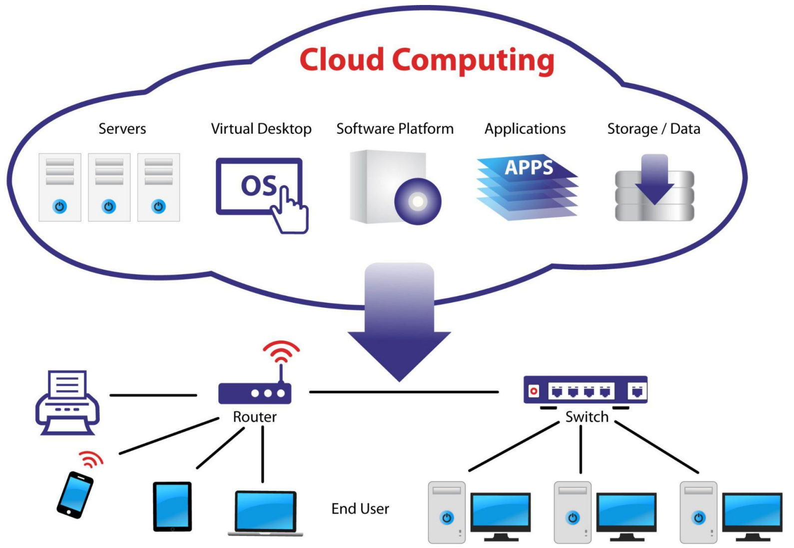
Resource Allocation in Cloud Computing Data Centers

Oregon State University
College of Engineering

What is Cloud Computing?

- Cloud service provider (CSP) provides pool of configurable computing resources [5]
- End-users invoke and release resources on a pay-per-use basis
- Users don’t need to know details of virtual/physical machines, task management [9]
- CSP guarantees performance to end-users under service level agreements (SLAs) [8]



https://www.cloudloadsolution.com/wp-content/uploads/2019/12/cloud_computing 3-scaled.jpg

Oregon State University
College of Engineering

Motivation for Improving Resource Allocation

[5,8,9]

- As cloud computing becomes more widespread, power/thermal cost increases exponentially
- Demand fluctuates unpredictably, which makes optimizing performance difficult
- CSPs must balance improving cost, energy usage, etc. with fulfilling end-user’s expectations and maintaining SLAs

Oregon State University College of Engineering

Three-Tier Architecture

Presentation tier(Front End)

- What the user sees

Application tier (Logic Tier)

- Where the information process occurs

Data tier(Back End)

- Where and how data is stored

Oregon State University
College of Engineering

Other Architectures and Topologies

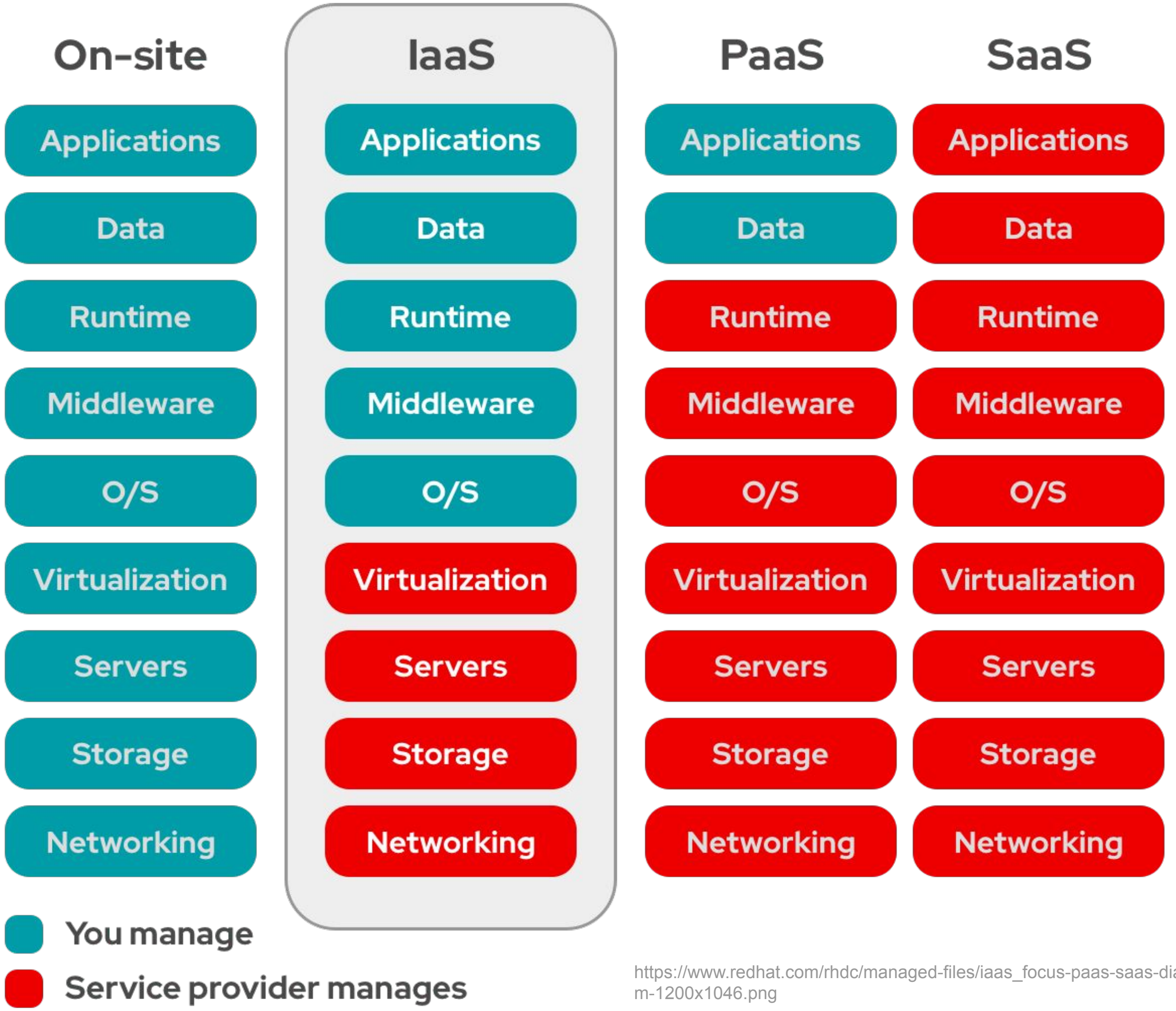
-Microservices Architecture

-Hybrid/Multi Cloud Topology

-Edge Computing Topology

Cloud Service Categories

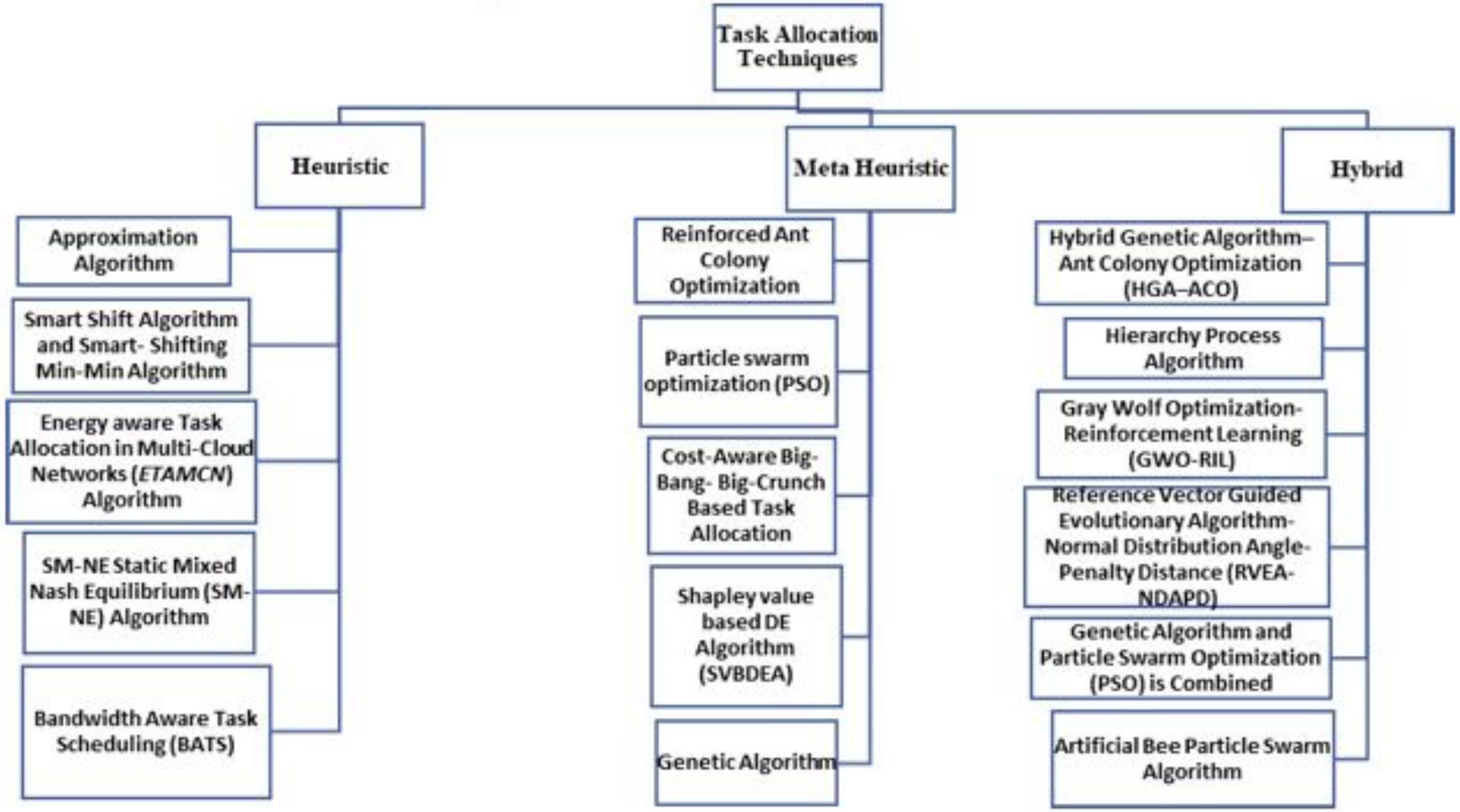
Oregon State University
College of Engineering



https://www.redhat.com/rhdc/managed-files/iaas_focus-paas-saas-diagram-1200x1046.png

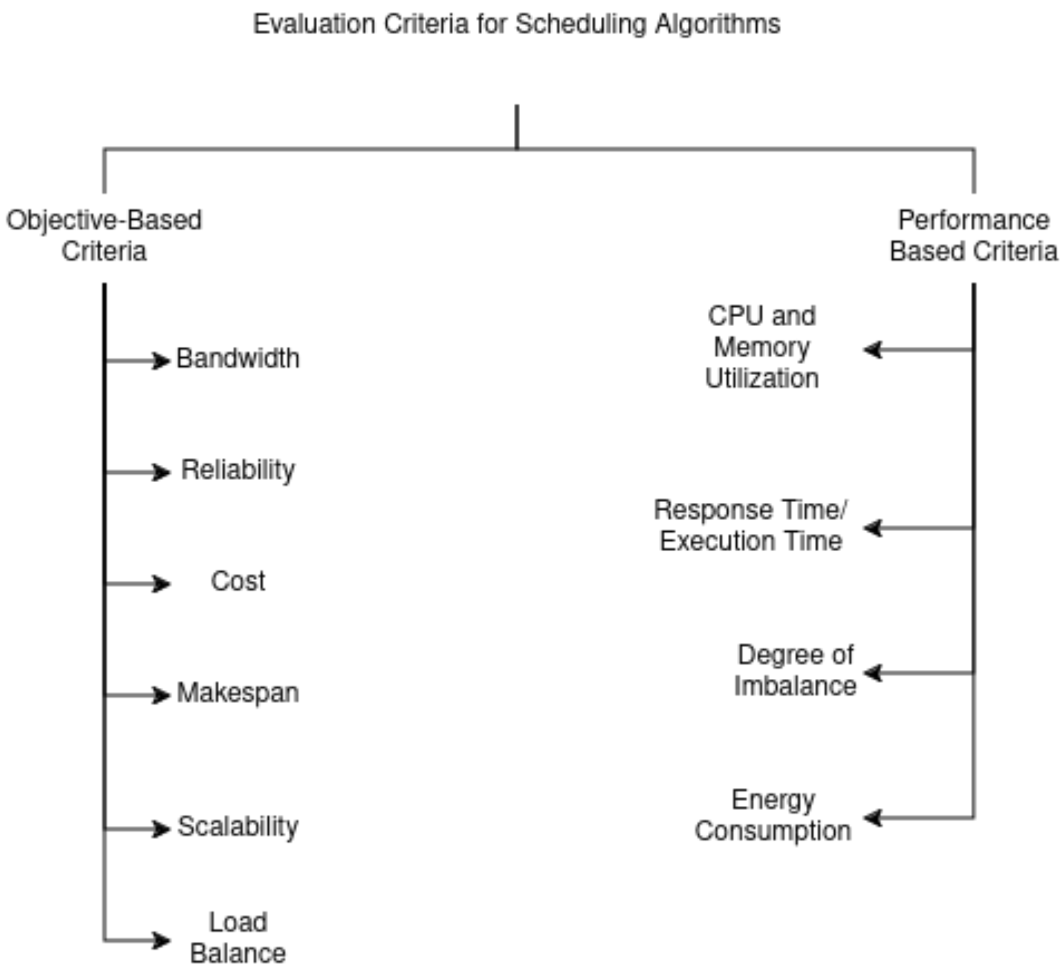
- Amazon Web Service(AWS), Microsoft Azure, and Google Cloud
- These larger models tend towards IaaS

Scheduling Techniques



[5]

Evaluation Metrics



Objective-Based Metrics

Oregon State University
College of Engineering

- Cost- applies to both end-user and provider. The service provider wants to maximize profit, while the end user wants to minimize expenses
- Makespan- Total time it takes for a given set of tasks to complete. Efficient scheduling = minimize makespan
- Scalability- Ability of the algorithm to meet the requirements of any number of end users or any number of tasks

Objective-Based Metrics

- Load Balance - Spreading tasks evenly across given resources
- Reliability - Likelihood a given task can be completed without failure

Performance Metrics

Oregon State University
College of Engineering

- CPU and Memory Utilization: How much memory and CPU are utilized running tasks under the scheduling algorithm
- Response Time: How long it takes for a task to actually execute and finish
- Degree of Imbalance: How imbalance load is between virtual machines
- Energy Consumption: How much energy is consumed running tasks under the scheduling algorithm

Oregon State University
College of Engineering

Heuristic Task Scheduling Algorithms

Heuristic Algorithms - Overview [5]

- Provides an approximate solution rather than optimal scheduling
- Derived using past information about the platform
- Attempts to capture relationship between metrics and hardware resource allocation, user workload patterns
- Handles estimated workload in coarse time scales (e.g., hours/days), maintains long-term workload

Oregon State University College of Engineering

Heuristic Algorithms - Advantages & Drawbacks [5,8]

Advantages:

- May be faster than traditional approaches
- Well-suited for online task-scheduling
- Generally simple to use

Drawbacks:

- Poor performance when there aren't many previous data
- Poor performance when data don't follow a particular distribution
- Expensive in terms of storage cost, processing time complexity

Example - Energy-Aware Task Allocation for Multi-Cloud Networks (ETAMCN) Algorithm [9]

- Expected Time to Completion (ETC) and Energy Consumption (EC) are calculated for each VM in each cloud
- Incoming tasks are stored in max-heap structure
  - Priority is calculated as task length divided by task deadline
- For each task, search for a set of VMs that can accommodate the task using the ETC matrix
- From this set, select the VM that minimizes energy consumption using the EC matrix
- Once the task is completed, the VM is updated in both matrices

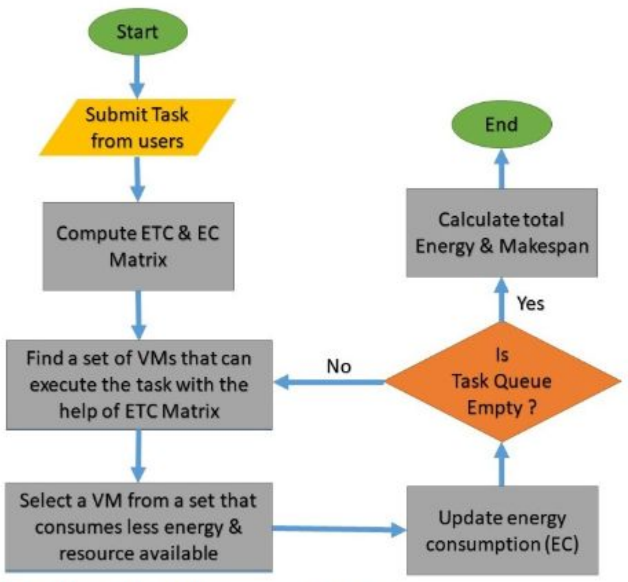


FIGURE 3. Flow-chart of the proposed algorithm.

Comparison to Other Heuristic Algorithms [9]

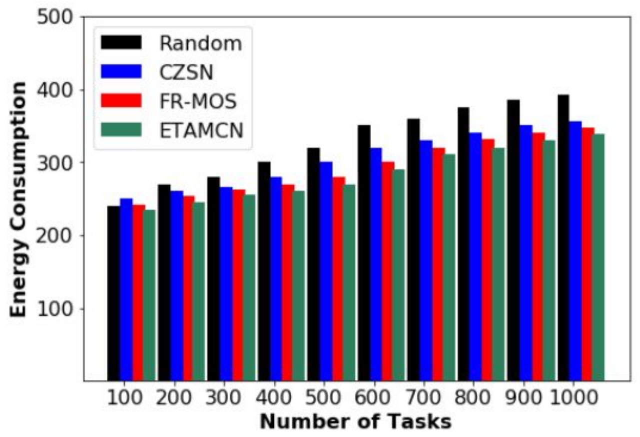


FIGURE 4. Energy Consumption Vs the number of Tasks where the number of VM is 80.

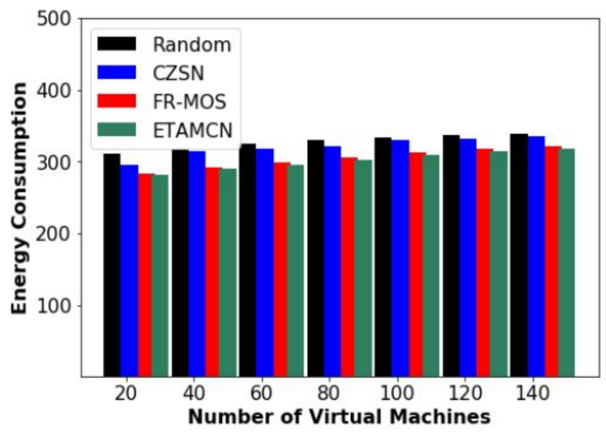


FIGURE 6. Energy Consumption Vs the number of VMs where the number of the task is 500.

TABLE 4. Simulation summary of energy consumption with respect to energy and number of tasks.

Scenario-1			Scenario-2		
No. of Tasks	Energy Consumption	SLA Violation	No. of VMs	Energy Consumption	SLA Violation
200	2318	2.7	20	310	11
400	3125	4.9	40	318	6.2
600	3719	7.9	60	325	3.8
800	4607	13.8	80	330	3.3
1000	5632	24.2	100	336	2.1

Meta-Heuristic Task Scheduling Algorithms

Meta-Heuristic Algorithms - Overview

- Deals with high level problems that may not have a clear solution.
- The large amount of data it takes to lead to an exact solution can lead to a heuristic solution.
- Known to simulate natural processes to achieve solutions

Meta-Heuristic Algorithms - Advantages & Drawbacks

Advantages:

- Can provide exact answers
- Can process large or complex solution space.

Drawbacks:

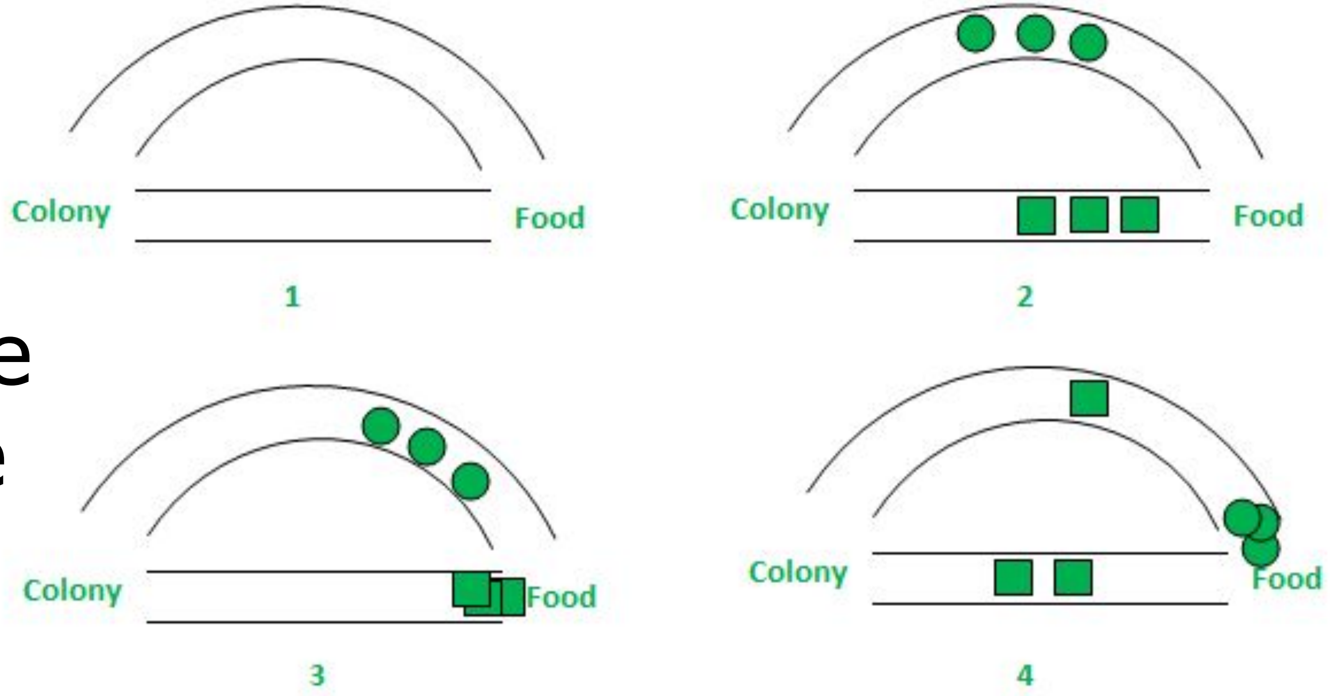
- Takes a long time
- Takes more processing resources

Example-Ant Colony Algorithm

Oregon State University
College of Engineering

-Simulates the natural process of ant colony foraging pattern

-The ants move through a path and leave behind a “pheromone”. The more the pheromones on a path, the more likely an ant will take the path.[6]



https://www.geeksforgeeks.org/introduction-to-ant-colony-optimization/

Oregon State University
College of Engineering

Hybrid Task Scheduling Algorithms

Hybrid Algorithms - Overview

- Combination of multiple optimization algorithms to increase amount of objectives covered
- Goal is to take strengths of each algorithm and combine them into one
- Can be a combination of meta-heuristic and heuristic, two or more heuristic algorithms, two or more meta-heuristic algorithms

Hybrid Algorithms - Advantages & Drawbacks

Advantages:

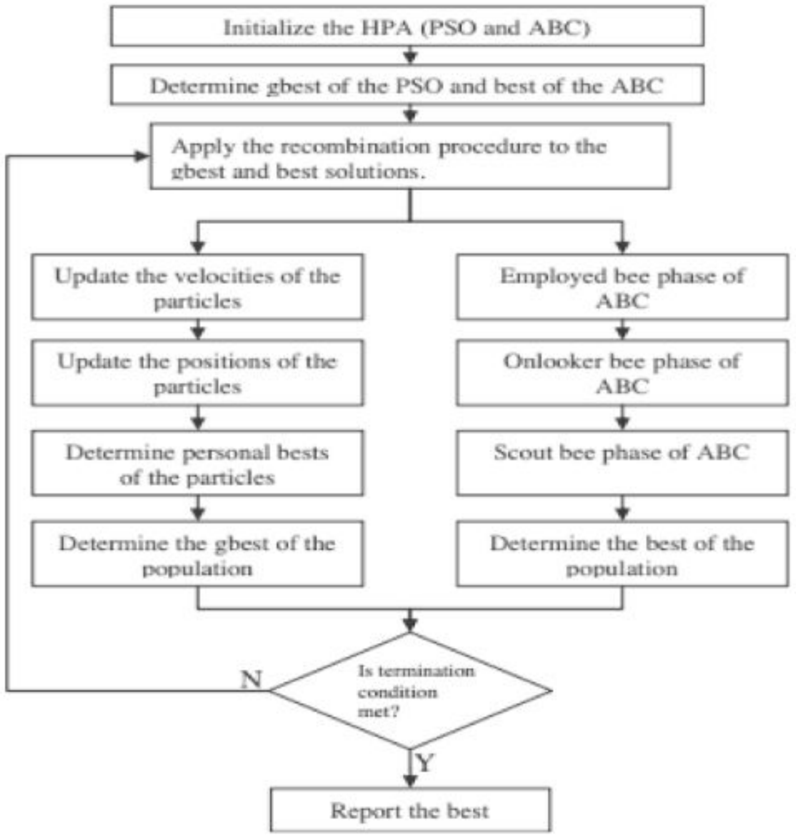
- Allow for multi-objective coverage
- Address deficiencies/trade-offs more effectively than a single algorithm does [10]

Drawbacks:

- increased complexity of implementation/analysis
- Execution time often increases despite convergence on solution being faster [11]

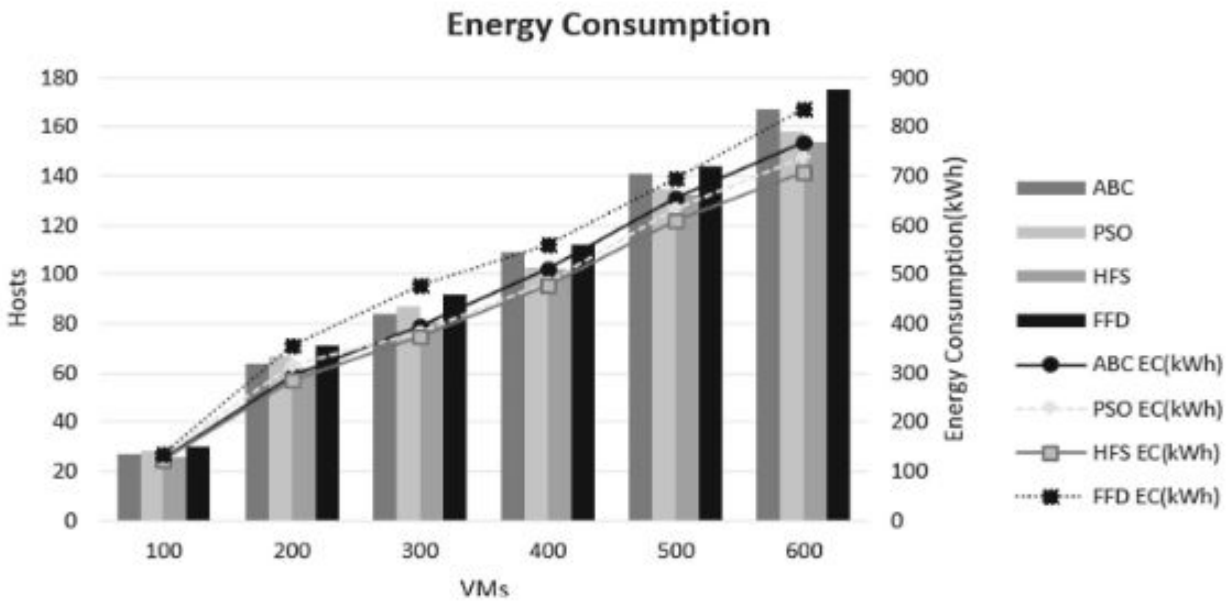
Example - Hybrid Artificial Bee Particle Swarm

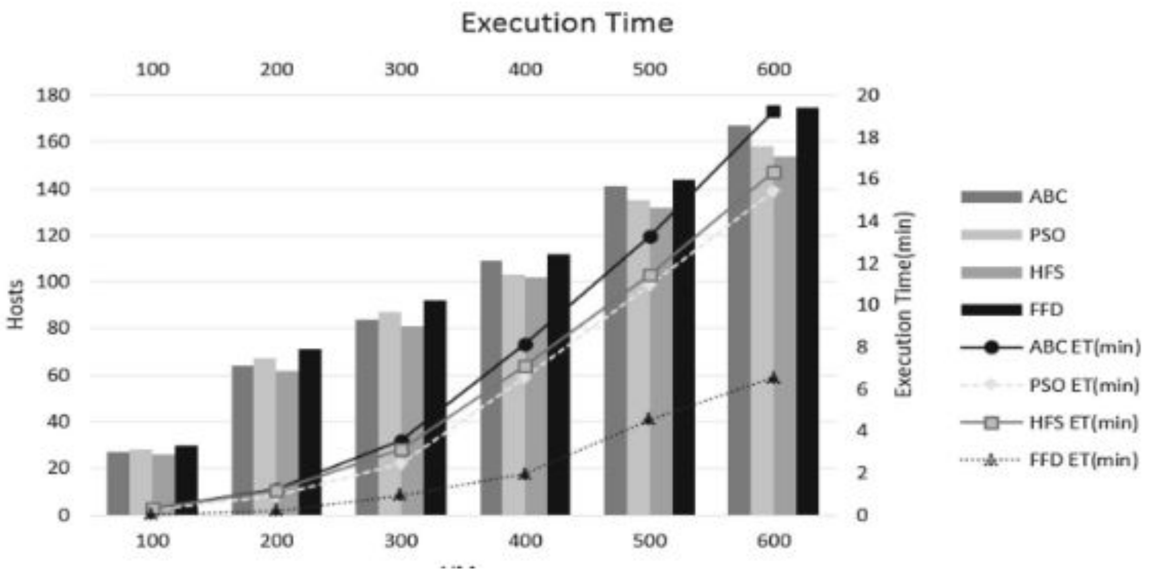
- Combination of Particle Swarm and Artificial Bee Colony algorithms
- Goal is to maintain performance while reducing energy consumption (Load Balancing)
- Combining these two algorithms addresses each of their deficits



[12]

Hybrid ABC-PSO: Evaluation Metrics





[12]

Research Challenges

-Maintaining scalability with growing number of servers and clients

-Dynamic Environments and real time solutions

-Different QoS goals for different clients

-Minimizing energy usage at scale

Hierarchy of Organization

Virtual Machine (VM) - dynamic resources for CPU, memory, bandwidth, etc… multiple per physical machine

Physical Machine (PM) - Static resources for CPU, memory, bandwidth etc…

Cluster - group of physical machines located together in a server/data center.

Network - group of connected clusters as available resources

Scalability

-NP-hard problem: complexity scales exponentially as problem size increases

-VM's per PM
-PM's per Cluster
-Clusters per Network
- Clients

Dynamic Environments

- Resource requests and open resources change in real time

- Solutions to allocation problems must respond in real time

- Predictive rather than reactive

Different QoS goals for different clients

- Scheduling must prioritize different goals for different clients [2].

- Makespan, user-cost, reliability, energy use, resource utilization

Oregon State University
College of Engineering

Energy Consumption at Scale

- Idle servers consume 50% peak power [3]
- Many PM’s with few VM’s to few PM’s with many VM’s
- Sleep Idle servers - overhead to wake up [3]
- Predict server load and prepare environment ahead of time

References

[1] T. A. L.Genez, L. F. Bittencourt, and E.R.M. Madeira. 2012. Workflow scheduling for SaaS/PaaS cloud providers considering two SLA levels. *In Proceedings of the IEEE Network Operations and Management Symposium*. 906–912.

[2] Z.-H. Zhan *et al.*, "Cloud computing resource scheduling and a survey of its evolutionary approaches," *ACM Computing Surveys*, vol. 47, no. 4, pp. 1–33, Jul. 2015. doi:10.1145/2788397

[3] M. Dabbagh, B. Hamdaoui, M. Guizani and A. Rayes, "Energy-Efficient Resource Allocation and Provisioning Framework for Cloud Data Centers," in *IEEE Transactions on Network and Service Management*, vol. 12, no. 3, pp. 377-391, Sept. 2015, doi: 10.1109/TNSM.2015.2436408

[4] What is Three-Tier Architecture | IBM. (n.d.). https://www.ibm.com/topics/three-tier-architecture

[5] Chauhan, N., Kaur, N., Saini, K. S., Verma, S., Alabdulatif, A., Khurma, R. A., Garcia-Arenas, M., & Castillo, P. A. (2024, February 20). A systematic literature review on task allocation and performance management techniques in cloud Data center. arXiv.org. https://arxiv.org/abs/2402.13135

[6] GfG. (2020, May 17). Introduction to ant colony optimization. GeeksforGeeks. https://www.geeksforgeeks.org/introduction-to-ant-colony-optimization/

[8] A. Hameed *et al.*, "A survey and taxonomy on energy efficient resource allocation techniques for cloud computing systems," *Computing*, vol. 98, no. 7, pp. 751–774, Jun. 2014, doi: https://doi.org/10.1007/s00607-014-0407-8

[9] S. K. Mishra *et al.*, "Energy-Aware Task Allocation for Multi-Cloud Networks," *IEEE Access*, vol. 8, pp. 178825–178834, Dec. 2020, doi: https://doi.org/10.1109/access.2020.3026875

[10] R. M Singh *et al.,* "Towards Metaheuristic Scheduling Techniques in Cloud and Fog: An Extensive Taxonomic Review" in *ACM Computing Surveys,* vol. 55, no. 3, pp. 1-43, Feb. 2022. [Online] https://doi.org/10.1145/3494520

[11] X. Yang, Ed. "Hybrid Metaheuristic Algorithms: Past, Present, and Future" in *Recent Advances in Swarm Intelligence and Evolutionary Computation,* Cham, Switzerland: Springer, 2015, ch. 4, pp 71-83. [Online] https://doi.org/10.1007/978-3-319-13826-8_4

[12] J. Meshkati, F. Safi-Esfahani, "Energy-aware resource utilization based on particle swarm optimization and artificial bee colony algorithms in cloud computing" in *The Journal of Supercomputing,* vol. 75, pp. 2455-2496, May 2019 [Online] doi: https://doi.org/10.1007/s11227-018-2626-9

Oregon State University
College of Engineering

Questions?

Contributions

Kira - Evaluation Metrics, Hybrid Algorithms
Morel - Definition of cloud computing, motivations, heuristic algorithms
Rand - Cloud service architecture, topology, and categories, Meta-heuristic algorithms
Sean - Open research areas and challenges