ISSN:0975-9646

Deepak K Gaur et al/ (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 12 (2) , 2021, 42-47

# Effective Fault prediction using classifier analysis for cloud environment

Deepak K Gaur[#1], Ajitabh Mahalkari[*2]

[#] *Department of Computer Science and Engineering*

*Sushila Devi Bansal College of Technology, Indore ( M.P)-45333, India*

[1]deepakme153@gmail.com

[2]arjun.parihar@sdbct.ac.in

***Abstract*— Cloud computing had opened a new horizon for utilization of resources and their computing. Emerging model of cloud computing to manage distributed large scale infrastructure resources is becoming very popular. Entities in both industry and individuals are investing a huge time and effort to investigate and develop its features. Fault tolerance is the ability to handle the faults and remove failures in such a way that the system will perform its functions unaware that any methods have been taken to handle the faults.**

**Cloud Computing promises real costs effective and agility to organizations. Cloud computing serves the demands of a number of individuals and organizations. The cloud environment consists of virtual machines. Virtual machines utilize the infrastructure capacity of resources. The performance of virtual machines monitors can be monitored in terms of response time, throughput, and bandwidth.**

**The proposed model utilizes these performance statistics to detect the faults after that fault is recognized by data analytics and machine learning algorithms. Fault prediction can be helpful to treat faults well in advance and appropriate action can be taken to make the system more reliable. We also developed a prototypes for data collection and training using .net and python libraries. Statistical results found satidfactory at the intial stages of research and experiments**

***Keywords*— Cloud Computing, Fault Tolerance, prediction, statistics, analysis.***

## I. INTRODUCTION

Cloud Computing is getting popular day by day, and also proved to be useful to organization, due to this there is a tremendous need of fault tolerance in cloud computing. Fault tolerance is group of actions and events that require avoiding the impact of the fault occurred in the system. Fault tolerance includes techniques to make system robust enough to handle the faults

Key motivation of using fault tolerance in cloud computing may includes failure recovery, lower costs, and improvised performance. Cloud computing has become first choice to design, develop and deploy distributed applications because of the features like virtualization, flexibility and elasticity. Quality of service is been managed by cloud computing in view of dependability and availability.

As it is clear that fault tolerance is one of the key business requirements of cloud computing, is the reason of motivation to involve more researchers to develop efficient fault tolerance strategies. Fault tolerance is an important key issue in cloud computing that why amount of research is being carried out on fault tolerance in cloud computing

Fault tolerance basically concerns with the policies and techniques to avoid faults in the system so that business operations do not face the down time. Client need continuity and availability, which can interrupted due to fault in the system.

Faults in the cloud computing may include hardware faults, software faults, network faults, process faults etc. Fault tolerance policies may include reactive fault tolerance, proactive fault tolerance, adaptive fault tolerance etc.

It's really challenging for CSPs to maintain high availability in cloud, so that customers can use the services of cloud uninterrupted. Information or data losses because of faults in the system may result in big loss in enterprises and may also lose the trust in the system. Faults can vanish out the availability of the data and services, so important to handle faults in the system properly.

Need of better and efficient fault tolerance model in terms of accuracy, detection and response time. If faults are not properly handled, it may cause serious damage to the system, and makes the system unreliable. Early detection of faults makes system capable to handle the faults before it hampers the system functions and can be well treated in time.

## II. LITERATURE SURVEY

In this paper [18], by Preeti Kumari et al. emphasizes on the importance of fault tolerance in cloud computing. Authors presented a comprehensive review of fault tolerance in clou computing. They presented a thorow review of fault tolerance related issues and chalanges need to be addressed in fault tolerance. Paper gives a state- of-the art review of fault tolerance in cloud computing.

In this paper [20], by Moin Hasan et al. , authors mentioned that fault tolerance ia very importance and critical factor to ensure reliability of the system. Implementation of fault tolerance policy in cloud not only needs specific knowledge of its application domain , but a comprehensive analysis of backgroung and various prevalent techniques also. This work [20] gives a comprehensive and systematice analysis of the various fault tolerance approaches. Authors provided intense survey of various fault tolerance solutions available in cloud computing. The best part in the work is, they also presented a qualified view of the available solutions.

In this paper [21], by Chiang et al. , they emphasizes on the reliability of cloud computing services as a fault tolerance in concerned. Authors addressed the issue in Fault Diagnosis Agreement (FDA). FDA is uses to establish a fault free service connection to the consumers. The work[21] also uses the concept of Early Stopping Protocol (ESP). They also proposed a Early Diagnosis Cloud Agreement (EDCA) to address the issues of Byzetien Agreement [BA] and FDA. The main focus of the work is to remove faulty servers, so to provide unintrupted services.

In this work [22], by Han et al., authors mentioned the importance of fault tolerance in reliability cloud systems.Traditional fault tolerance mechanisms cannot full fill the ever changing demand of consumers of cloud computing. In this work[22], authors proposed ARCHER as, fault tolerance scheduling algorithm to address the issues of fault tolerance. Proposed solution, not only deal with the traditional methods like backup and checkpointing which leads to better resource utilization. But also apply scheduling to VM tasks based on the conditions. Proposed algorithm proven to be usefull through experimental results.

In this work [24], by Bhupesh Kr Dewangan et al., focuses on the energy consumption of the cloud resources. The fault tolerance mechanism add on the resources to avoid faults that increases the overall energy comsuption of the cloud systesm, at the same time faults need to be handled and resources need to be optimized for usage and energy consumption. In this work[24], authors addressed both key concepts thats is Fault tolerance and resource utilization in cloud computing. They proposed method that take care of the faulty resources in cloud infrastructure to avoid fault conditions. They utilizes the Antilion Optimization Algorithm to find the optimal resource with higher quality of services and non violation of service level agreements(SLAs).

In this work[19], by Bashir Mohammed et al., They considered failure preditions as very critical issue in cloud computing. In IaaS its even more challenging to predict the faults in advance and handle them appropriately. They mentioned that, as per the needs of modern computing systems, a strong and robust fault tolerance technique capable to predict the faults in advance to avoid potential failures in the sytem. In this work[19], they proposed a fault prediction model for high performance cloud infrastructure. They used the concept of Auto-Regressive Moving Average(ARMA) to predict the faults conditions in future. The solution was able to give 95% of accuracy and also is practically adaptable for real time systems.

## III. Problem Identification

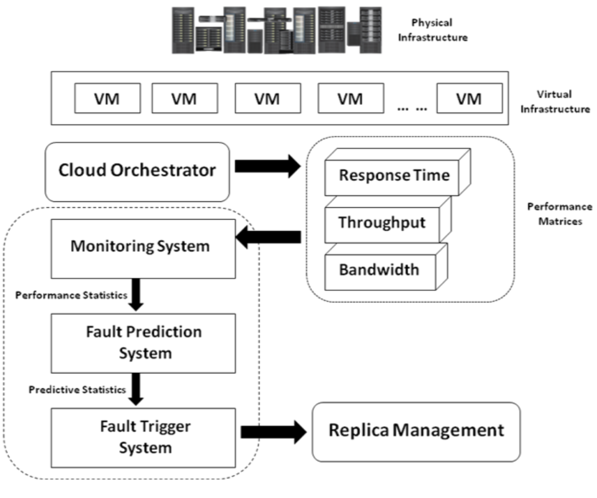
After review of literature available in the field of fault tolerance in cloud computing, we have identified some problems and issues that need to be addressed. Following are the problems identified,

- It’s really challenging for CSPs to maintain high availability in cloud, so that customers can use the services of cloud uninterrupted.
- Information or data losses because of faults in the system may result in big loss in enterprises and may also lose the trust in the system.
- Faults can vanish out the availability of the data and services, so important to handle faults in the system properly.
- Need of better and efficient fault tolerance model in terms of accuracy, detection and response time.
- If faults are not properly handled, it may cause serious damage to the system, and makes the system unreliable.
- Early detection of faults makes system capable to handle the faults before it hampers the system functions and can be well treated in time.

## IV. Proposed Solution

Different fault solutions are analyzed and compared and proposed a new model for early prediction of faults using data analytics and machine learning algorithms. The cloud environment consists of virtual machines. Virtual machines utilize the infrastructure capacity of resources. The performance of virtual machines monitors can be monitored in terms of response time, throughput, and bandwidth. The proposed model utilizes these performance statistics to detect the faults after that fault is recognized by data analytics and machine learning algorithms. Detected fault can be treated well in advance and appropriate action can be taken to make the system more reliable.

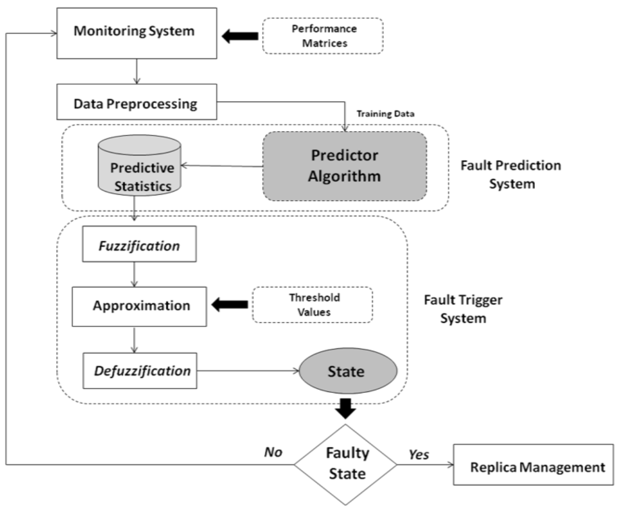
The proposed model consists of three major components in bracket subsystems. The monitoring system monitors the performance parameters of virtual machines; these parameters are transferred to the fault prediction subsystem. The fault prediction subsystem comma predictor algorithm uses these statistics to create predictive statistics. Thought component is fault triggering subsystem. In this, projective statics undergo analysis. Some threshold values need to be provided so that fault can be recognized based on the threshold values. If fault detected is considered to be a faulty state, appropriate handling subroutine need to be involved for example replica management.




Virtual machines possess various utilizations and capacities of computing resources such as CPU, RAM, network data transmission, and storage volume. To establish the suitable tolerance scheme, the proposed engineering is focused to early distinguish the failures of the cloud ahead of time. To accomplish this objective, the prediction technique is used to give information on the predictive monitoring statistics. From that point onward, the faults are perceived by using the data analytics and machine learning algorithms. Basically, the organization of the proposed model is as follows:

- **Monitoring System:**

This part of the proposed model collects the performance matrices as training data from virtual machines. The primary source of the statistics is cloud orchestrator, its manager of all the virtual resources available. Monitoring system continue sly monitors the allocation and de-allocation of the resources .Monitoring system collects the matrices and provides these statistics to predictor algorithm. Parameters provided by monitoring system treated as training data for predictor algorithms.




- **Fault Prediction System:**

The fundamental task of monitoring component is to gather system parameters as the information dataset for the prediction process. Periodically, the monitoring system provides the info statistics to the fault prediction system.

The prediction system, which is based on the Gaussian process regression(GPR), is responsible to deliver the predictive statistics of the following monitoring epoch. This futuristic utilization information is subsequently used in the fault trigger component. • The fault trigger, as its name, does the fault detection based on the data analytics and machine learning algorithms. At whatever point a failure is distinguished, the result is sent to the fault tolerance system to recover ahead of time. Because the suitable actions would be applied in the beginning phase of failure, the consequence of potential errors is assumed control over much better.

The working of the prediction depends on the monitoring statistics, which reflect the historical utilization of virtual resources. Assuming that the input data are a limited collection of time location x = [x1, x2, x3, . . . xn], a finite set of random variable y = [y1, y2, y3, . . . yn] represents the corresponding joint Gaussian distribution of historical monitoring statistics. Theoretically, the statistics consist of the physical resources utilization (CPU usage, RAM usage, disk I/O) and the network metrics (response time, bandwidth, throughput, request/response ratio) with regard to the time order.

For Example, five linguistic terms are used to describe the value of response time: very fast, fast, normal, slow, very slow. The value of parameter variables, explains how the determination process is conducted. Depending on the different factors, the value of parameters may vary.

- **Fault Trigger System**

This component is used to recognize the failures in the cloud computing system based on data analytics and machine learning algorithms and predictive statistics. Naturally, the predictive statistics give the futuristic awareness of the system status. With this capacity, the fault tolerance would bode well and more proper to cover the likely issues.

On the other hand, the alluring advantage of data analytics and machine learning algorithms is the approximate reasoning with an imprecise proposition based on the data set in the proposed approach, the aforementioned predictive information is used as the input data to early distinguish the likely issue. So as to do that, the first step of establishing the fault trigger component is to distinguish the desired metrics. Based on the studies of system performance and network measurement, the most key metrics can be considered as underneath: Response time, Throughput and Bandwidth.

## V. RESULT ANALYSIS

After proposing a solution to underlined problem, we developed a prototype to find the proof of concept. The cloud environment consists of virtual machines. Virtual machines utilize the infrastructure capacity of resources. The performance of virtual machines monitors can be monitored in terms of response time, throughput, and bandwidth. The proposed model utilizes these performance statistics to detect the faults after that fault is recognized by Fuzzy Logic algorithms. Detected fault can be treated well in advance and appropriate action can be taken to make the system more reliable.

The proposed model consists of three major components in bracket subsystems. The monitoring system monitors the performance parameters of virtual machines; these parameters are transferred to the fault prediction subsystem. The fault prediction subsystem comma predictor algorithm uses these statistics to create predictive statistics. Thought component is fault triggering subsystem. In this, projective statics undergo fuzzification. Some threshold values need to be provided so that fault can be recognized based on the threshold values. If fault detected is considered to be a faulty state, appropriate handling subroutine need to be involved for example replica management.

We run the prototype for three hundred different cases to develop a data set, which will be used to train our model for accuracy in fault prediction. Table 1 shows the different performance parameters collected corresponding to different cases.

| Case | CPU Utilization (%) | RAM(MB) | Performance (Response) | File size (Bytes) | Upload Speed(kbps) | Download Speed(kbps) |
|---|---|---|---|---|---|---|
| 1 | 8 | 18 | 36 | 46096 | 31.63 | 2.39 |
| 2 | 3 | 16 | 36 | 46590 | 30.39 | 1.79 |
| 3 | 4 | 15 | 47 | 21785 | 30.14 | 5.67 |
| 4 | 13 | 18 | 29 | 38106 | 34.48 | 3.25 |
| 5 | 8 | 34 | 43 | 34245 | 34.69 | 5.98 |
| 6 | 11 | 12 | 28 | 40688 | 34.77 | 8.54 |
| 7 | 8 | 29 | 44 | 20964 | 36.3 | 2.21 |
| 8 | 9 | 43 | 31 | 10743 | 39.06 | 7.79 |
| 9 | 8 | 14 | 43 | 26113 | 34.6 | 2.76 |
| 10 | 8 | 45 | 46 | 38787 | 32.19 | 0.02 |
| 296 | 8 | 13 | 41 | 16290 | 36.83 | 6.89 |
| 297 | 2 | 38 | 35 | 19772 | 32.21 | 1.98 |
| 298 | 7 | 28 | 39 | 42691 | 37.15 | 6.3 |
| 299 | 1 | 16 | 46 | 11696 | 33.32 | 3.46 |
| 300 | 6 | 20 | 41 | 17791 | 30.82 | 6.15 |

**TABLE 1.1**: SHOWING DIFFERENT CASES (300) SCENARIOS WITH DIFFERENT SYSTEM ATTRIBUTES. ACCORDING TO THAT WE TOOK DECISION OF FAULT OCCURRENCE.

| | CPU | RAM | PF | FS | Pkt | upspeed | dwspeed | Fault |
|---|---|---|---|---|---|---|---|---|
| **count** | 303 | 303 | 303 | 303 | 303 | 303 | 303 | 303 |
| **mean** | 7.59736 | 27.76568 | 37.22112 | 30426.88 | 20.84819 | 35.12518 | 5.079472 | 0.537954 |
| **std** | 4.403069 | 10.03411 | 7.33814 | 12089.21 | 8.270473 | 2.787091 | 2.773688 | 0.499382 |
| **min** | 1 | 12 | 25 | 10196 | 7 | 30.08 | 0.02 | 0 |
| **25%** | 3 | 18 | 31 | 19005.5 | 13 | 32.72 | 2.715 | 0 |
| **50%** | 7 | 27 | 37 | 31530 | 22 | 35.29 | 5.31 | 1 |
| **75%** | 11 | 36.5 | 43 | 41254 | 28 | 37.44 | 7.43 | 1 |
| **max** | 15 | 45 | 50 | 49874 | 34 | 39.97 | 9.99 | 1 |

**TABLE 1.2**: SHOWING DIFFERENT CASES (300) SCENARIOS WITH DIFFERENT SYSTEM ATTRIBUTES. ACCORDING TO THAT WE TOOK DECISION OF FAULT OCCURRENCE.

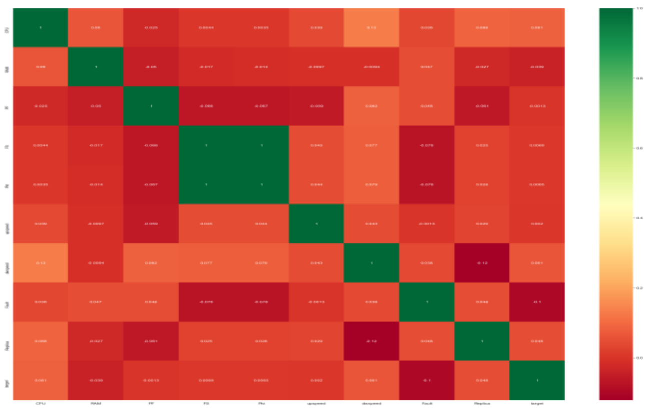
- **Data Analysis**

Data set is prepared through the different performance parameters collected after executing the prototype for different cases. The data is not yet ready to be used in the model training and predictions of the system faults. Data set need to be Pre-processed so that any anomalies can be removed, which can potentially affect the fault predictions. Although data set contains finite and unstructured data

- Heat Map Analysis

The proposed solution is intended for forecasting the faults in the system. Heat map analysis shows the possibilities of the system faults, based on the learning and classification done on the data set collected. As the data set is made up of the parameters collected from the system performance it can give collect insights of the system status and utilizations.

In heat map analysis we have used different color codes to show the possibilities of the faults in the system based on the dataset. Here, we have color 'green' showing there is not at all possibilities of the faults, color 'orange' will tell us about the smooth running status of the system. Color 'red' is notion of faults here; whatever is the combination of the parameters collected the fault in the system may occur in future also. This analysis shows the clear states that, if the values of the parameters is similar to the values or also near by the system need to be at alert and try be proactive against the system faults.



| | CPU | FS | Pkt | Fault | target |
|---|---|---|---|---|---|
| **0** | 0.091597 | 1.298268 | 1.35062 | 1 | 0.914529 |
| **1** | -1.045854 | 1.339198 | 1.35062 | 1 | 0.914529 |
| **2** | -0.818364 | -0.716025 | -0.708286 | 0 | 0.914529 |
| **3** | 1.229047 | 0.636255 | 0.623947 | 0 | 0.914529 |
| **4** | 0.091597 | 0.316351 | 0.260611 | 0 | 0.914529 |

**TABLE 3**: SHOWING DIFFERENT CASES (300) SCENARIOS WITH DIFFERENT SYSTEM ATTRIBUTES. ACCORDING TO THAT WE TOOK DECISION OF FAULT OCCURRENCE.

- **Classifier Analysis**

Classifier here, is showing the statistical analysis of the cases we collected to form the dataset. The analysis of the data set shows the parameters and its finite values that can be used for fault analysis. These values are very useful to predict the future faults. These values will indicate the possibilities of the future system failures so that the system administrator can plan the nessaccasry actions to save the system. This classification analysis is very important in the sense that, system can be safeguard from the downtime and also reduce the system damages because of the faults.

Table 3 showing the classifier analysis of the data collected from the current state of the system in the form of different parameters.

## VI. Conclusions and Future Work

Its complex to predict and manage faults proactively in the cloud environment because of its volatile nature. At the same time organizations expect robust policies against the faults in the system so that they can reliably deploy there business applications and data on the third party cloud service provider. Prediction of the faulty conditions will play a key role here that if the cloud service providers can nearly forecast the future faults. Knowing the approximate conditions in which the possibilities of the system failure will help the providers to prepare the system against the faults.

This action will lead to increase in the reliability of the system. Our proposed solution tried to build with keeping this in mind that in cloud environment different parameters like CPU, RAM usage, and network states may become the indicatives of the performance of the system. These parameters and its values can be used to find the future faults.

We have developed prototypes of the proposed solution in .net and python; we have collected data and form a data set. This training data set is used to predict the future faults occurrences. The research is still in the initial stages, but statistical results found satisfactory.

This domain of fault prediction and analysis has a potential scope of research. We have initiated the formulation of the solution for the prediction of faults in the system proactively and manage it on time in the cloud environment. We have used data analytics and machine learning algorithms for better analysis and prediction.

In the future, we can even extend the work to better and accurate predictions using the optimum algorithms for the data analysis. In future we may think of the larger dataset, because larger data set will be helpful for better training of model and lead to better predictions.

## References

[1] Peter Mell, Timothy Grance, ”The NIST Definition of Cloud Computing”, *Recommendations of the National Institute of Standards and Technology*, Special Publication 800-145, 2010

[2] Yashpalsinh Jadeja, Kirit Modi, “Cloud Computing - Concepts, Architecture and Challenges”, *International Conference on Computing, Electronics and Electrical Technologies [ICCEET]* 2012

[3] Jayadivya S K, Jaya Nirmala S. “Fault tolerant workflow scheduling based on replication and resubmission of tasks in Cloud Computing”, *International Journal on Computer Science and Engineering (IJCSE)* 2012

[4] Anju Bala, Inderveer Chana, “Fault Tolerance- Challenges, Techniques and Implementation in Cloud Computing” *IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No 1*, January 2012

[5] Deepak Poola, Kotagiri Ramamohanarao, and Rajkumar Buyya,”Fault-Tolerant Workflow Scheduling Using Spot Instances on Clouds” *ICCS 2014. 14th International Conference on Computational Science Volume 29, 2014, Pages 523–533*

[6] Ajitabh Mahalkari, Ritu Tondon “A Replica Distribution Based Fault Tolerance Management For Cloud Computing” *(IJCSIT) International Journal of Computer Science and Information Technologies*, Vol. 5 (5), 2014, 6880-6887

[7] Deepak Poola, Kotagiri Ramamohanarao, and Rajkumar Buyya,”Fault-Tolerant Workflow Scheduling Using Spot Instances on Clouds” *ICCS 2014. 14th International Conference on Computational Science Volume 29, Pages 523–533,* 2014

[8] R. Raju,J. Amudhavel, M. Pavithara et al. “A Heuristic Fault Tolerant MapReduce Framework for Minimizing Makespan in Hybrid Cloud Environment”

[9] Ji Wang, Weidong Bao, et al. “FESTAL: Fault-Tolerant Elastic Scheduling Algorithm for Real-Time Tasks in Virtualized Clouds*”* *IEEE TRANSACTIONS ON COMPUTERS*, 2014 DOI: 10.1109/TC.2014.2366751

[10] Ajitabh Mahalkari, Ritu Tondon “Performance Evaluation Of Quality Metrics Based Fault Prediction model For Replica Management In Cloud Computing” *(IJCSIT) International Journal Of Computer Science And Information Technologies*, Vol. 6 (3) , 2015, 2512-2518.

[11] J.Soniya et al. “Dynamic Fault Tolerant Scheduling Mechanism for Real Time Tasks in Cloud Computing” *International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) –* 2016

[12] Xiaomin Zhu et al. “Fault-Tolerant Scheduling for Real-Time Scientific Workflows with Elastic Resource Provisioning in Virtualized Clouds” *IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS*, 2015 DOI 10.1109/TPDS.2016.2543731

[13] Muhammad Abdulhamid et al. “Fault tolerance aware scheduling technique for cloud computing environment using dynamic clustering algorithm” *Neural Computing and Applications* · July 2016 DOI: 10.1007/s00521-016-2448-8

[14] Guoqi Xie, et al. “Quantitative Fault-tolerance for ReliableWorkflows on Heterogeneous IaaS Clouds”, *IEEE TRANSACTIONS ON CLOUD COMPUTING*, NOV 2016 DOI 10.1109/TCC.2017.2780098

[15] SatjotKaur, Gurbhej Singh “Review on Fault Tolerance Techniques in Cloud Computing” *International Journal of Engineering and Management Research ISSN (ONLINE): 2250-0758* Volume-7, Issue-3, May-June 2017

[16] Yanchun Wang, et al. “Formulating Criticality-Based Cost-Effective Fault Tolerance Strategies for Multi-Tenant Service-Based

Systems" *IEEE TRANSACTIONS ON SOFTWARE ENGINEERING* DOI 10.1109/TSE.2017.2681667

[17] Xiwei Qiu et al. "Correlation Modeling and Resource Optimization for Cloud Service With Fault Recovery" *IEEE TRANSACTIONS ON CLOUD COMPUTING* DOI 10.1109/TCC.2017.2691323

[18] Priti Kumari, Parmeet Kaur "A survey of fault tolerance in cloud computing" *Journal of King Saud University – Computer and Information Sciences* (2018)

[19] Bashir Mohammed et al. "Failure Analysis Modelling in an Infrastructure as a Service (Iaas) Environment" *Electronic Notes in Theoretical Computer Science* 340 (2018) 41–54 Elsevier

[20] Moin Hasan, Major Singh Goraya "Fault tolerance in cloud computing environment: A systematic survey*" Elsevier Computers in Industry* 99 (2018) 156–172

[21] Mao-Lun Chiang et al. "An Agreement under Early Stopping and Fault Diagnosis Protocol in a Cloud Computing Environment" *IEEE Translations and content mining are permitted for academic research only* DOI:10.1109/ACCESS.2018.2859351

[22] Haoran Han et al. "Fault-Tolerant Scheduling for Hybrid Real-TimeTasks Based on CPB Model in Cloud" *IEEE ACCESS 2017* DOI 10.1109/ACCESS.2018.2810214

[23] PeiYun Zhang et al. "An Online Fault Detection Model and Strategies Based on SVM-Grid in Clouds" *IEEE/CAA JOURNAL OF AUTOMATICA SINICA*, VOL. 5, NO. 2, MARCH 2018

[24] Bhupesh Kumar Dewangan et al. "Self-characteristics based Energy-Efficient Resource Scheduling for Cloud" *Elsevier International Conference on Pervasive Computing Advances and Applications* 152 (2019) 204–211

[25] CHUN GAN et al. "An Overview of Fault-Diagnosis and Fault-Tolerance Techniques for Switched Reluctance Machine Systems" IEEE *Digital Object Identifier 10.1109/ACCESS.2019.2956552*

[26] Lei Wang et al. "Architecture-based Reliability-sensitive Criticality Measure for Fault-Tolerance Cloud Applications", IEEE *Transactions on Parallel and Distributed Systems* 2019, DOI 10.1109/TPDS.2019.2917900