinctly a standout amongst the most complex undertaking because of the intrinsic heterogeneity and resource isolation. The proposed calculation is contrasted and the best in class calculations, i.e., First Come First Served (FCFS), EASY Backfilling, and Minimum Completion Time (MCT) planning strategies to assess the execution. Facilitate, a proof-of-idea analysis of genuine logical work process applications is performed to show the

versatility of the CWSA, which confirms the adequacy of the proposed arrangement.

An intriguing Dynamic Fault-Tolerant Scheduling (DFTS) Mechanism for Real-Time Tasks in CC is presented in Soniya et al. [20]. In the current framework, Primary Backup (PB) model is utilized, yet it doesn't contain any dynamic resource assigning instrument. Authors propose a dynamic resource allotting system with adaptation to internal failure to enhance resource use. Authors fuse a reinforcement covering system and proficient VM relocation technique for outlining novel mechanism perform in distributed computing. The proposed show goes for accomplishing both adaptations to non-critical failure and high resource usage in the cloud. The analysis was analyzed utilizing irregular engineered workload, and Google cloud follow logs to test the effectiveness of the proposed model.

In Vascak et al. [21] an Agent-Based CC Systems for Traffic Management is delineated. The requirement for a safe and financially productive traffic represents a test for making traffic service frameworks. This work manages to combine three fundamental ideas, in particular, cloud-based advancements, operator based methodologies and fuzzy based cognitive maps to tackle the limitation. The proposed framework was straightforwardly tried on a play area, and acquired outcomes were investigated using a few chose the algorithm. At last, some further conceivable outcomes of potential use and future research are specified.

A critical work of Sundar Rajan et al. [22] presented a Workflow Scheduling in CC Environment utilizing Firefly Algorithm. The work scheduling mechanism plays a key part in getting most extreme advantage from the resources that are given. Another critical component to being considered about distributed computing is Load adjusting. This controlling of fill guarantees that each elite machine does the extremely same measure of work at any quick of time. To ensure this, we need to prescribe on utilizing fill controlling. Here in this archive, author prescribes heuristic algorithm known as Firefly algorithm for compelling fill controlling in thinking handling. This foundation depends on the travel conduct of the fireflies which go searching for the nearest conceivable greatest choices. Author utilize Firefly algorithm to plan the occupations and subsequently uniformly disseminate the heap and like this diminish the general finish time.

A Traffic-Aware Task Allocation for Cooperative Execution in Mobile CC is said in Wang et al. [23]. In this author, designs a task allocation mechanism to the mobiles with the objective that the aggregate traffic brought is least, while the imperative of the resources on mobile is not damaged. Given that, author additionally intends to minimize the aggregate number of mobiles took an interest in undertaking execution. A dynamic programming is proposed to finish the over two destinations separately. The simulation comes about exhibit the adequacy of the proposed plots in minimizing the traffic and the quantity of required mobile devices.

A work of Zhang et al. [24] has an effect on The Modeling of Big Traffic Data Processing Based on CC. The progress of research for data collection mechanism and huge data

processing model has a vital commitment to the improvement of urban traffic information. The control of urban transportation relies on upon the successful handling of continuous traffic perception information, which is typically the way of information escalated. The author studied an extensive number of floating car data processing (FCD) traffic control in the distributed computing environment, to take care of the issue of the new distributed computing innovation to take care of the issue of urban traffic control framework. The research comes about demonstrate that the distributed computing innovation, for example, HBase and Map Reduce can give huge information stream figuring of generous utility, for example, adaptability and ongoing count execution can be precisely by the proposed data storage, service, and parallel preparing model. The appropriateness and practicability of distributed computing are assessed for two runs of the mill information figuring errands of urban traffic checking, in particular, FCD question, FCD outline.

V. EXISTING RESEARCH GAP

The existing traffic measurement and analysis methods cannot be easily extended to the data centers. The density of links at the data centers is much higher than the density of links at the enterprise network. This makes the worst case scenario for the existing methods to be applied to measure traffic across the CC network. Most of the existing methods are used for the enterprise level network data traffic management. These methods are capable of analyzing traffic across few hundred hosts. But a modular data center may have several thousand servers; this is where the existing methods fail. Since these methods are designed for the enterprise networks, they assume the flow patterns that are reasonable in the Internet and enterprise networks. But when it comes to cloud network it is not possible to assume a pattern because the data traffic will be varying. The variation is not linear nor is it predictable. The systems deployed on cloud platform needs to be scalable since the data traffic across the system may change anytime. The traffic management decisions are often made in a centralized manner. This leads to high complexity and poor scalability.

The existing system is witnessed with the abundant literature for load balancing techniques over cloud environment. The majority of the techniques are found to adopt an approach that uses highly complex and sophisticated design where various probabilities of normalizing traffic congestion were not focused effectively. Hence, there is a need of adopting a technique that can cost-effectively lowered down the overheads of the traffic using multi-tenancy technique, efficient scheduling, and distributed, etc.

VI. CONTINUATION OF FUTURE RESEARCH

The further research can be followed to solve the existing issues of the traffic management in CC. In future, the work can be extended as:

- A review of research can be done, and the significant standard techniques of load balancing deployed in the existing system and explore its research gap.
- To design a novel and a simple algorithm of multi-cloud tenancy for the purpose of redirecting the traffic

based on its requirement to reduce the massive load of task scheduling is need to be analyzed.

- Also, develop a novel mathematical model for addressing the routing and scheduling issues considering virtual machines over cloud are too considered.
- A framework with a novel class of algorithms for distributed load balancing system over a large number of a cloud environment for ensuring resource availability is needed to design.
- The performance of the designed system can be measured with the existing system of load balancing.

VII. CONCLUSION

CC is a platform that is cost-effective and also provides faster means of data transmission. Hence traffic management across the CC network becomes critical. In order to reduce the transmission delays, few of the above-described solutions are used. Managing the network traffic based on the real time scenario will be a more appropriate solution. The network must be deployed in a scalable manner. This not only helps in easier traffic management but also provides better Quality of Service (QoS) for the customers. This paper has discussed the significant factors related to the traffic management and load balance in cloud computing.

With this paper, we have discussed the existing research gap which addresses that no such work is presented with better load balance in cloud with dynamic variation in the load. Also, the existing techniques are cost effective. The future scope of research to solve the load balance issues were addressed with proper resource allocation.

REFERENCES

- [1] Armbrust, Michael, "A review of cloud computing," Communications of the ACM 53.4, pp.50-58, 2010.
- [2] Armbrust, Michael, et al. "Above the clouds: A berkeley view of cloud computing.", 2009.
- [3] Zhang, Qi, Lu Cheng, and Raouf Boutaba. "Cloud computing: state-of-the-art and research challenges." Journal of internet services and applications 1.1, pp.7-18, 2010.
- [4] T OGRAPH, B., and Y. RICHARD MORGENS. "Cloud computing." Communications of the ACM 51.7, 2008.
- [5] Maguluri, Siva Theja, R. Srikant, and Lei Ying. "Stochastic models of load balancing and scheduling in cloud computing clusters." INFOCOM, 2012 Proceedings IEEE. IEEE, 2012.
- [6] Rimal, Bhaskar Prasad, Eunmi Choi, and Ian Lumb. "A taxonomy and survey of cloud computing systems." INC, IMS and IDC (2009): 44-51.
- [7] Weiss, Aaron. "Computing in the clouds." Computing 16 (2007).
- [8] Fayoumi, Ayman G. "Performance evaluation of a cloud based load balancer severing Pareto traffic." Journal of Theoretical and Applied Information Technology 32.1 (2011): 28-34.
- [9] Berl, Andreas, et al. "Energy-efficient cloud computing." The computer journal 53.7 (2010): 1045-1051.
- [10] Marston, Sean, et al. "Cloud computing—The business perspective." Decision support systems 51.1 (2011): 176-189.
- [11] Tunc, Cihan, et al., "Value of Service Based Task Scheduling for Cloud Computing Systems", ICCAC, 2016.
- [12] Shen, Qinghua, et al. "Exploiting geo-distributed clouds for a e-health monitoring system with minimum service delay and privacy preservation." IEEE journal of biomedical and health informatics 18.2 (2014): 430-439.

- [13] de Abranches, Marcelo Cerqueira, and Priscila Solis. "An algorithm based on response time and traffic demands to scale containers on a Cloud Computing system." *Network Computing and Applications (NCA)*, 2016 IEEE 15th International Symposium on. IEEE, 2016.
- [14] Assi, Chadi, et al. "Towards scalable traffic management in cloud data centers." *IEEE Transactions on Communications* 62.3 (2014): 1033-1045.
- [15] Brindha, SK Jeya, J. Angela Jennifa Sujana, and T. Revathi. "Agent based bidirectional bidding mechanism for efficient scheduling of real time tasks in cloud computing." *Electrical, Electronics, and Optimization Techniques (ICEEOT)*, International Conference on. IEEE, 2016.
- [16] Doyle, Joseph, Robert Shorten, and Donal O'Mahony. "Stratus: Load balancing the cloud for carbon emissions control." *IEEE Transactions on Cloud Computing* 1.1 (2013): 1-1.
- [17] Patel, Ronak R., et al. "Improved GA using population reduction for load balancing in cloud computing." *Advances in Computing, Communications and Informatics (ICACCI)*, 2016 International Conference on. IEEE, 2016.
- [18] Ran, Chen, Shaowei Wang, and Chonggang Wang. "Balancing backhaul load in heterogeneous cloud radio access networks." *IEEE Wireless Communications* 22.3 (2015): 42-48.
- [19] Rimal, Bhaskar P., and Martin Maier. "Workflow Scheduling in Multi-Tenant Cloud Computing Environments." IEEE, 2015.
- [20] Soniya, J., J. Angela Jennifa Sujana, and T. Revathi. "Dynamic Fault Tolerant Scheduling Mechanism for Real Time Tasks in cloud computing." *Electrical, Electronics, and Optimization Techniques (ICEEOT)*, International Conference on. IEEE, 2016.
- [21] Vascak, Jan, Jakub Hvizdo, and Michal Puheim. "Agent-Based Cloud Computing Systems for Traffic Management." *Intelligent Networking and Collaborative Systems (INCoS)*, 2016 International Conference on. IEEE, 2016.
- [22] SundarRajan, R., V. Vasudevan, and S. Mithya. "Workflow scheduling in cloud computing environment using firefly algorithm." *Electrical, Electronics, and Optimization Techniques (ICEEOT)*, International Conference on. IEEE, 2016.
- [23] Wang, Xiumin, et al. "Traffic-aware task allocation for cooperative execution in mobile cloud computing." *Communications in China (ICCC)*, 2016 IEEE/CIC International Conference on. IEEE, 2016.
- [24] Zhang, Dongbo, Yanfang Shou, and Jianmin Xu. "The modeling of big traffic data processing based on cloud computing." *Intelligent Control and Automation (WCICA)*, 2016 12th World Congress on. IEEE, 2016