Hybrid Grouping Genetic Algorithm for Large-Scale Two-Level Resource Allocation of Containers in Cloud

Taiwo Akindele, Boxiong Tan, Yi Mei & Hui Ma   Victoria University of Wellington, New Zealand

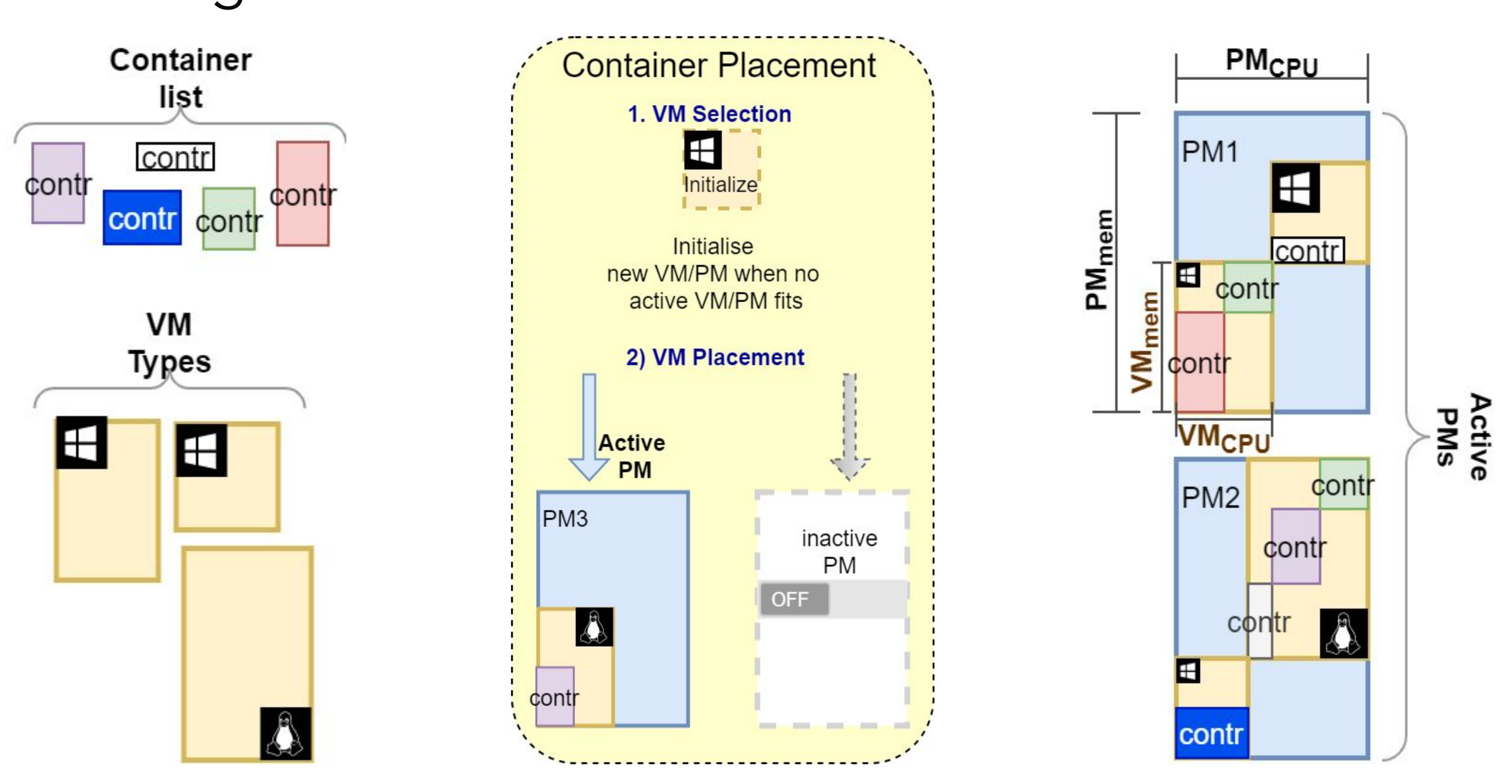
Background

- Carbon footprint of data centers are a concern at ~1% of all electricity use and projected to grow to 2-3% by 2030 [1, 2]
- Over 40% of total energy consumption of data centers is consumed by its servers (i.e physical machines)
- Data center servers typically run at less than 50% CPU utilisation on average (often range as low as 15 to 40% on average)
- Overall energy consumption of cloud data centers can be minimised by optimising server resource allocation and utilization, so that less machines are active in a cloud data center
- Two-level resource allocation is required for the increasingly popular two-level virtualization approach combining cloud container OS virtualization with virtual machines (VM) to deploy applications on shared datacenter physical machines (PM)

Problem

Given:

- A large batch of containers, each deploying a cloud task or application defined by their CPU and memory requirements, to be initially allocated resources in the cloud
- A set of fixed VM types each defined by CPU and memory capacity
- Homogenous PMs in the cloud

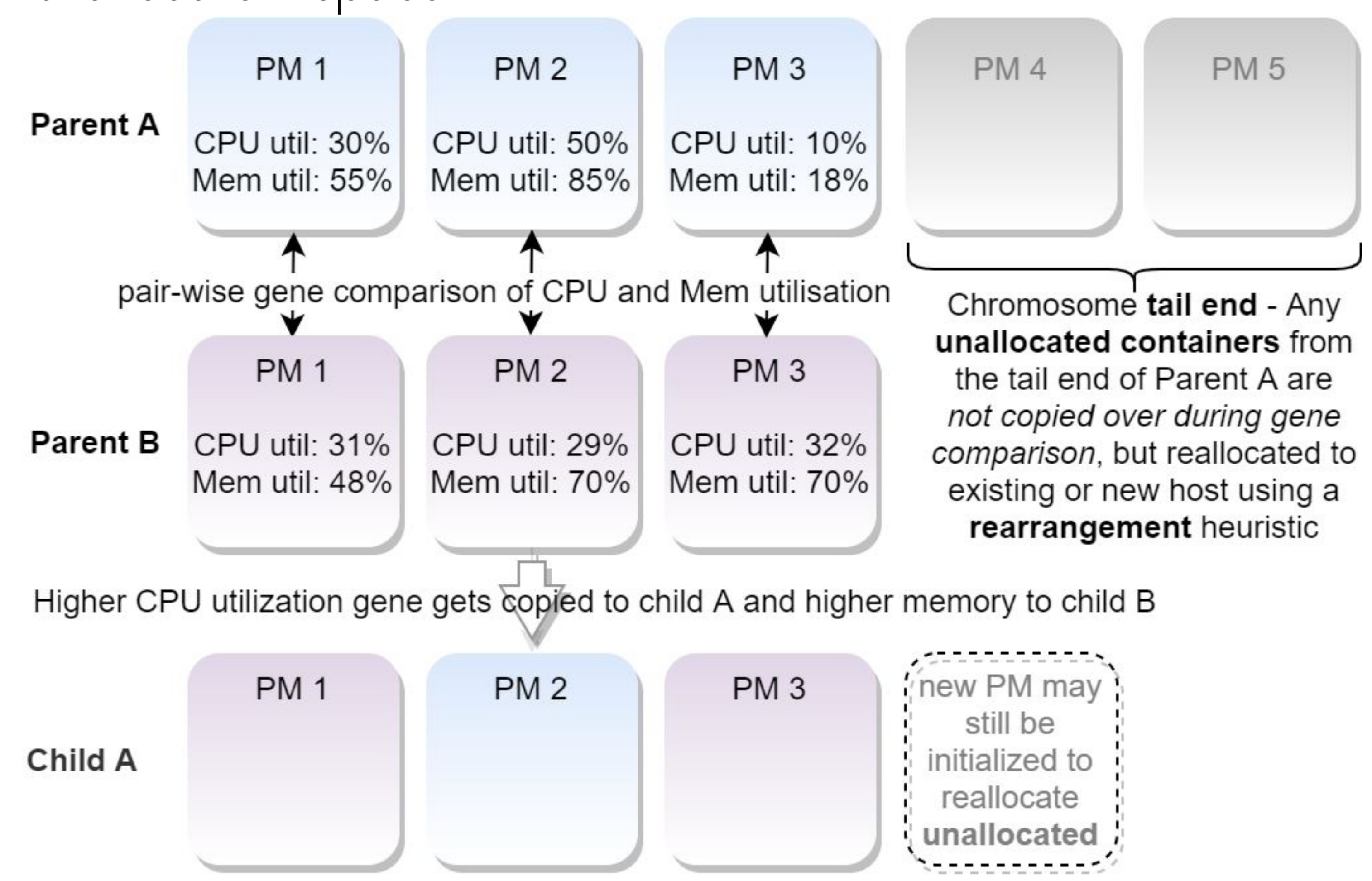


Objective: Find two-level resource allocation for containers and VMs, minimising the energy consumption of total utilized PMs in the allocation.

Method

We propose a hybrid Grouping Genetic Algorithm (GGA) [4] with a new crossover operator and mutation operators enhanced with *problem specific heuristics*.

- *Fixed Length crossover (FLX)*: We introduce a new crossover operator to promote diversity in the population and exploration of the search space



- *UnpackBF* mutation: unpack a PM chosen by roulette wheel with increased probability when PM utilization is poor, and rearrange it's containers using *Best Fit(BF) VM/PM heuristics*
- *MergeLV* mutation: 2 smallest VMs in a PM are merged or replaced with the largest VM to fit (*LV heuristic*)

Experiment and Results

Compared proposed algorithms with benchmark GGA framework [3]:

- FLXGGA is a GGA with proposed fixed-length crossover, while FLXGGA-BF/LV has additional Best Fit and Large VM heuristics to improve performance for large scale problems
- Had 8 test cases of differently sized problems, 40 runs of each algorithm, and all algorithms used the same GA settings

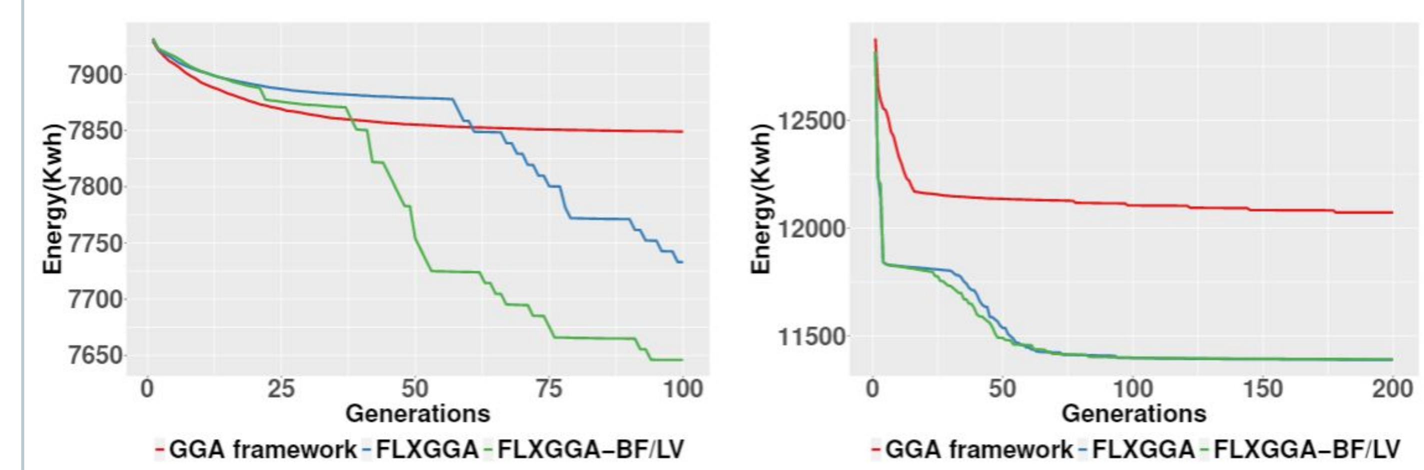
The following table compares the resulting mean energy consumption in (Kwh) and standard deviation:

- Proposed algorithms outperformed GGA framework [3] with significantly lower energy consumption in 5 out of 8 test instances

	Size	GGA[3]	FLXGGA		FLXGGA-BF/LV		
Real world VM types	500	3844.8 ± 1.70	3843.9 ± 1.96	+	3850.3 ± 5.76	-	-
	1000	8035.8 ± 186.40	7838.7 ± 59.37	+	7817.7 ± 115.76	+	=
	1500	12072.3 ± 124.30	11388.6 ± 9.26	+	11391.7 ± 84.43	+	=
	5000	37850.5 ± 378.93	36587.5 ± 219.65	+	35935.9 ± 142.02	+	+
Synthetic VM types	500	3851.6 ± 1.03	3860.1 ± 2.60	-	3864.8 ± 3.75	-	-
	1000	7849.1 ± 3.19	7733.0 ± 183.29	=	7646.0 ± 185.16	=	+
	1500	11423.0 ± 136.02	11378.9 ± 7.78	+	11401.2 ± 11.85	+	-
	5000	35982.2 ± 276.3	36058.2 ± 11.14	=	35735.37 ± 161.38	+	+

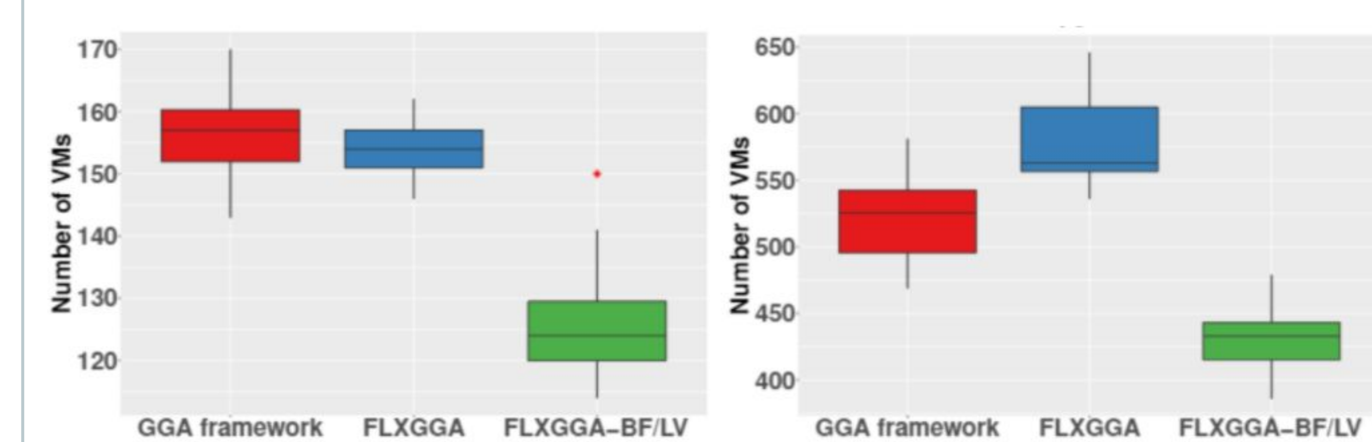
Results

- FLXGGA-BF/LV is the preferred algorithm for large test cases of 1000 containers or more.



(a) Synthetic VM types, 1000 containers   (b) Realworld VM types, 1500 containers

- Convergence curves above illustrate how proposed GGA are able to escape local optima where GGA framework can not



(a) Real World VM types, 1500 containers   (b) Real World VM types, 5000 containers

- FLXGGA-BF/LV algorithm with Large VM heuristic is effective in finding solutions using significantly fewer VMs. This amounts to less wasted resources on VM overhead and better energy consumption.

Summary

Proposed FLXGGA and FLXGGA-BFLV algorithm demonstrate improved exploratory as well as exploitative capabilities than our baseline GGA framework, and either of them was able to find solutions with significantly better energy consumption in 5 out of 8 test cases, including large-scale test instances.

References

1. Koot, M., and Wijnhoven, F. Usage impact on data center electricity needs: A system dynamic forecasting model. Applied Energy 291 (2021), 116798
2. MASANET, E., SHEHABI, A., LEI, N., SMITH, S., AND KOOMEY, J. Recalibrating global data center energy-use estimates. Science 367, 6481 (2020), 984–986
3. Tan, B., Ma, H., Mei, Y.: A group genetic algorithm for resource allocation incontainer-based clouds. In: Paquete, L., Zarges, C. (eds.) Evolutionary Computa-tion in Combinatorial Optimization. pp. 180–196 (2020)
4. Falkenauer, E.: A hybrid grouping genetic algorithm for bin packing. Journal ofheuristics2(1), 5–30 (1996)

taiwo.akindele@ecs.vuw.ac.nz