ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 7, July 2014

Multiple Correlation Coefficient Approach for VM Migration

Sabina Singh

Assistant Professor, Computer Science Engineering Department, Chandigarh University,

Gharuan, Punjab. India

Abstract: In this research paper we have extended maximum correlation algorithm for detection of overload thresholds so that undue migration does not occur and the fault tolerance mechanism efficiently works. The improved/extended algorithm basically not only conducts correlation analysis but correlation with path analysis. By doing this we are able to find grain level fault tolerance as the overload may take multiple paths while processing multiple workload task. The result shows that by doing correlation with path analysis we are able to reduce host, VM reallocation time and able to avoid undue migration.

Indexterms: Cloud computing, VM-migration, correlation, overload detection, path analysis.

I. INTRODUCTION

Mitigation and migration is the way to make cloud more resilient, when it is under pressure of work or suffering from adversity. Infrastructure as a Service (IaaS) is changing how computing resources are being managed, delivered and consumed. Cloud providers are able to benefit from economies of sale by pooling large amounts of computing resources and running the underlying infrastructure at higher utilization levels than traditional-non-cloud-data centers. Providers also have the opportunity to over-commit their resources, relying on natural fluctuations-peaks and valleys-in computing demands across individual users.

Over commitment in IaaS clouds allows providers to sell more resources than they actually have. A physical machine with 8 CPUs, for example, can be sold to 16 users, each being offered one CPU per virtual machine (VM). In such a case, the physical machine has an over-commit ratio of 2. The provider is then hedging that-on average- each user only needs half of the requested resources. While a high over-commit ratio increases the average utilization of the underlying resources, it also increases the risk of provider-induced overload. This type of overload occurs when (over-committed) users demand enough of their resources such that – in combination –they exhaust all available physical capacity. In the example above, if two users demand 100 percent of their leased CPU resource, while the rest remain exactly at 50 percent, then some users will not get their fair share. Depending on the scheduling policy, some users (possibly all) will experience provider-induced overload. In the case of memory, provider induced overload can trigger dramatic drop in application performance [1]. In general, this problem can be solved by migrating certain VMs on the overloaded machine to other- relatively under-loaded-machines. The difficulty arises when the entire datacenter is over-committed. Deciding on which VMs to move and where to move them is important because migration should mitigate the overload without raising the risk of further overloads on the destination machines. It should also avoid negatively impacting the underlying network.

II. RELATED WORK

Anton Beloglazov and Rajkumar Buyya [2] - To understand the implications of the online nature of the problem, we conduct competitive analysis and prove competitive ratios of optimal online deterministic algorithms for the single VM migration and dynamic VM consolidation problems. Furthermore, we propose novel adaptive heuristics for dynamic consolidation of VMs based on an analysis of historical data from the resource usage by VMs. The proposed algorithms significantly reduce energy consumption, while ensuring a high level of adherence to the Service Level Agreements (SLA). Another direction for future research is the investigation of more complex workload models, e.g. models based on Markov chains, and development of algorithms that will leverage these workload models. Besides the reduction in infrastructure and on-going operating costs, this work also has social significance as it decreases carbon dioxide footprints and energy consumption by modern IT infrastructures.

III. GAP ANALYSIS

There is also a need to maintain idealize thresholds or the SLA thresholds, which needs to be maintained no matter what is the local threshold of each sub-data set and need to take this into account, even if it comes out to be a case of static threshold, or dynamic threshold, which time has proven are not enough to provide correct method for taking decision on mitigation and migration. There is far more need to correlate multiple factors really taking decision on mitigation and migration. Apart from CPU utilization pattern other factors like bandwidth, delay also need correlation under pressure of over loading.

IV. NEED AND SIGNIFICANCE OF RESEARCH

One way to mitigate such issues is to design algorithms that can detect over utilization of resources and predict any adversity in datacenter. Once this is done, we can build a fall back VM allocation policy or in simple words fault tolerance mechanism which may be reactive or proactive in nature. Therefore there is an urgent need to build VM allocation policy that works as fault tolerance algorithm to migrate and consolidate the virtual machines dynamically and efficiently based on non-linear nature of distribution of workflow jobs in tasks.

V. PROPOSED SOLUTION

The solution of the problem is based on understanding the utilization pattern of the host. Since, dynamic workload exists in which CPU usage varies with time , in this case, if CPU usage exceeds the available capacity of CPU, it is considered that there exists deviation from idealized threshold (SLA). Thus, there is a need of re-allocation or migration of VMs or simply a fall back mechanism that runs datacenter without performance degradation based on work variability.

Multiple-maximum Correlation Policy

Our proposed approach is an extension of the multiple correlation analysis used [2], it is a method of describing complex sequential relationship between measurements. Conceptually, the variables (measurements) used in the model are assumed to be all the measurements that matters. These are assigned to different levels in a sequence or cascade of influence, where the earlier levels affect the subsequent ones, but the reverse does not happen. In this sequence, all measurements in prior levels affect all subsequent levels, and the scale of the influence is described by the path coefficient, and partial correlation between measurements in the same level assumes that there are as yet unexplained common preceding influences. Mathematically, it is called path analysis [3], which consists of a repeated sequence of multiple correlation calculations from a correlation matrix, following the cascade of influences. This is carried out one level at a time, using all the measurements in preceding levels as independent variables, and the Standardized Partial Regression Coefficients (Path Coefficients) represents the size of each influence. This is followed by calculating Partial Correlation Coefficients between all the variables in the same level, corrected for all preceding variables as well. These Partial Correlation Coefficients represents common influences that have not as yet been explained by the model. For example table 5.1 shows correlation matrix that shows the relationship between the CPU usage and VMs.

TABLE 5.1 CORRELATION MATRIX

	VM1	**VM2**	**VM3**	**VM4**	**VM5**	**VM6**	**VM7**
VM1	1	0.5	0.2	0.2	0.3	0.6	0.4
VM2	0.5	1	0.1	0.1	0.3	0.7	0.5
VM3	0.2	0.1	1	0.4	0.5	0.1	0.5
VM4	0.2	0.1	0.4	1	0.6	0.1	0.4
VM5	0.3	0.3	0.5	0.6	1	0.3	0.7
VM6	0.6	0.7	0.1	0.1	0.3	1	0.8
VM7	0.4	0.5	0.5	0.4	0.7	0.8	1

Let VM7 represent one of the VMs that is currently considered for being avoided for migration. but , it might happen that .Given that CPU utilization of VM1,VM2 may be correlated with those of VM3 and VM4 as also have similar pattern of CPU utilization, and VM5 and VM6 also may have similar ,CPU utilization pattern to VM1 and VM2, we need a multiple correlation analysis to correct for all the inter-correlations, and identify the direct influence (Partial Correlation Coefficient) each member has on the selection VM7 as candidate which will have low performance and further degradation due over utilization . The VM7, is the machine is which is expected to have maximum load of the work or may have overload, and thus might degrade in performance as time comes, there for, a fallback policy is required if the VM7 fails or is overloaded with workload, hence, we run our fault tolerances to overcome such issue, here, this machine will be avoided in future re-allocation of machines or migrations of machines and the one which has the best performance will be considered for VM migration.

Experimental results

Following experiments were conducted to identify that how good is the VM migration policy. Experimental results shows in the form of tabular form and graphical form.

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 7, July 2014

TABLE-6.1 EVALUATION OF FAULT TOLERANCE MECHANISM

Experiment Number	Overloading Detection Algorithm	VM Migration Algorithm	Safety Parameter	No Of VM/Host	Workload Type
1	Local Regression	Max Correlation	0.5	50	Random
2	Local Regression	Max Correlation	0.6	50	Random
3	Local Regression	Max Correlation	0.7	50	Random
4	Local Regression	Max Correlation	0.8	50	Random
5	Local Regression	Max Correlation	0.9	50	Random
6	Local Regression	Max Correlation	1	50	Random
7	Local Regression	Max Correlation	1.2	50	Random
8	Local Regress	Extended maximum correlation	0.5	50	Random
9	Local Regression	Extended maximum correlation	0.6	50	Random
10	Local Regression	Extended maximum correlation	0.6	50	Random
11	Local Regression	Extended maximum correlation	0.7	50	Random
12	Local regression	Extended maximum correlation	0.8	50	Random
13	Local Regression	Extended maximum correlation	0.9	50	Random
14	Local Regression	Extended maximum correlation	1	50	Random
15	Local Regression	Extended maximum correlation	1.2	50	Random

Table 6.1 shows that overload detection algorithm, with VM migration algorithm, safety parameter, No. of VM's, Safety parameter, workload type. Local regression algorithm is used with maximum correlation policy and with proposed multiple maximum correlation policy for 50 VMs .The VMs are allocated according to the resource requirements defined by the VM types.

TABLE - 6.2 RESULTS OF EXPERIMENTS

Exeriment Number	Number of VM migrations	VM selection mean	VM selection standard deviation	host selection standard deviation	VM reallocation mean	VM reallocation standard deviation	Safety Parameter
1	257	0.00137	0.007	0.00531	0.00022	0.00182	0.5
2	361	0.00283	0.009	0.00561	0.00016	0.0016	0.6
3	817	0.00766	0.014	0.00535	0.00104	0.00392	0.7
4	1457	0.0103	0.015	0.00552	0.00141	0.00486	0.8
5	1735	0.01228	0.012	0.00512	0.00174	0.00491	0.9
6	2430	0.0135	0.015	0.00606	0.00394	0.00683	1
7	3081	0.01326	0.02	0.00589	0.00658	0.00846	1.2
8	273	0.00109	0.006	0.00499	0.0002	0.0017	0.5
9	377	0.00201	0.009	0.00559	0.000155	0.0012	0.6
10	817	0.00166	0.013	0.005	0.001	0.00301	0.7
11	1457	0.0111	0.014	0.005	0.00114	0.00409	0.8
12	1735	0.01122	0.011	0.005	0.00155	0.003999	0.9
13	2430	0.01335	0.014	0.00588	0.00355	0.00513	1
14	30[1	0.01325	0.02	0.005	0.006	0.005999	1.2

Table 6.2 shows the 14 experiments. Experimental results shows that it reduces the mean-time, standard deviation mean, VM standard mean. VMs utilize less resources according to the workload data, creating opportunities for dynamic consolidation but if there is overloading or adversity it again forms another type of challenge in terms of defining the VM reallocation or migration policy after the overloading and adversity has been detected, which is the core part of any fault tolerance mechanism in cloud computing.

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 7, July 2014

TABLE - 6.2 RESULTS OF EXPERIMENTS

Exeriment Number	Number of VM migrations	VM selection mean	VM selection standard deviation	host selection standard deviation	VM reallocation mean	VM reallocation standard deviation	Safety Parameter
1	257	0.00137	0.007	0.00531	0.00022	0.00182	0.5
2	361	0.00283	0.009	0.00561	0.00016	0.0016	0.6
3	817	0.00766	0.014	0.00535	0.00104	0.00392	0.7
4	1457	0.0103	0.015	0.00552	0.00141	0.00486	0.8
5	1735	0.01228	0.012	0.00512	0.00174	0.00491	0.9
6	2430	0.0135	0.015	0.00606	0.00394	0.00683	1
7	3081	0.01326	0.02	0.00589	0.00658	0.00846	1.2
8	273	0.00109	0.006	0.00499	0.0002	0.0017	0.5
9	377	0.00201	0.009	0.00559	0.000155	0.0012	0.6
10	817	0.00166	0.013	0.005	0.001	0.00301	0.7
11	1457	0.0111	0.014	0.005	0.00114	0.00409	0.8
12	1735	0.01122	0.011	0.005	0.00155	0.003999	0.9
13	2430	0.01335	0.014	0.00588	0.00355	0.00513	1
14	30[1	0.01325	0.02	0.005	0.006	0.005999	1.2

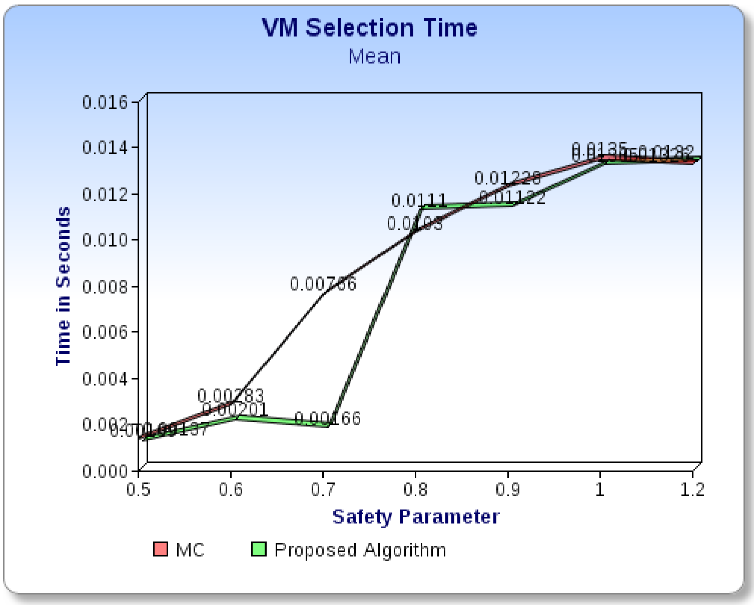

Figure 6.1- Graph of VM Selection Time Mean

It can be seen from the above graph in Figure 6.1, that the mean selection time for the Fault tolerance algorithm for the proposed algorithm is coming below the value of the Maximum Correlation (MC) algorithm in most of the cases but at the time when the safety parameter value is around 0.8 , the value goes above the MC value ,and then finally it has similar magnitude value when it comes to safety parameter value of 1.2.From the above line curve we can also infer that the proposed algorithm overall VM selection policy is performing better as it is taking less average time in selecting the new VM placement when utilization varies due to safety parameter.

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 7, July 2014

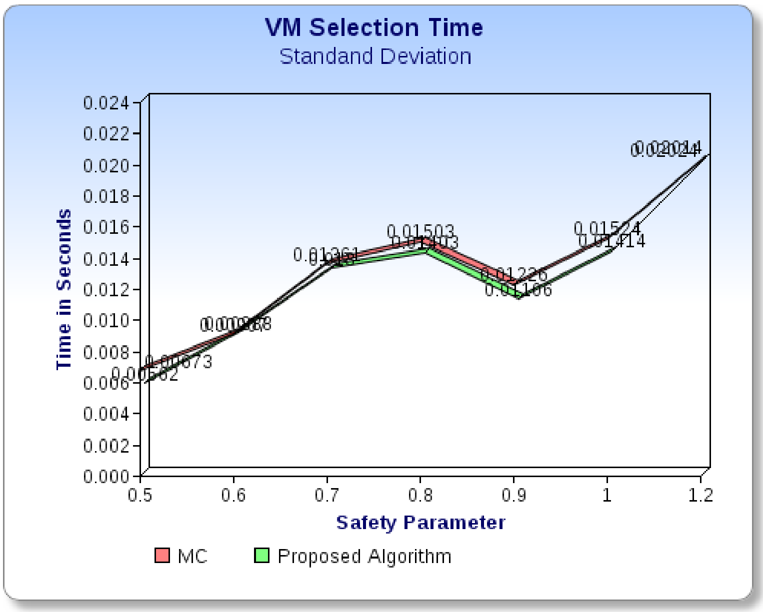

Figure 6.2- Graph of VM Selection Time Standard Deviation

From the above graph Figure 6.2 also we can apparently infer that proposed algorithm shown in green line is performing on the most of the levels of utilization, the curve showing lower deviation from the average time taken in VM selection, this may be attribute the fact the overloading detection algorithm (Local Regression) is same while initial selection of the VMs, but it is standard deviation value falls below as average value is also.

CONCLUSION

It is apparent from the performance graphs that extended maximum correlation is basically reducing VM selection time mean. Also the maximum correlation algorithm is extended with path analysis.

VI. CONTRIBUTIONS TO RESEARCH

- Proposed a VM migration policy.
- Design the approach for overloaded detection.
- Implementation of VM migration policy in cloud environment.

VII. FUTURE RESEARCH

- This work shows implementation of the reactive fault tolerance approach. It can be used for proactive fault tolerance
- It can be used to handle multiple tasks.
- It can further be used to merge with scheduling algorithms.

REFERENCES

[1] D.Williams,H.Jamjoom,Y.-H.liu,and H.Weatherspoon, "Over-Driver:Handling Meomry Overload in an Oversubscribed Cloud," in proceedings of the 20111 ACM SIGPLAN,SIGOPS International Conference on Virtual Execution Environments,2011

[2] Beloglazov,A., and Buyya,R.,"Optimal Online Deterministic Algorithms and Adaptive Heuristics for Energy and Performance Efficient Dynamic Consolidation of Virtual Machines in Cloud Data Centers".

[3] Lleras,C., "path analysis".

BIOGRAPHY

Sabina Singh is currently working as Assistant Professor in Chandigarh University, gharuan,Punjab. She is an alumni of Thapar University, Patiala (Punjab), having obtained her Masters from the prestigious institution in Computer Science Engineering. She has done her Bachelors in Technology in Computer Science Engineering from Punjab Technical University, Jalandhar (Punjab).