Parareal Algorithm in Cloud Environment

Hongtao Xiao, Prof. Eric Aubanel. Faculty of Computer Science

Contribution

We investigate how to adjust the parareal algorithm in the cloud environment to solve time-dependent partial differential equations (PDEs). The master-worker paradigm is modified for work load allocation according to the different speeds of virtual cores on Amazon EC2.
Our approach is proved to be feasible; it saves overall time of the PDE solver.

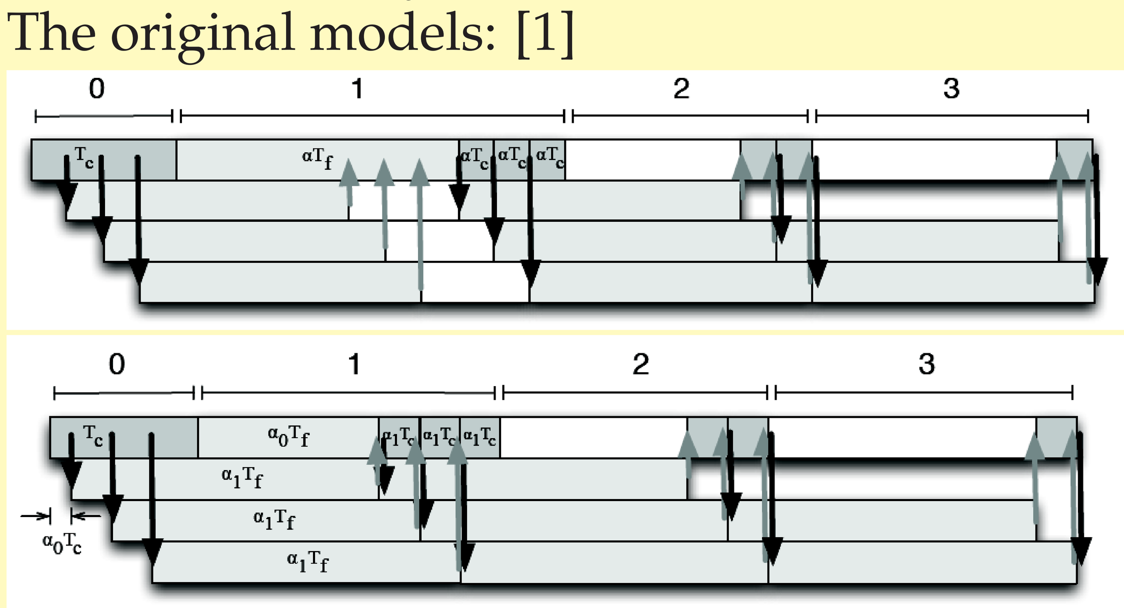
Objectives

Our research has following objectives:

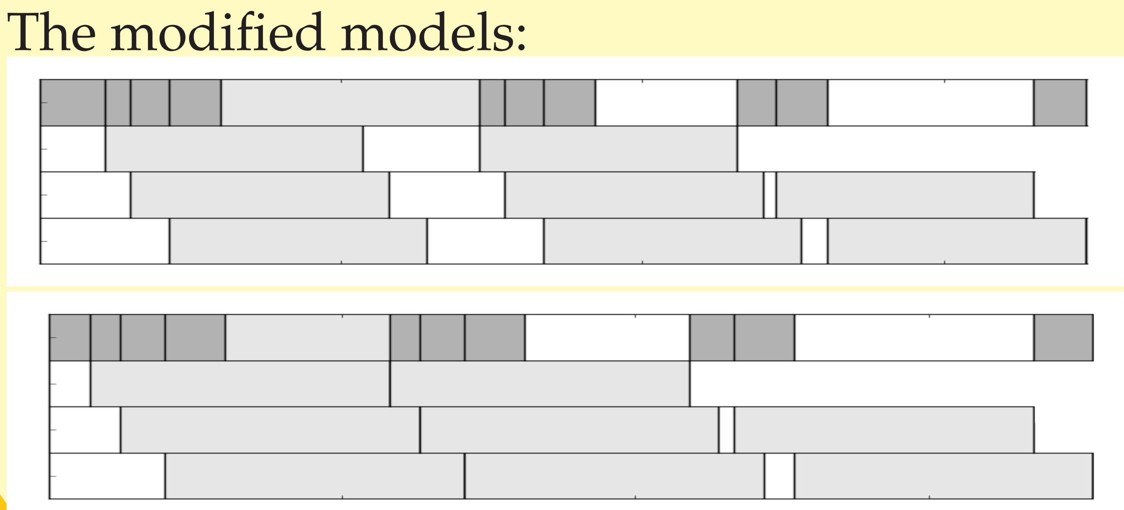
1. Adjust the parareal algorithm for heterogeneous platforms (i.e. cloud environments). This involves a manager-worker-based task schedule algorithm in heterogeneous environments.
2. Demonstrate the feasibility of the parareal algorithm in cloud environments using the example of the two-dimensional heat equation.
3. For the parareal algorithm for the two-dimensional heat equation, investigate performance characteristics and variations on Amazon EC2 due to different instance types (and their combinations), different locations (or availability zones), different networks, or other factors. We expect this would further demonstrate the feasibility of the parareal algorithm in heterogeneous environments.
4. Compare performance, speedup, and efficiency of a cloud with that of a computing cluster taking account of differences in hardware specifications (e.g. CPUs).

Methodology

The models we use are modified from [1]. The original models assume homogeneous CPU speed and no communication overhead.
For cloud environment, the first assumption does not hold any more.
The original models: [1]



The modified models:



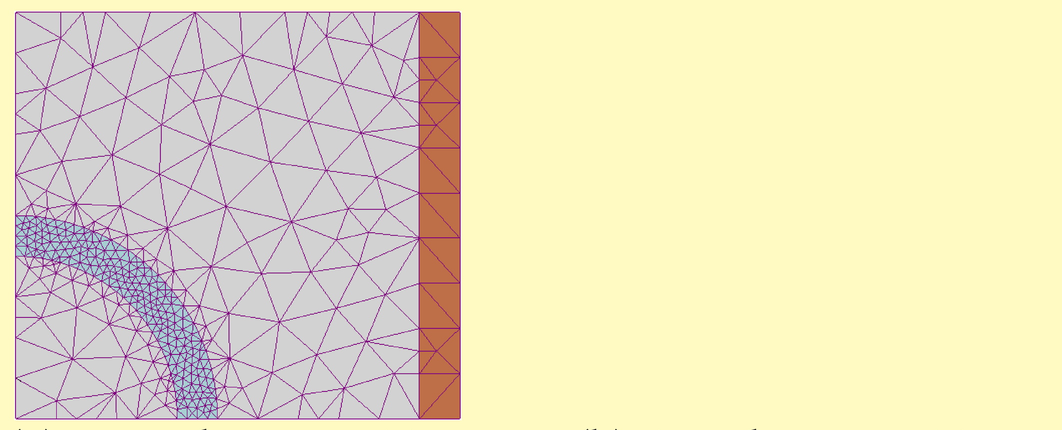
References

[1] E. Aubanel, "Scheduling of tasks in the parareal algorithm," in press, Parallel Computing, 2010.

[2] J. J. Rehr, F. D. Vila, J. P. Gardner, L. Svec, and M. Prange, "Scientific computing in the cloud," *Computing in Science & Engineering*, vol. 12, no. 3, pp. 34–43, 2010.

Parareal algorithm

Parareal algorithm is a less-known algorithm for solving time-dependent PDEs. Unlike conventional numerical algorithms, parareal decomposes the time dimension.



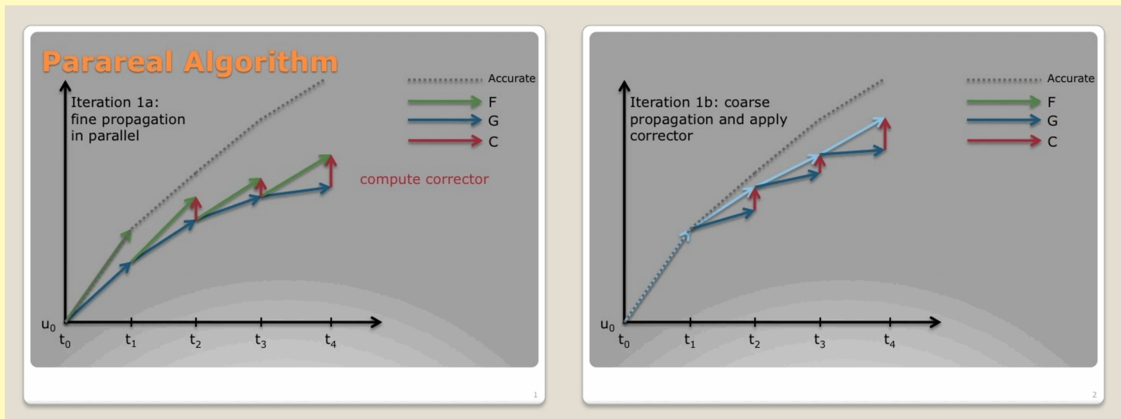
(a) space decomposition (b) time decomposition

(source: Wikipedia, [1])

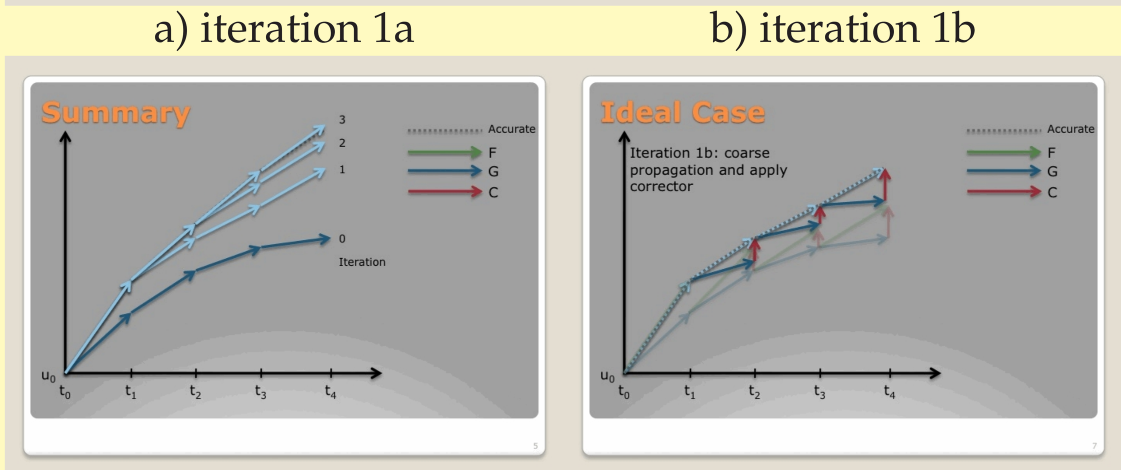
The benefits with time decomposition (vs. space dimension) include:

- further decomposition for long time evolution PDEs
- low communication between iterations

Parareal algorithm uses a prediction - correction format. A faster coarse solver is used as the predictor. A slower fine solver is used as the corrector. The whole process is as follows: [1]



a) iteration 1a b) iteration 1b



c) summary d) ideal case (iter 1b)

Run time results

Our experiments consist of: (1) sequential solver on a single machine; (2) parallel solver with original master-worker paradigm; (3) parallel solver with modified master-worker paradigms.
The experiment results of parallel solver on a heterogeneous set of machines are in the following table.

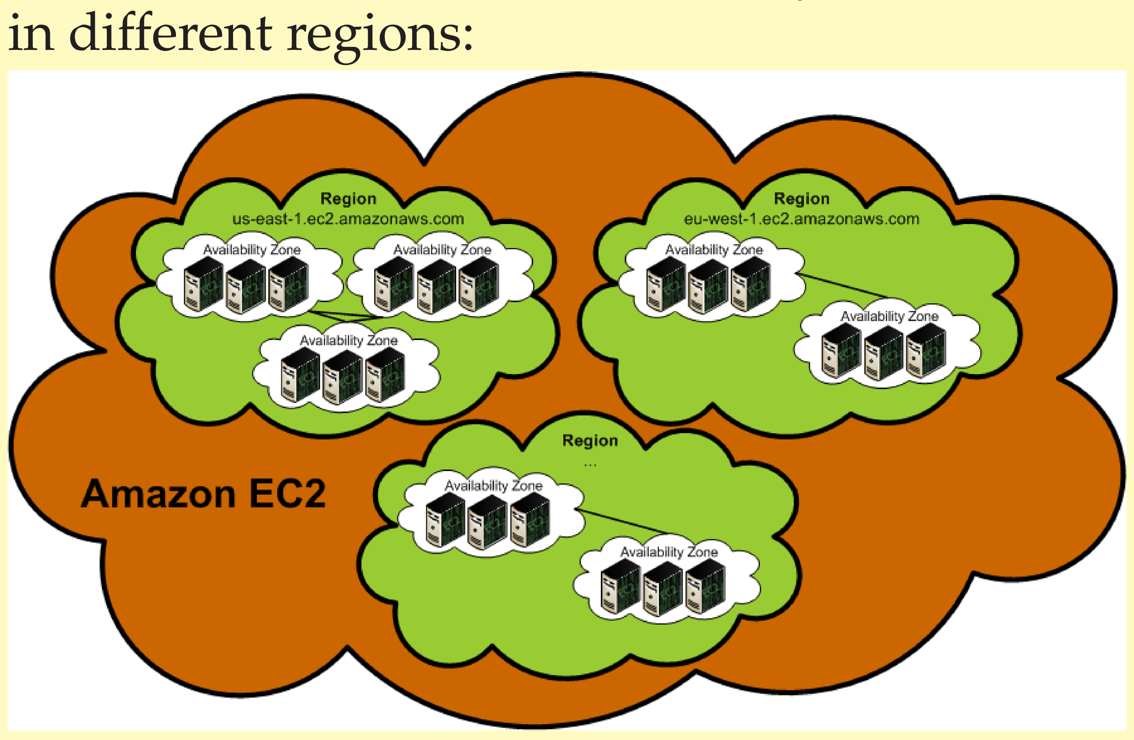
N	k	Equal Overlap			Impr. Overlap		
		exp.	orig.	mod.	exp.	orig.	mod.
4	1	5.67	4.32	5.51	5.47	3.96	5.42
	2	10.53	7.76	10.24	10.38	7.37	10.35
	3	15.43	11.21	15.03	15.36	10.89	15.34
8	1	6.59	5.18	6.37	6.13	4.41	5.78
	2	11.53	8.61	11.11	11.25	7.83	10.62
	3	16.55	12.05	15.90	16.49	11.36	15.53
16	1	8.49	6.93	8.06	7.06	5.26	6.61
	2	13.47	10.37	12.73	12.19	8.67	11.42
	3	18.51	13.80	17.47	17.40	12.20	16.30
32	1	12.23	10.39	11.56	9.05	7.07	8.39
	2	17.16	13.81	16.27	14.13	10.49	13.24
	3	22.10	17.23	21.04	19.27	14.03	18.15

Cloud environment

Currently cloud service is provided mainly in following forms:

- SaaS: Google Docs
- PaaS: Facebook
- IaaS: **Amazon EC2**, Microsoft Azure

Amazon EC2 provides IaaS and divides the servers to different availability zones in different regions:



(source: Amazon)

The benefits of cloud (vs. clusters) include:

- financially beneficial
- readily accessible
- user-friendly interfaces
- "zero" downtime

Cloud is not suitable for typical High Performance Computing (HPC) applications. The scaling in cloud environments decreases dramatically due to high communication overhead often found in HPC and poor networking among cloud processors. [2]

However, parareal algorithm is ideal for Cloud environment due to its low communications.

Latency (non)effects

One assumption for the modified model is that the latency can be ignored and the communication overhead is negligible. Our experiments record the run time when the master is always in one availability zone (us-east-1c), and the workers in different availability zones. Following table shows that the run time is tolerant to latencies in different availability zones. The modified model (mod') is used in the experiments.

worker azone & latency (in μs)	k	Equal Exp.	Impr. Exp.
us-east-1b	1	6.98	6.79
853.04	2	13.16	13.00
	3	19.45	19.31
	4	25.74	25.70
us-east-1c	1	6.65	6.83
162.65	2	12.65	13.11
	3	18.77	19.47
	4	24.83	25.93
us-east-1d	1	7.07	7.22
418.89	2	13.55	13.91
	3	20.15	20.71
	4	26.79	27.51

Funding

This work by supported by NSERC (Natural Sciences and Engineering Research Council of Canada).