Advancing Cloud Network Management with Smart Monitoring Technique

Sudipta Sahana, Rajesh Bose, Debabrata Sarddar

Abstract*: The proliferation of Cloud Computing has opened new and attractive offerings for consumers. Cloud Service Providers promote and market packages of cloud computing services that cater to diverse opportunities and user applications. While this has obvious advantages, there are certain factors that are a cause for concern. Monitoring the underlying infrastructure that supports the entire fabric of cloud computing is an aspect that requires a great deal of attention. The aspect of monitoring takes on a great deal of significance when performance and robustness of cloud service on offer is taken into consideration. Although research has been conducted into various cloud computing monitoring techniques, there is scope and room yet for an integrated cloud monitoring solution that can fulfill the requirements of cloud administrators to ensure optimal performance of the underlying infrastructure of a cloud computing network. In this paper, we propose a unified monitoring model that is essentially a composite framework involving hardware and network layers. Studies conducted during our experiments suggest that our unified cloud monitoring approach can significantly aid in reducing overall carbon emissions while helping meeting compliance and audit norms by ensuring that the underlying cloud infrastructure is monitored closely.

***Keywords* : Cloud Network, Smart Monitoring , Traffic, Server utilization, Cloud Storage**

I. INTRODUCTION

The National Institute of Standards and Technology defines Cloud Computing thus, "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models." Given the current level of proliferation and marketability of Cloud Computing, this definition presents a holistic definition of how we view and consume cloud computing services. It is the very nature of Cloud Computing that requires detailed examination and continuous monitoring of the underlying infrastructure that supports all functions and offers the features that makes this technology so attractive and ubiquitous. The concept of computing on a cloud platform relies primarily on availability of Internet and the underlying network infrastructure, as well as on hardware systems and platforms that are most supported by virtual machines. The ability to virtualize resources and physical hardware infrastructure has given Cloud Computing a unique advantage of optimizing resources and restraining costs. Given the almost instantenous and on-demand nature of Cloud Computing to provide resource provisioning for users, the reliability and manageability of the physical hardware infrastructure and setup that are interconnected takes on a more prominent role. The broad spectrum of Cloud Computing users encompasses businesses, scientific and academic institutions, software developers, and common users. The flexibility and the ability to bring in tremendous computing and data storage power at the press of a few buttons makes Cloud Computing attractive and commercially lucrative than traditional on-premise deployments.

While consumers eagerly flock to Cloud Computing in droves and continue to do so in greater numbers progressively citing superior response times, excellent performance, superior data handling and backup facilities, for administrators this has become a major challenge. For Cloud Service Providers (CSPs), seeking an optimum balance of available resources and users' demands is a top priority. Without question the issue of putting in place an agile system that is sensitive to changes, can translate to profitability or lack of it for any CSP. In an era where cloud computing commands a great deal of interest, its detractors can only point out to any identifiable systemic faults with its underlying infrastructure consisting of the physical hardware running at multiple data centers located at diverse geographical locations across the globe and connected via the Internet.

Research scholars have conducted studies that show that even a single-point failure at any network node can cascade into multiple faults. While most cloud computing systems today support failover options to route user data and requests across alternate routes, there remains the question of an all-inclusive monitoring system that can help monitor and manage hardware and network at the same time. Studies have already shown that there are grounds for such monitoring systems to exist not only to aid CSP administrators but also help present a clear picture to users.

Financial implications and cost factors concerning Cloud Computing demand that CSPs and users have the best of both worlds. For CSPs, maximizing profitability can be achieved in a way by lowering costs. From the point of view of users, response times and availability of processing power hold the key in persuading them to move from traditional on-premise deployments.

Revised Manuscript Received on January 15, 2020

Sudipta Sahana, Department of CSE, JIS College of Engineering, Kalyani, Nadia, W.B. India 741235 Email: ss.jisce@gmail.com

Rajesh Bose, 2Simplex Infrastructures Ltd, Kolkata, W.B. India Email: bose.raj00028@gmail.com

Debabrata Sarddar, Department of CSE, University of Kalyani, Nadia, W.B. India 741235 Email: dsarddar1@gmail.com

Although many still hold the view that traditional on-premise deployments offer faster responses and more computing power, that argument is usually offset by supporters of cloud computing in the form of reduced operating expenses and almost zero capital investments. However, if CSP administrators can monitor and manage hardware resources as well as network infrastructure that support cloud computing services offered to users across diverse geographical locations, the case of Cloud Computing as the more favoured platform can only grow from strength to strength.

Consumers of cloud computing services and CSPs themselves are bound by Service Level Agreements (SLA). It is incumbent upon CSPs to provide a certain grade or level of service. Failure to adhere to agreed and set standards influencing services consumed by users, can be a serious cause for concern to CSPs. Resulting delays and violations can be detrimental to CSPs as they may be required to compensate for the losses to clients as per the terms of SLAs. Currently, there are monitoring tools and systems that allow CSP administrators to check on hardware systems and networks. Our research into such systems shows that there is scope for a holistic approach to this problem. In the course of our work conducted using existing models by research scholars working in this field, we have been able to design and develop a layered approach as a potential solution to the problem faced by CSP administrators. Our proposed model is composed of two layers and is designed as a unified cloud monitoring solution. In this paper, we show how we have introduced a single point for administrative focus to enable CSP administrators obtain a quick overview of situation unfolding in realtime. With the help of our model, CSP administrators shall also be able to make informed judgements on managing the underlying hardware and network infrastructure of cloud computing services that are being consumed. As a result of our proposed model's agile and robust management system, our experiments reveal that there is considerable scope for improvement insofar as carbon emissions and green cloud computing are concerned.

Our paper is organized in four sections: Section 2 that is titled "Related Work" discusses relevant works conducted by researchers in this area; Section 3 contains details of our proposed technique, and is titled "Proposed Work"; in Section 4, we present our result and analysis under the title "Result Analysis"; and finally, we conclude our study under the section titled "Conclusion" submitted in Section 5.

II. RELATED WORKS

The significance of Cloud Computing is bolstered by two vital considerations. The first being economic and the other technical. The authors [1] in their research have highlighted thee need to lower total cost of ownership [TCO] and increased suppleness and responsiveness in terms of the quality of service offered. With Cloud Computing, balancing economic factors with technical aspects poses a significant challenge as there are myriad of actual physical resources and hardware to be monitored. Monitoring a cloud computing infrastructure is, therefore, an exercise that must be undertaken continuously with focus on physical resources consisting of CPU, memory, storage space, and networking availability combined with software performance and fine-tuning. The authors have observed that the initial steps to monitoring Cloud Computing infrastructure needs to consider the encapsulation of heterogeneity of actual underlying physical hardware through use of virtualization. Thus, a capable monitoring system should be designed around virtualization and recording of the state of the virtualized resources at periodic and frequent intervals. Considering the elastic and scalable nature of Cloud Computing, the authors observe that any Cloud Monitoring system must be as dynamic as the set of resources that may change based on users' requirements. The authors propose two forms of Cloud Monitoring at high- and low-level. At the high-level, monitoring can take place through middleware and applications provided by CSPs and consumers through third-party software. In Cloud Computing, software-as-a-service (SaaS) is one of popular commercial deployment models. The factor of quality of service, or QoS, takes on a greater significance for consumers and, thus, high-level monitoring is a preferred route. However, for CSPs, performance of underlying hardware and network infrastructure is more of a concern. Therefore, low-level Cloud Computing monitoring is a necessity as a poorly performing physical infrastructure is bound to affect the abstraction layer of virtualized resources at which Cloud Computing is shaped and its services consumed.

Cloud Computing is an aggregation of actual physical computing hardware combined with network resources. Through virtualization such hundreds or even thousands of interconnected nodes are used to form a virtualized layer that eventually lead to high scalability and superior performance. Given these numbers, it is not unusual for intermittent failures and rapid degradation of performance because of such failures. The authors [2] discuss the need for a state monitoring framework that discusses the quality of reporting of Cloud Computing monitoring systems. In their work, the authors show how transmission of information on node failures can be a bottleneck to appropriate functioning of Cloud Computing monitoring system and the resultant relevance of information produced based on message dynamics as described in their research. Their paper describes the use of defining a critical state that a monitored system can reach. The authors have proposed a model that periodically checks the state of a monitored system and whether it has reached a critical level at any given point of time. The efficacy of such Cloud state monitoring scheme has been shown to yield results during sudden load peaks on resources by auto-scaling web applications as part of experiments conducted.

The concept of offering monitoring-as-a-service (MaaS) on Cloud has been examined and discussed for the benefit of CSP administrators and consumers alike[3]. The model proposed by the researchers studies the correlation between single- and multi-tenant service. As in research shown concerning state monitoring[2], the authors discuss the applicability of MaaS to monitor state of application, hardware, or network. The authors show that it is possible to design an effective model through resource-aware topology planning. In this design, the efficacy of cloud monitoring is increased through local search based approach to enable service scalability.

Research has shown that there are significant benefits in terms of clustering virtual machines that exhibit similar behavior patterns insofar as resource utilization and consumption are concerned. While there have been several methods in use to cluster virtual machines, such mechanism does not reach the required reactive levels in situations where virtual machines are required to be added or removed based on dynamic nature of Cloud Computing dynamics involving consumers' needs. In a research[4], the authors propose an adaptive approach to clustering virtual machines based on time series and degrees of uncertainty that may result from clustering processes.

The applicability of virtual network functions (VFNs) has a great deal of significance in the context of geographically distributed Cloud Computing systems and infrastructure. Researchers[5] have shown that reduction of capital and operating expenditures on networks for Cloud Computing can be overcome using their proposed approach. The model proposed in their paper uses an on-demand monitoring system across multiple domains consisting of data center and network support systems that consist radio access and mobile core networks.

In their research[6], the authors discuss how their proposed model involving Weight-Based Optimizing Routing, can significantly reduce consumption of power by migrating processor-intensive tasks to cloud. Their work which takes on added significance in the context of green cloud computing focuses on two subordinate features consisting of mobile and cloud networks. In their model, the authors introduced a weight-based approach for their algorithm to select optimum node availability based on path cost and channel availability. Their work complements research[7] conducted on load balancing on cloud networks using hierarchical search optimization. The work as presented in that paper is unique for two reasons. First, being identification of packets transmitted on underlying Cloud Computing networks and sifting through for instances of packet flooding. Second, a hierarchical structure is proposed to identify and retrieve for any particular resource on Cloud.

The question asked of high-performance computing (HPC) in Cloud Computing environment is its continued sustainability while keeping a low carbon footprint. Given the state of environmental pollution, public awareness and economic concerns mandate that performance metrics be revised and rescaled. At this juncture, therefore, presence of appropriate Cloud Computing monitoring system takes on an increased significance. In research studies[8][9], the authors discuss and propose methods to streamline power consumption using algorithms that balance server loads in Cloud Computing data centers. Incidentally, the aspect of data retrieval also begins to take on a greater significance when taken in context of many millions of users who rely on Cloud Computing to store data. Enterprises and organizations prefer to go the Cloud Computing route because of convenience and ease in not only storing data, but also retrieval of such at a later date[10].

As noted and documented in research conducted[11], for Cloud Computing data centers to conserve power at every possible stage, power usage and heat sensitivity issues need to be addressed. Research studies[8][9] concluded that rationalizing power consumption and, thus, bring in degree of control over heat emissions resulting from peak loads on servers, can be possible using proposed load balancing techniques. In tandem with how server load balancing optimization is approached, considering that Cloud Computing networks has to handle search requests from system applications, software, or direct consumer requests, search engine optimization is considered extremely important. The authors from their research studies[12][13] concluded that it is possible to enhance search engine optimization through a novel route table algorithm and fog computing. The concept of fog computing allows a significant portion of computing processes to be conducted by devices at the edge of the network. Thus, much of the computing loads that would have otherwise sent Cloud Computing data centers to work at peak performance for extended lengths of time, can be systematically processed. The resulting energy efficiency level achieved helps economize available resources and, consequently, increases equipment life.

Arguably, gossip protocol has caught the undivided attention of researchers in Cloud Computing. The advantages that this procedure offer can be said to be significant. Researchers[14][15] have shown that using gossip protocol, allows for a more efficient and speedy transmission of data from one node to another. Given that Cloud Computing infrastructure can involve many nodes numbering in hundreds or even more, savings in terms of energy and increase in performance is multiplied. The results and analysis of researchers consolidate findings from contemporary studies conducted in this field, and indicate that gossip protocol plays a significant role in enhancing Cloud Computing performance at both hardware and network levels.

III. PROPOSED METHODOLOGY

In our research, we focus primarily on two layers of cloud computing. The first being at the Data Center layer, and the other being at the Network layer. Of vital interest to our research is Gossip protocol that we have adopted to manage and control network layer. In our proposed model, we have used Gossip protocol to switch from a node that is currently down in the network to another working node without any significant loss or high latency in an ongoing communication process. We have also integrated monitoring parameters to monitor temperature of servers, processor utilization limit, storage problem and identify any significant hardware faults. In an approach that mirrors existing research[2], we set threshold values for different components within the underlying hardware infrastructure of a Cloud Computing network. Any breach in threshold values will automatically set off a trigger to control any runaway situation that might cause a slowdown in the system. We also introduce Billboard Manager[13] as a control terminal for Cloud Computing system administrators.

At the Data Center layer, we look at identifying and managing hardware aspects including temperature control. The latter being of vital interest considering that power utilization factors and optimum load balancing techniques are at stake, our proposed model uses the following algorithm at this layer:

Algorithm for Temperature Control

Since each server has a limit of heat tolerance, we set a threshold value of temperature sensors installed at 45^0C. A trigger is automatically activated when the threshold limit is breached. The algorithm has been designed to send the server to an idle state immediately when a breach is detected.

The server is only called into action or kept at standby mode only when its temperature is at the level of ambient temperature within Data Center. The algorithm also prevents more than one server to make a break for idle state so as to prevent Data Center operations to slow down.

Server temperature crosses the threshold limit.

Step 1: Information is sent to the supervisor.
Step 2: Supervisor checks the flag value.
Step 3: If flag is not set, continue from step 4 else go to step 8
Step 4: Does not allocate a more job to the unhealthy server.
Step 5: Server completes the queued job and goes to idle state.
Step 6: After reaching a cooled state the server is called back to action based on the trigger called and informs the supervisor about the ready status.
Step 7: Request not granted due to one server is already at the same state.

Algorithm for Utilization Limit issue

It is not uncommon for servers to suffer failures resulting out of performance overloads brought about by many simultaneous and constant processing requests. To obviate such a possibility, we set a threshold value of 95%. A CSP administrator monitoring and managing the system will not find the server ready to accept any more processing requests once it breaches this threshold mark. In this section of the algorithm, we introduce Gossip protocol to determine the state and readiness of servers available throughout a Cloud Data Center network.

Step 1: Server utilization reaches the threshold value.
Step 2: Trigger got activated and inform supervisor not to assign any task.
Step 3: Supervisor acts the same until next gossip message received.
Step 4: Server reaches the check point and gossip runs again.
Step 5: Supervisor got informed and assign task normally to the said server.

Storage exceeded and hardware fault issue

In the case of storage space, we set a threshold at 95% of the total available storage. A breach in the threshold mark shall automatically notify a CSP administrator for addition of new disks. The model can be adapted to send a message every hour till action is taken to reduce the storage utilization level from the 95% utilization mark.

Hardware faults can be tracked by a dedicated server connected to probes or specific software. Embedded systems can be used to report on hardware health. Using Gossip protocol, messages can be transmitted and broadcasted to all concerned CSP administrators via Billboard Manger. Once, a server hardware is restored, Billboard Manager can then use Gossip protocol to change the status to readiness for the particular server hardware.

At Network End

Networks are susceptible to three types of failures raning from intermediate node down, heavy traffic and network congestion, to outright link breakdowns and failures. We have made use of Gossip protocol in our design of the algorithm to report on failure as well as restoration of service. The algorithm for monitoring and managing the network layer of operations is as follows:

Algorithm for intermediate node down or Link break down

Step 1: Communication node suffers failure.
Step 2: Message sent to Billboard Manager automatically to set alert on node failure.
Step 3: CSP administrators are made aware of the problem on Billboard Manager.
Step 4: Following restoration, all nodes are notified of working state of the node.
Step 5: Billboard Manager reports on status on network based on Gossip protocol.

Algorithm for Heavy Traffic

Step 1: Network congestion development is detected.
Step 2: Node sends gossip message to the associated nodes for link avoidance due to congestion.
Step 3: Node deals with the queued processes.

Once node is free and is available for communication, message is transmitted accordingly.

IV. RESULT ANALYSIS AND DISCUSSION

Our proposed methodology focuses on managing the two layers of hardware and networking on which the primary foundation of Cloud Computing is constructed. We have used Gossip protocol to respond and cater to the dynamic nature of Cloud Computing and incidences of failures in the underlying infrastructure. Our simulated experiments show that our proposed approach can significantly aid in reducing the rate of carbon emissions. Similarly, the number of concurrent processes can also be increased.

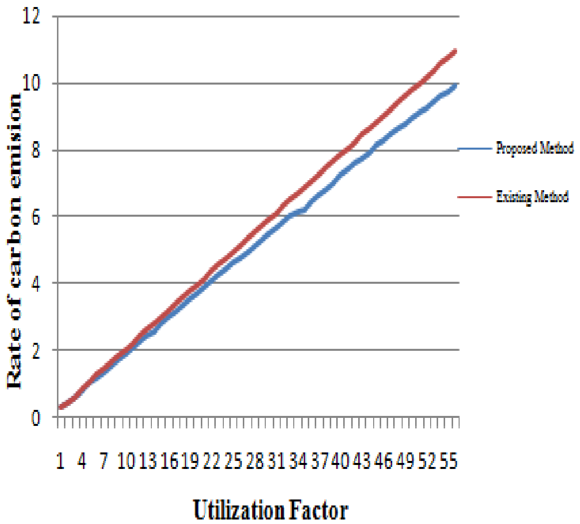
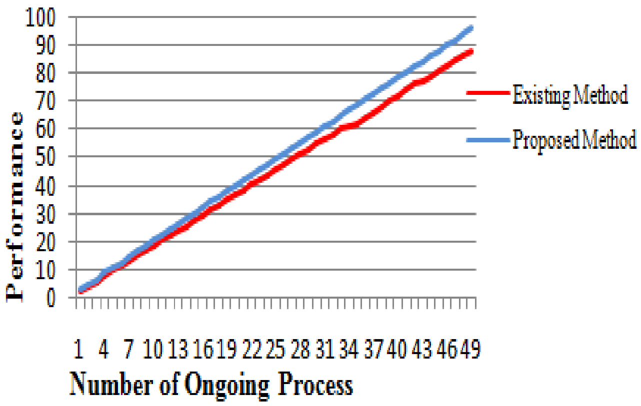


Fig 1. Carbon Emission Rate vs. Utilization Factor



Performance vs. Number of Ongoing Process

V. CONCLUSION

The essence of Cloud Computing is its elasticity and scalability. Given that service level agreements that are enforceable can be a matter of litigation, CSPs today are cognizant of the need to put in place effective monitoring system. While existing research has shown that it is possible to design and implement monitoring systems that are either hardware focused or network-oriented, our proposed model offers to monitor both layers using Gossip protocol and Billboard Manager. The resulting combination promises to offer a steady stream of information on underlying hardware, storage, and network infrastructure to any CSP administrator.

ACKNOWLEDGMENT

The authors are indebted to JIS College of Engineering and Dept. of CSE, University of Kalyani for providing invaluable lab support to conduct the work.

REFERENCES

1. Giuseppe Aceto, Alessio Botta, Walter de Donato, Antonio Pescapè "Cloud monitoring: A survey" in Computer Networks Volume 57, Issue 9, 19 June 2013, Pages 2093-2115 https://doi.org/10.1016/j.comnet.2013.04.001
2. Shicong Meng, Arun K. Iyengar, Kisung Lee, Isabelle M. Rouvellou, Balaji Palanisamy "Reliable State Monitoring in Cloud Datacenters" in IEEE Fifth International Conference on Cloud Computing, 2012 DOI 10.1109/CLOUD.2012.10
3. Shicong Meng, Ling Liu, Senior Member "Enhanced Monitoring-as-a-Service for Effective Cloud Management" in IEEE Transactions on Computers, Vol. 62, No. 9, September 2013
4. Claudia Canali, Riccardo Lancellotti "An Adaptive Technique to Model Virtual Machine Behavior for Scalable Cloud Monitoring" in IEEE Symposium on Computers and Communications (ISCC) 2014 DOI: 10.1109/ISCC.2014.6912613
5. Julius Muelleri, David Palma, Giada Lande, Joao Soares, Bruno Parreira, Thijs Metsch, Peter Gra, Alexander Georgiev, Yahya Al-Hazmil, Thomas Magedanzl, Paulo Simoes "Monitoring as a Service for Cloud Environments" in IEEE international conference 978-1-4799-5051-5/14/$31.00 ©2014 IEEE
6. D. Sarddar, R. Bose, S. Sahana "A Novel Approach on Weight Based Optimized Routing for Mobile Cloud Computing" in Brazilian Journal of Science and Technology DOI:10.1186/s40552-015-0008-xpublished on 27 June 2015
7. D. Sarddar, R. Bose, S. Sahana "An Enhanced Cloud Network Load Balancing Approach Using Hierarchical Search Optimization Technique" in International Journal of Hybrid Information Technology Vol.8, No.3, pp.9-20 ISSN: 1738-9968http://dx.doi.org/10.14257/ijhit.2015.8.3.02(2015)
8. R. Bose, S. Sahana,D. Sarddar "An Energy Efficient Dynamic Schedule based Server Load Balancing Approach for Cloud Data Center" in International Journal of Future Generation Communication and NetworkingISSN: 2233-7857 Vol. 8, No. 3, pp. 123-136http://dx.doi.org/10.14257/ijfgcn.2015.8.3.12 (2015)
9. S. Sahana, R. Bose, D Sarddar "Server Utilization Based Smart Temperature Monitoring System for Cloud Data Center" in International Conference of International Conference on Industry Interactive Innovations in Science,
10. Sudipta Sahana, Rajesh Bose and Debabrata Sarddar "Harnessing RAID mechanism for enhancement of Data Storage and Security on Cloud" in Brazilian Journal of Science and Technology (2016) 3:12 DOI 10.1186/s40552-016-0024-5 published on 22nd March 2016.
11. Muhammad Tayyab Chaudhry, Teck Chaw Ling, Atif Manzoor, Syed Asad Hussain, Jongwon Kim "Thermal-Aware Scheduling in Green Data Centers" in ACM Computing Surveys · February 2015 DOI: 10.1145/2678278
12. R. Bose, S. Sahana, D. Sarddar "An Enhanced Storage Management Scheme with Search Optimization for Cloud Data Center" International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 12 (2015) pp. 32141-32150
13. Sudipta Sahana, Rajesh Bose and Debabrata Sarddar "Fog-Based Hierarchical Search Optimization" in CSI 2017, CCIS 836, pp. 674–681, 2018. https://doi.org/10.1007/978-981-13-1343-1_55 (2017)
14. Fetahi Wuhib, Mads Dam, Rolf Stadler "A Gossiping Protocol for Detecting Global Threshold Crossings" in IEEE Transactions on Network And Service Management, Vol. 7, No. 1, March 2010
15. S. Sahana, R. Bose, D Sarddar "An Enhanced Search Optimization Proposal Based On Gossip Protocol For The Cloud" in International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 19 (2017) pp. 8436-8442

AUTHORS PROFILE

Sudipta Sahana, is an assistant professor of a renowned engineering college of West Bengal. For more than 4 years, he has worked in this region. He has passed his M.Tech degree in Software Engineering and B.Tech Degree in Information Technology from West Bengal University of Technology with a great CGPA/DGPA on 2010 and 2012 respectively. He is recently working in Ph.D. in the domain of "Cloud Computing". He is a member of the Computer Science Teachers Association (CSTA), Computer Society of India (CSI) and also a member of International Association of Computer Science and Information Technology (IACSIT). He has published many research papers in this domain.

Dr. Rajesh Bose, is an IT professional employed as Deputy Manager (IT) with Simplex Infrastructures Limited, Data Center, Kolkata. He graduated with a B.E. in Computer Science and Engineering from Biju Patnaik University of Technology (BPUT), Rourkela, Orissa, India in 2004. He went on to complete his degree in M.Tech. in mobile communication and networking from Maulana Abul Kalam Azad University of Technology, West Bengal (Formerly known as West Bengal University of Technology) - WBUT, India in 2007. He completed his PhD degree from the University of Kalyani in Computer Science and Engineering in 2018. He has also several global certifications under his belt. These are CCNA, CCNP-BCRAN, and CCA (Citrix Certified Administrator for Citrix Access Gateway 9 Enterprise Edition), CCA (Citrix Certified Administrator for Citrix Xen App 5 for Windows Server 2008). His research interests include cloud computing, wireless communication and networking. He has published more than 75 technical papers in different journals and conferences that have been cited worldwide. Mr. Bose also has experience in teaching at college level for a number of years prior to moving on to becoming a full time IT professional at a prominent construction company in India administering virtual platforms and systems at the company's data center.

Debabrata Sarddar, Assistant Professor in the Department of Computer Science and Engineering, University of Kalyani, Kalyani, Nadia, West Bengal, INDIA. He has done Ph.D. at Jadavpur University. He completed his M. Tech in Computer Science & Engineering from DAVV, Indore in 2006, and his B.E in Computer Science & Engineering from NIT, Durgapur in 2001. He has published more than 75 research papers in different journals and conferences. His research interest includes wireless and mobile system and WSN, Cloud computing.