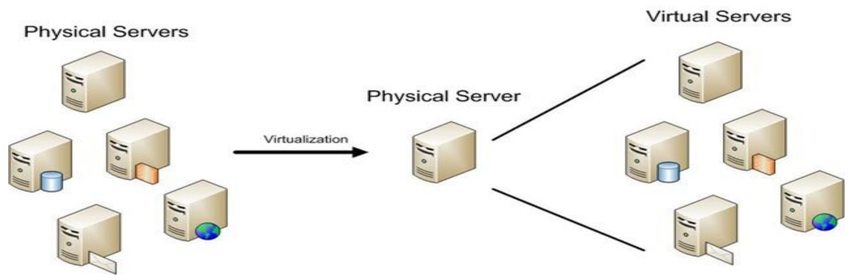
SERVER CONSOLIDATION ALGORITHMS USING LIVE MIGRATION OF VIRTUAL MACHINES

Pathikrit Ghosh*, E. Soundararajan, M. Sai Baba
Scientific Information Resource Division, Indira Gandhi Centre for Atomic Research, Kalpakkam
***Department of Computer Science, Birla Institute of Technology and Science, Pilani**

Abstract

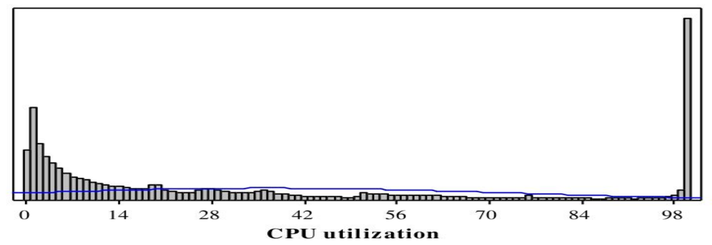
Modern data centres face hassles like server sprawl which is a result of under-utilization of servers. Virtualization using server consolidation is one of the ways these problems can be tackled. This method provides a uniform pool of resources where a single server can host multiple computing environments encapsulated as a virtual machine, all of which are isolated from each other. In such a virtualized setup, it is possible to use lesser nodes and save on resources like power to do the same work. This work aims at finding an efficient server consolidation algorithm using live migration of virtual machines to meet the needs of different data centers. This problem is essentially similar to a d-vectored bin-packing problem, which is NP-hard. Hence we use heuristics to find a solution to this problem. Three heuristic algorithms analyzed here are – Sercon, Khanna's algorithm and Sercon with load balancing. Sercon algorithm aims at server consolidation using memory and CPU usage as parameters. It migrates the VMs on a physical node with the least score onto some other nodes such that this node has no load and hence can be gotten rid of. This way it ensures minimum number of servers are used. Khanna's Algorithm is a polynomial time heuristic which also aims at packing VMs as tightly as possible by maximizing the variance on its parameters across all the physical hosts. Our algorithm incorporates load balancing(among the consolidated servers) into the Sercon algorithm. The platform used to implement is CentOS-7. The aim is to generate test-cases by simulating loads on cluster of virtual machines and use that as an input to analyze the algorithms. The work aims to provide a conclusion and a suggestion on the environments where one algorithm should be preferred over the another.

Server Consolidation in a virtualized setup



- Server Consolidation is a process by virtue of which we can reduce the total number of host servers in a data center.
- This helps in reducing power consumption, data center footprint and other monetary costs.
- It is of mainly two types, namely, batching and periodic.
- Batching involves migration from a non-virtualized cluster to a virtualized cluster.
- Periodic process involves dynamic reallocation of Virtual Machines (VMs) without any interference in the applications running on the VMs.

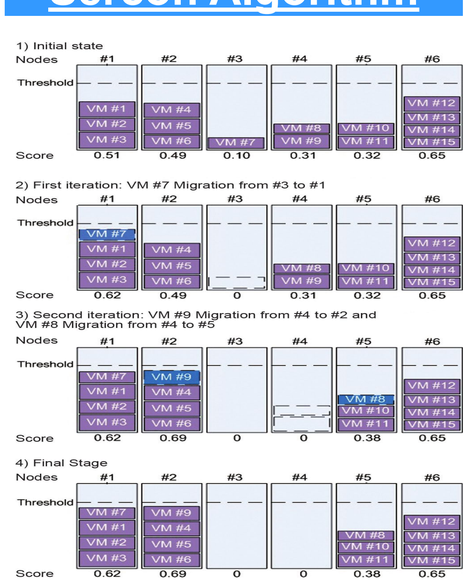
Need for Server Consolidation in data centers



Server Sprawl, that is massive under-utilization of servers in the current virtualized setup.
The figure above shows data from CPU utilization form 100 “Planet lab” nodes over a period of 10 days.
This highlights the current problem and emphasizes the need for server consolidation.

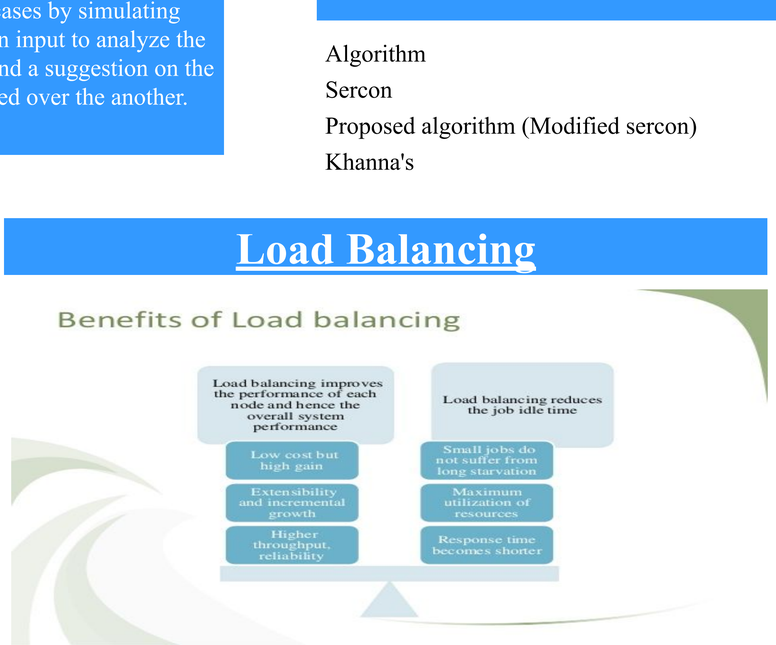
Algorithm	Parameters Used to analyze	Aim
Sercon	CPU, Memory	Server Consolidation, Minimize migrations
Proposed algorithm (Modified sercon)	CPU, Memory	Server Consolidation, Load Balancing
Khanna's	CPU, Memory	Sever Consolidation, Hotspot Mitigation, Coldspot Mitigation

Sercon Algorithm



- We sort the nodes based on loads in Physical Machines(PMs) in decreasing order.
- This is done by giving a score to each PM based on their CPU and memory loads.
- The VMs on the last(least loaded) node in this list is then selected as a candidate for migration.
- The VMs are also then sorted in the decreasing order of their scores.
- We the try to allocate VMs one by one on the first(most loaded) node, and if the attempt is unsuccessful, then the second node, and so on.
- Migration occurs if and only if all candidate nodes on a particular candidate PM can be migrated to another PM.
- The diagram above shows an illustration for the process of migration due to Sercon algorithm.

Load Balancing



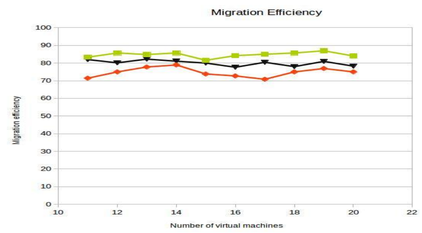
Proposed algorithm with load balancing

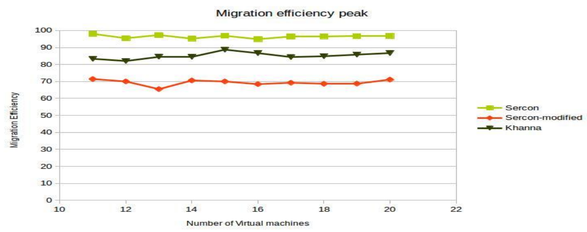
- First all PMs are sorted according to their scores.
- The scores are assigned to each PM and VM, similar to a Sercon algorithm.
- Among the possible target machines for migration, VMs are migrated onto a machine which reduces the overall variance of the system.
- Threshold figures(SLA requirements) are followed at each step.
- Scalability of this is difficult for big data centers due to large number of migrations.

Khanna Algorithm

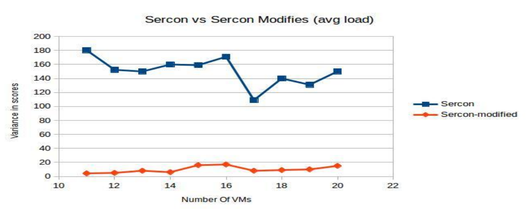
- It is a dynamic Algorithm that suggests a new migration schedule every 't' intervals of time.
- Reduces hot-spots by moving VMs to other PMs that can handle the load.
- Eliminated cold spots by trying to increase variance among physical machines ultimately resulting in releasing a node.
- Server Consolidation is done by migrating VMs from less loaded machines to heavily loaded ones without caring for migration efficiency.

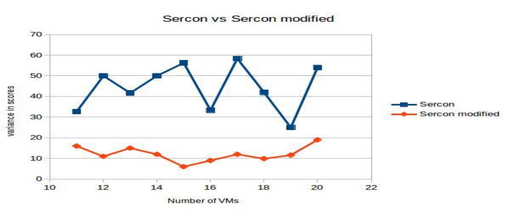
Results





The above figures shows migration efficiency during the period of average load and during peak load period. From the graphs we can infer that Sercon in the most efficient algorithm while Khanna also has a good migration efficiency. Our proposed algorithm suffers a bit on migration efficiency.





The figure above shows variance among the scores in physical machines comparing Sercon and our proposed algorithm. The variance is at least 15% better using the same number of servers post consolidation both on average and peak load.

Conclusion

Large data centers should use Sercon while small data center can use our modified Sercon algorithm. Our algorithm is not suitable for large data centers due to its poor migration efficiency and the high cost associated with it. In centers where the process of server consolidation needs to be done very frequently Khanna is a good choice as its a dynamic algorithm that runs every finite interval of time.

References

- Aziz Murtazaev and Sangyoon Oh, “Sercon: Server Consolidation Algorithm using Live Migration of Virtual Machines for Green Computing”, *IETE Technical review, June 2011.*
- *Gunjan Khanna,Kirk Beaty, Gautam Kar,Andrzej Kochut,“Application Performance Management in Virtualized Server Environments”,10th IEEE Conference on Network*
- C.Clark, K. Fraser,S. Hand,J. Hansen, E. Jul, C. Limpach, I. Pratt and A. Warfiled, “*Live Migration of Virtual Machines” , NSDI 2005*