ENERGY-EFFICIENT RESOURCE UTILIZATION ALGORITHMS IN CLOUD DATACENTER SERVERS

Research brown bag presentation

Presenter: Kenga Mosoti Derdus
Faculty: FIT

DEFINITION OF TERMS

Cloud computing

virtualization

Physical machine (PM)

Virtual machine (VM)

VM consolidation

Datacenter

Datacenter server

Cloud/application workloads

Homogeneous workloads

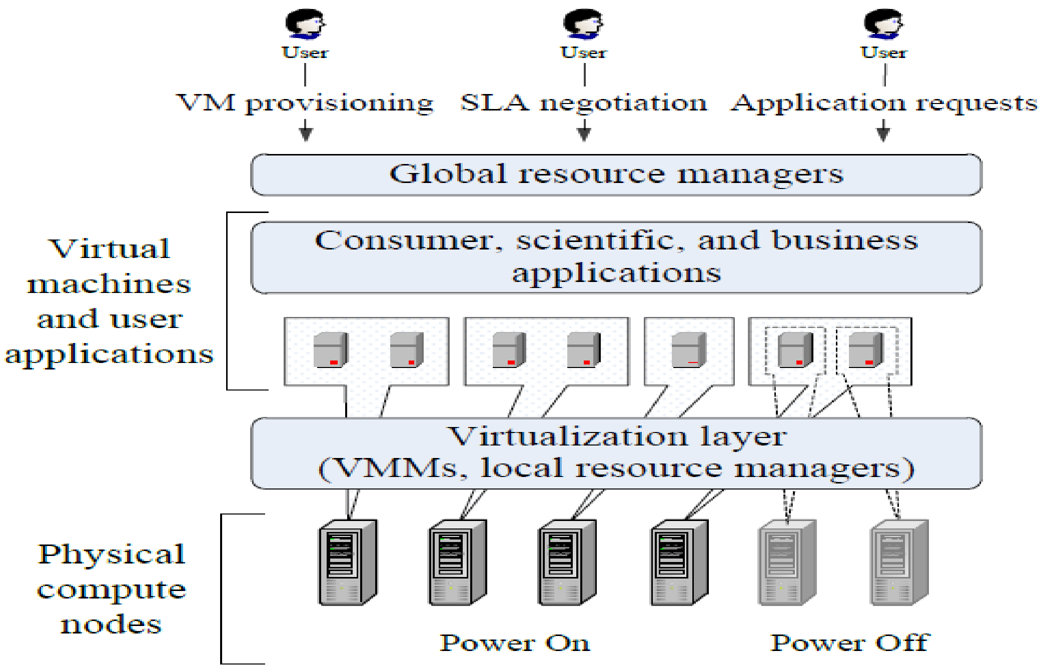
Heterogeneous workloads

Computing resources

Cloud Service Provider (CSP)

Cloud users/customer

Computing proportionality



BACKGROUND

Cloud computing is an important IT paradigm for providing pay-as-you-go service (Chaima, 2014)

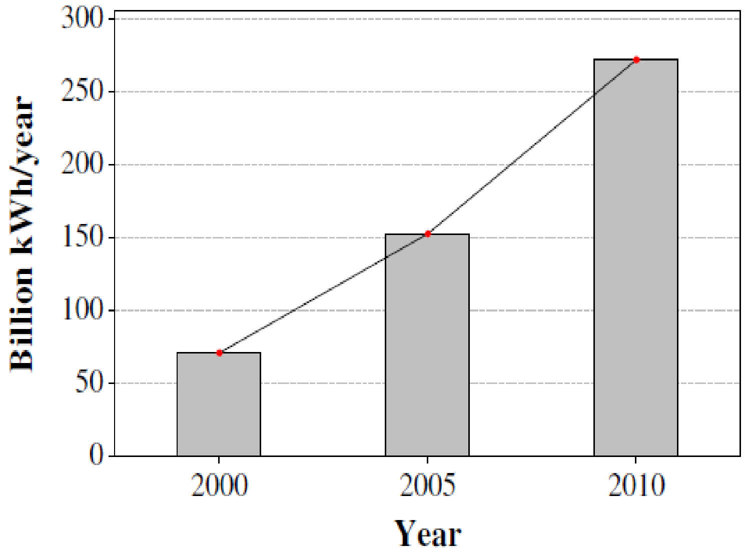
56 % of enterprise workloads were processed in cloud (Ullah, Hassan and Khan, 2017)

However, cloud datacenters consume a lot of electrical energy – stands at 3% of global consumption (Rallo, 2014)

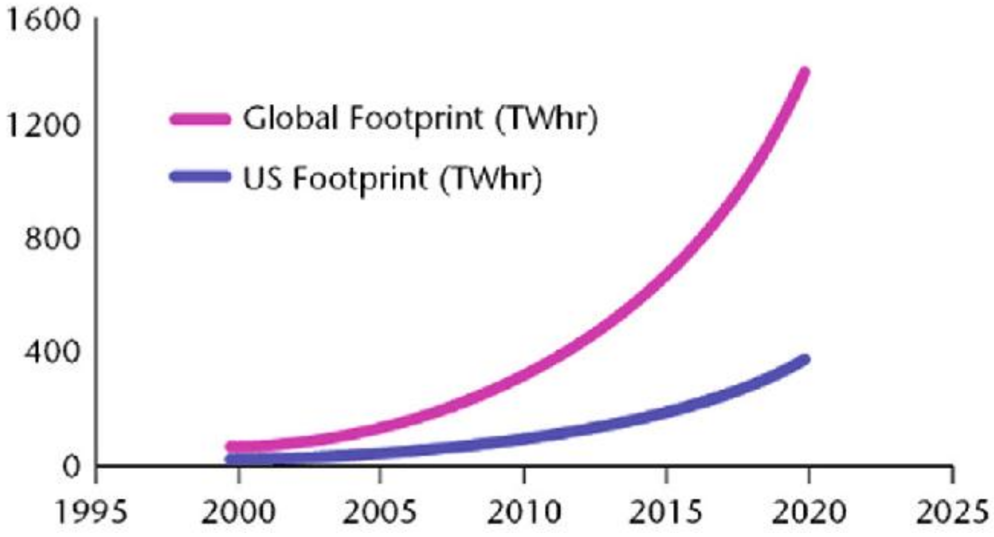
An average datacenter (covering around 9300 M^2.) consumes as much energy as 25, 000 households in the United States (US).

Result is high operating costs (electricity bills), which reduces profits, increases Total Cost of Ownership (TCO) of datacenter infrastructure and increases carbon dioxide emission to the environment

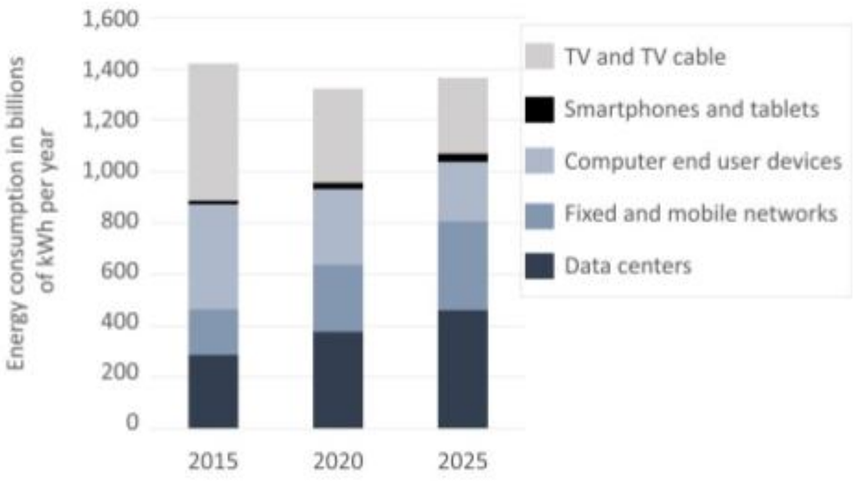
BACKGROUND



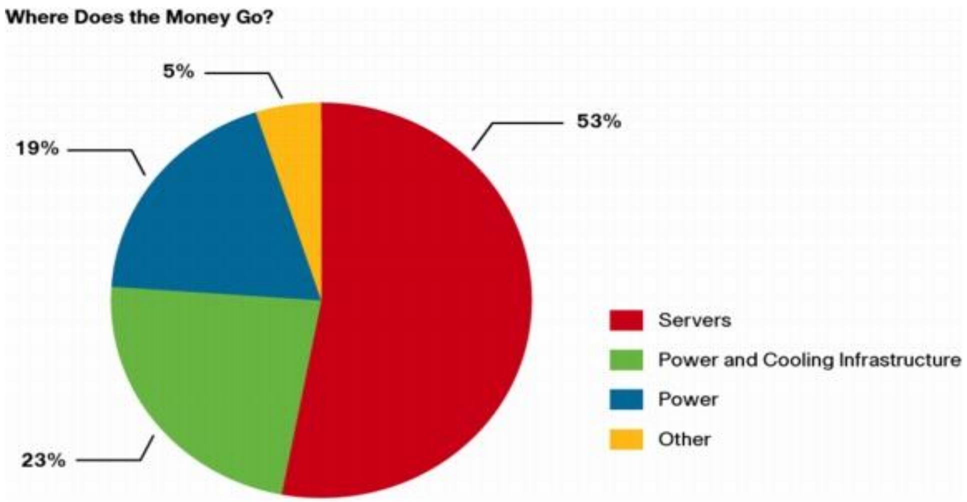
Billion kWh/year
300
250
200
150
100
50
0
2000
2005
2010
Year



1600
1200
800
400
0
Global Footprint (TWhr)
US Footprint (TWhr)
1995
2000
2005
2010
2015
2020
2025

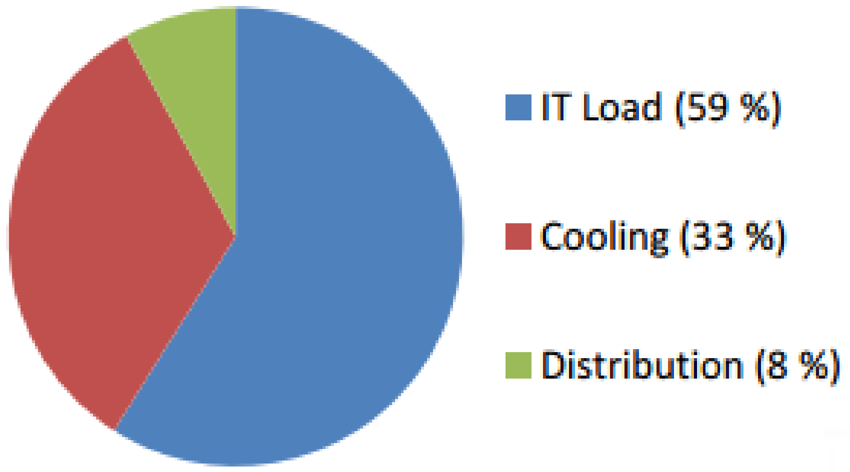


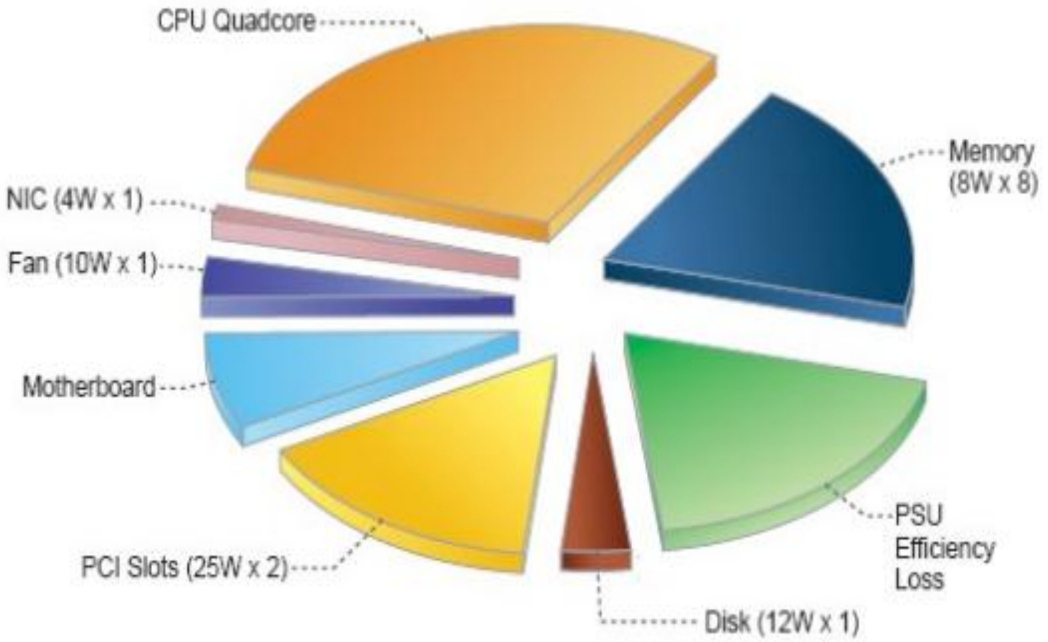
Energy consumption in billions of kWh per year
1,600
1,400
1,200
1,000
800
600
400
200
0
2015
2020
2025
TV and TV cable
Smartphones and tablets
Computer end user devices
Fixed and mobile networks
Data centers



Where Does the Money Go?
5%
53%
19%
23%
Servers
Power and Cooling Infrastructure
Power
Other

WHERE DOES THE POWER GO





WHAT IS THE PROBLEM

Energy consumption and wastage leads to high operating costs of CSPs, high cost of using cloud services (from users), carbon emission

What causes energy wastage?: low server utilization

Why is it difficulty to ensure higher server utilization?: performance unpredictability, application workload interference, resource over provisioning from unexperienced users.

Are there any solutions currently? Which ones?: Yes, Workload/VM consolidation, DVFS, Hardware optimization by manufacturers and Server switching

Why are current solutions not working?

VM consolidation – does not consider workload characteristics

DVFS – only good for CPU and not other resources such as RAM

Server switching – how do we know when to switch off and on?

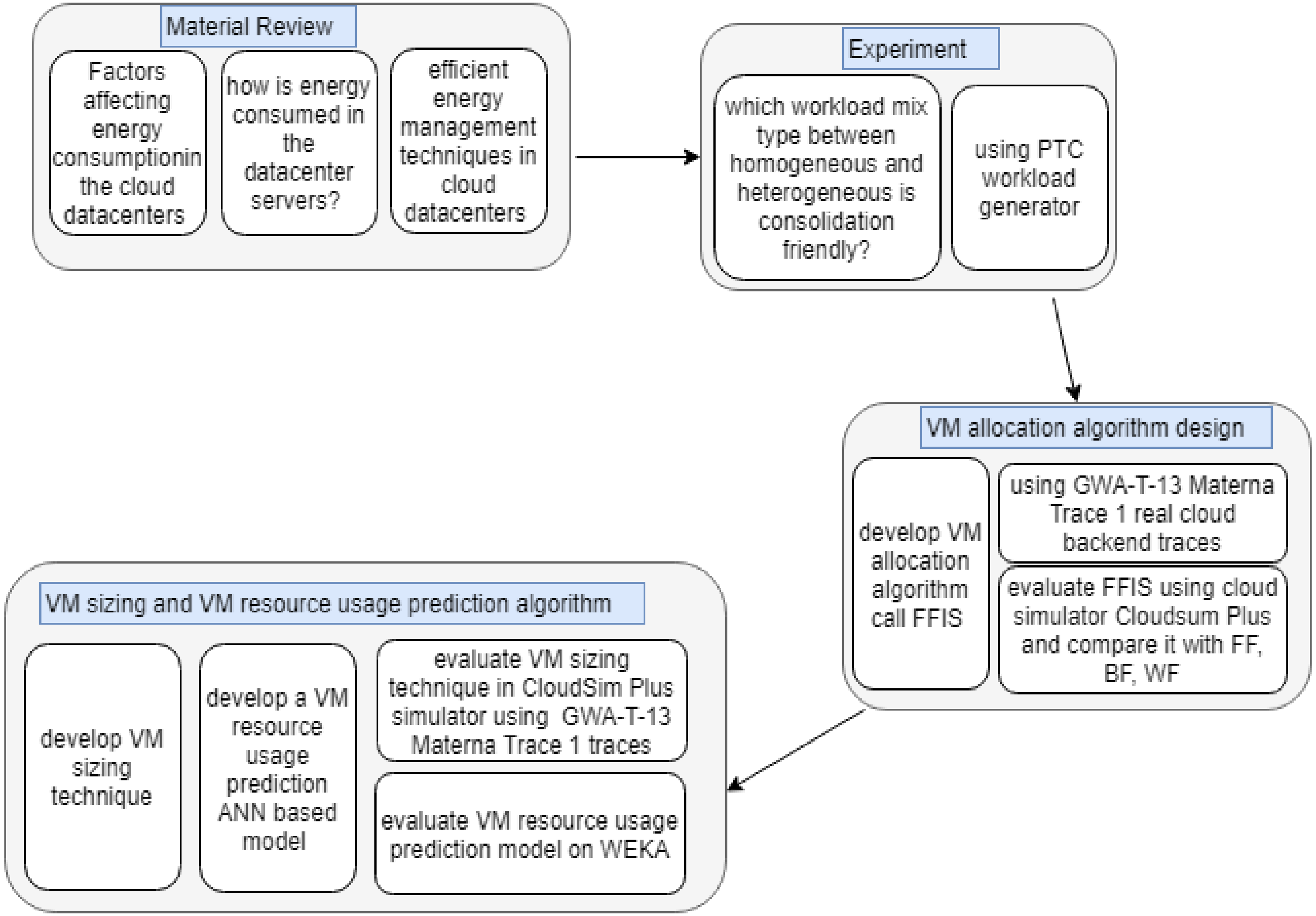
PURPOSE OF THE STUDY

The purpose of this study to achieve the following:

- Investigate the performance and power profiles of different cloud workloads
- Design, implement and evaluate a VM allocation algorithm
- Design, and evaluate prediction algorithms ate a VM sizing and VM resources

PROCEDURE

Experimental design was used

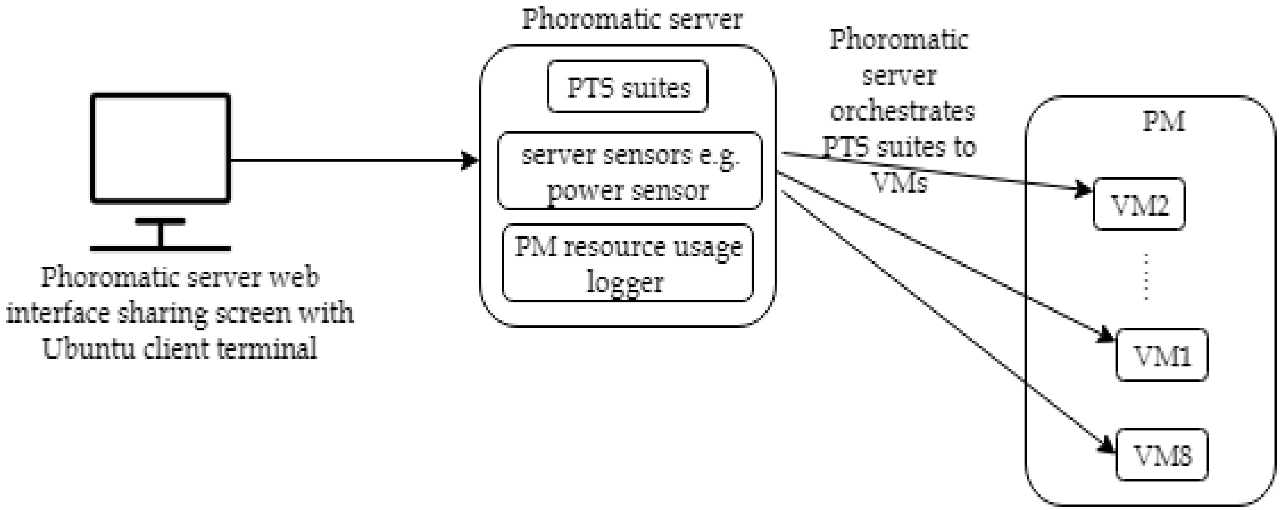


SOURCE OF DATA DATACENTER

This table shows the configurations of datacenter, which produced data used

No. of hosts	**49**
No. of VMs	520
No. of CPUs	69 (454 cores)
Memory size (in GB)	6780
Hypervisor	VMware ESX
No. of cores allocated per VM	Varying (1,2,4,6 and 8)
Memory size allocated per host (in GB)	Varying (2,4,8 and 16)
Host static power	60 % of host peak power

PERFORMANCE AND POWER PROFILES OF DIFFERENT CLOUD WORKLOADS



Results:

Heterogeneous workloads are consolidation friendly as compared to homogeneous workloads

Workload consolidation limiting factor for CPU intensive workloads and disk intensive workloads is power consumption and performance respectively.

VM ALLOCATION ALGORITHMS

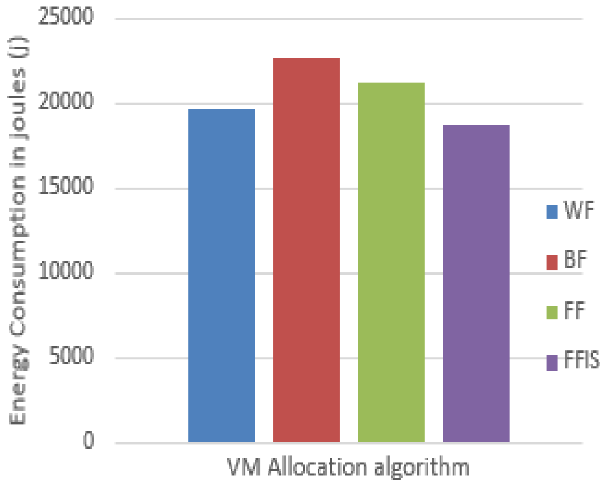
AIM: Design a workload-aware VM allocation algorithm, FFIS

Implementation: Algorithms implemented on CloudSim Plus

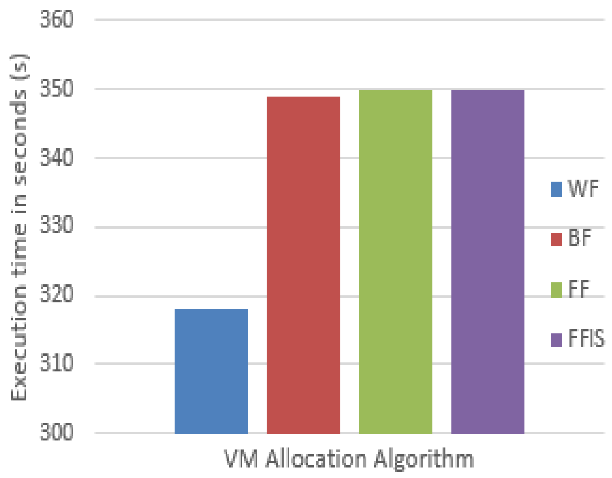
Source of data used: GWA-T-12

Technique used: workload Clustering using K-means

Results:



a) Total amount of energy consumed by all the 46 hosts for different VM allocation algorithm



b) Total execution time for different VM allocation algorithm

VM SIZING ALGORITHM AND VM RESOURCE USAGE PREDICTION

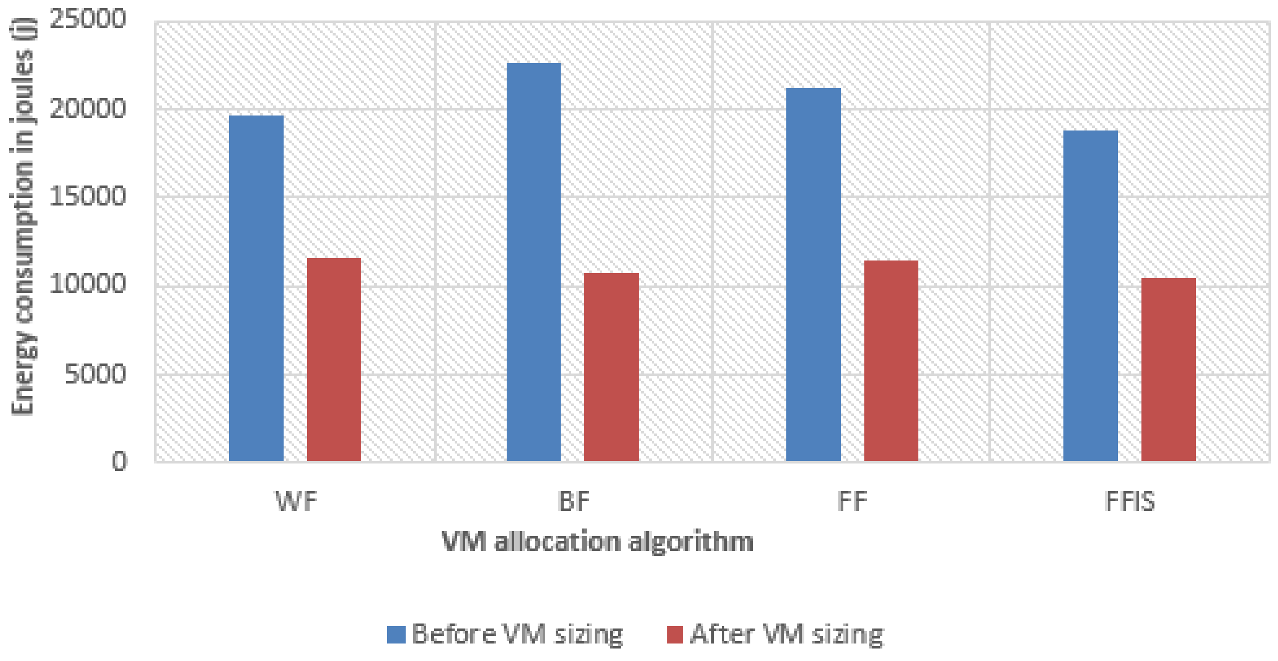
AIM: Design a VM sizing algorithm to address resource over-provisioning

: Design neural network model for VM resource usage to address resource usage unpredictability

Implementation: Algorithms implemented on CloudSim Plus

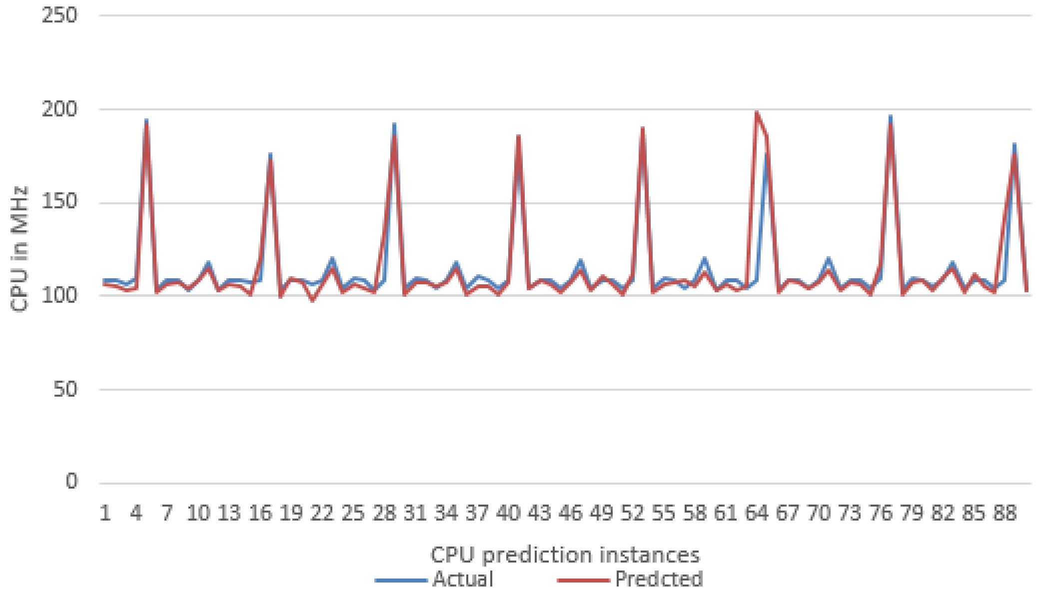
Source of data used: GWA-T-12

Results:

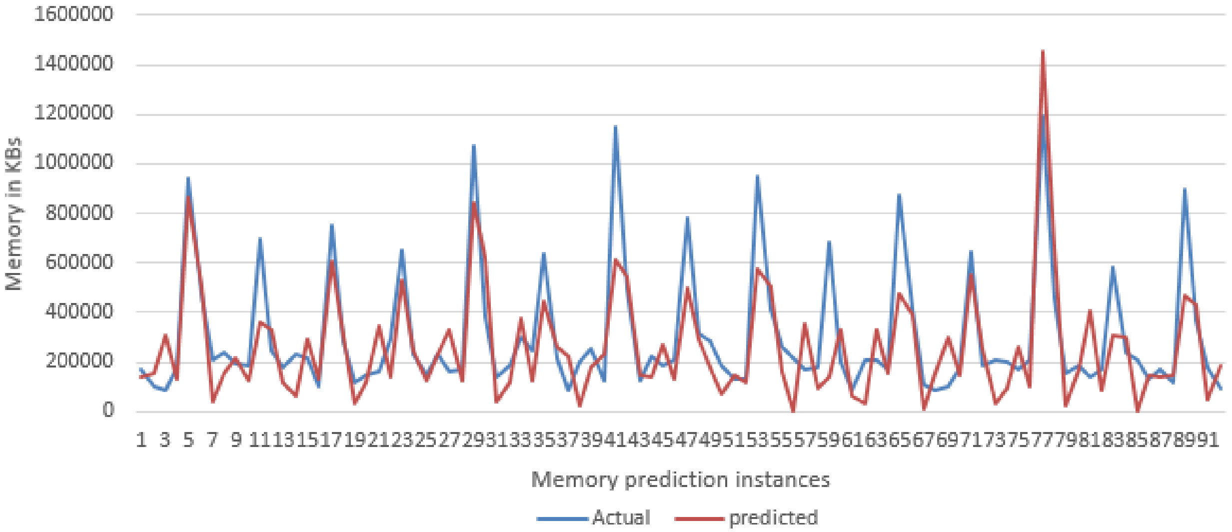


VM SIZING ALGORITHM AND VM RESOURCE USAGE PREDICTION

Performance of ANN model



VM 172	MAE	6.7	142359
	MAPE	5.8	58.9
	RMSE	17.7	201971
	SR	31	39.8



CONCLUSIONS

Heterogeneous workloads are consolidation friendly as compared to homogeneous workloads

Workload consolidation limiting factor for CPU intensive workloads and disk intensive workloads is power consumption and performance respectively.

Workload-aware VM allocation algorithm has a big potential in reducing energy wastage in data centers

VM sizing techniques have a potential in addressing resource over-provisioning

ANN models can achieve great success in forecasting future resource usage for timely provisioning to ensure QoS.

THANK YOU