Grid-Enabled Hydropad:
a Scientific Application for Benchmarking GridRPC-Based Programming Systems

Michele Guidolin* Alexey Lastovetsky†

Heterogeneous Computing Laboratory
School of Computer Science and Informatics
University College Dublin (UCD)

*michele.guidolin@ucd.ie
†alexey.lastovetsky@ucd.ie

Introduction

Grid Computing is widely used for scientific applications

* GridRPC is a standard API promoted by Open Grid Forum
* Several Grid middleware are GridRPC compliant

What are the potential performance of GridRPC based systems?

Introduction

Grid Computing is widely used for scientific applications

* GridRPC is a standard API promoted by Open Grid Forum
* Several Grid middleware are GridRPC compliant

What are the potential performance of GridRPC based systems?

Different GridRPC systems are tested with artificially engineered applications

- ✔ tasks with high computation
- ✔ execution of many parallel tasks
- ✔ minimal data communication between tasks

Results that they have an equally good performance

Introduction

Grid Computing is widely used for scientific applications

* GridRPC is a standard API promoted by Open Grid Forum
* Several Grid middleware are GridRPC compliant

What are the potential performance of GridRPC based systems?

Different GridRPC systems are tested with artificially engineered applications

- ✔ tasks with high computation
- ✔ execution of many parallel tasks
- ✔ minimal data communication between tasks

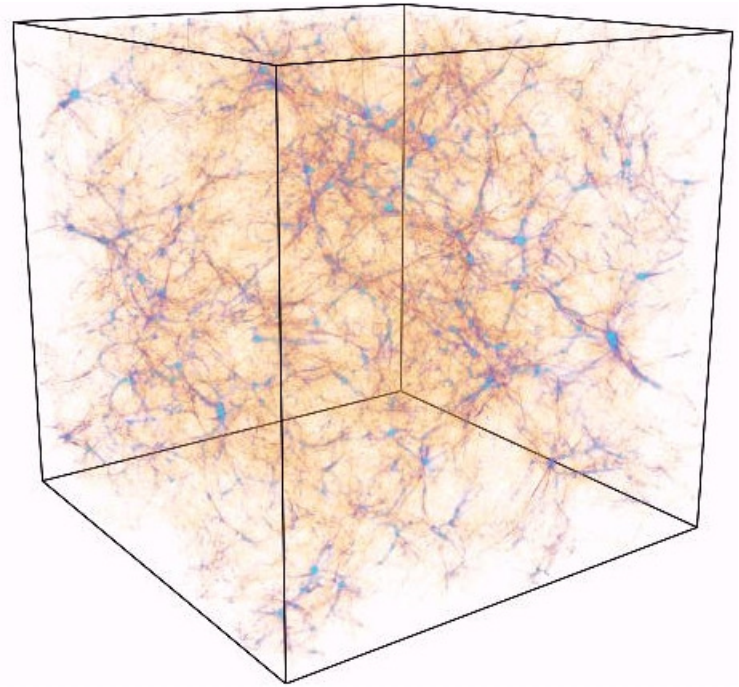
Results that they have an equally good performance

We need a real-life, non Grid friendly, application as a benchmark!

- ✔ To show the eventual limits and benefits of a GridRPC middleware
- ✔ To show that Grid computing can be used to speed-up a wide range of applications

Hydropad

- ✔ An astrophysical application that simulates the clustering of galaxies from the big bang till present.



It is based on the Lambda-CDM model which assumes that the universe is composed of two matters:

- ✔ Baryonic matter (all bright objects)
- ✔ Dark matter (gravitational mass)

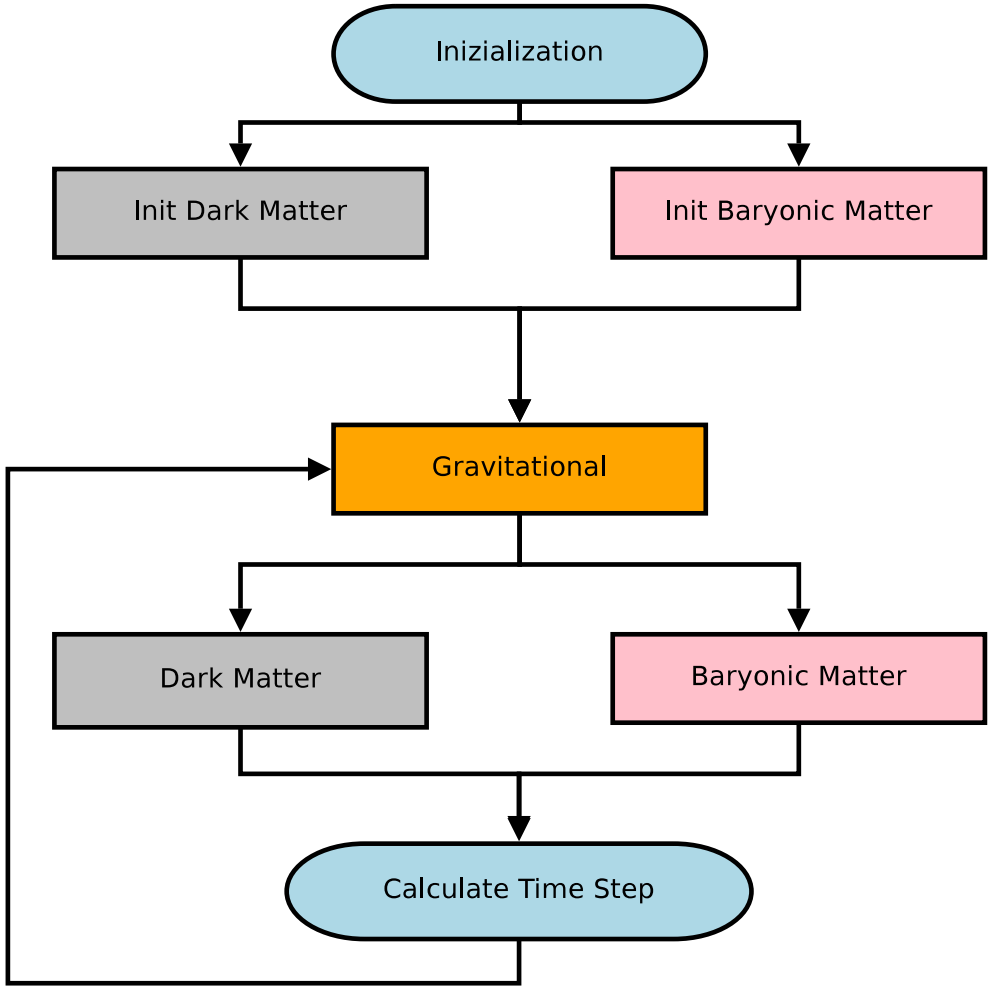
Hydropad is not an ideal application to get top performance in a Grid environment

- ✔ relatively low complexity of its tasks (log-linear at maximum)
- ✔ large amount of input and output data moved between tasks.

Hydropad Internal Structure

It consists of four parts

- Initialisation
- Gravitational **(FFT)**
- Dark Matter **(N -Body)**
- Baryonic Matter **(PPM)**



Characteristics

- ✔ $C_{bm} \gg C_{dm}$
- ✔ $D_{bm} \ll D_{dm}$.

Experiments Motivation

We compare the three versions of Hydropad:

1. Local sequential version
   (Comparison base)
2. GridSolve version
3. SmartGridSolve version
   (Our own extensions of GridRPC model)

Experiments Motivation

We compare the three versions of Hydropad:

1. Local sequential version
   (Comparison base)
2. GridSolve version
3. SmartGridSolve version
   (Our own extensions of GridRPC model)

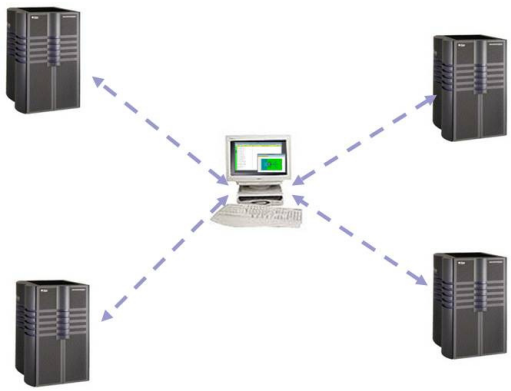
We analyse how Grid-enabled versions of Hydropad can improve the performance

* Computation
* Communication
* Memory footprint

GridSolve vs. SmartGridSolve

GridSolve:

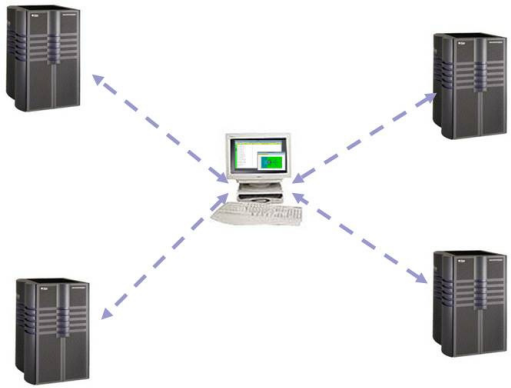
- ✔ Maps tasks individually
- ✔ Uses client-server/ star network data communication



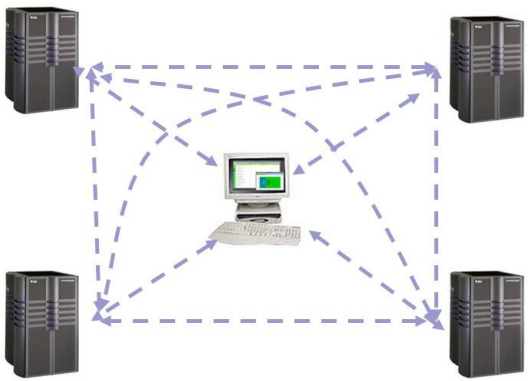
GridSolve vs. SmartGridSolve

GridSolve:

- ✔ Maps tasks individually
- ✔ Uses client-server/ star network data communication



SmartGridSolve: a new middleware that extends the execution model of GridRPC to overcome its limitations



- ✔ Maps a group of tasks
- ✔ Uses fully connected network data communication

Experimental Setup

The hardware configuration:

- Client machine with 256MB of memory and 248MFlops
- Two heterogeneous servers with 1GB of memory
  498 and 531 MFlops respectively
- The server-to-server link speed is 1GB/s
- Two client-to-server link setups: 100MB/s and 1GB/s

The initial parameters and data sizes for the problems:

Problem ID	N_p	N_g	Data Size
P1	120^3	60^3	73MB
P2	140^3	80^3	142MB
P3	160^3	80^3	176MB
P4	140^3	100^3	242MB
P5	160^3	100^3	270MB
P6	180^3	100^3	313MB
P7	200^3	100^3	340MB
P8	220^3	120^3	552MB
P9	240^3	120^3	624MB

GridSolve Results Table

(a) Local

P.ID	Time Step	Paging
P1	14.09s	No
P2	29.95s	No
P3	35.29s	No
P4	55.13s	Light
P5	61.63s	Light
P6	83.66s	Yes
P7	128.55s	Yes
P8	227.89s	Heavy
P9	280.07s	Heavy

(b) GridSolve C - 1GB/s

Time Step	Paging	S_p v Loc
7.20s	No	1.96
15.51s	No	1.93
16.48s	No	2.14
29.11s	No	2.14
29.07s	No	2.12
36.74s	Light	2.28
48.06s	Yes	2.67
77.91s	Heavy	2.92
91.75s	Heavy	3.06

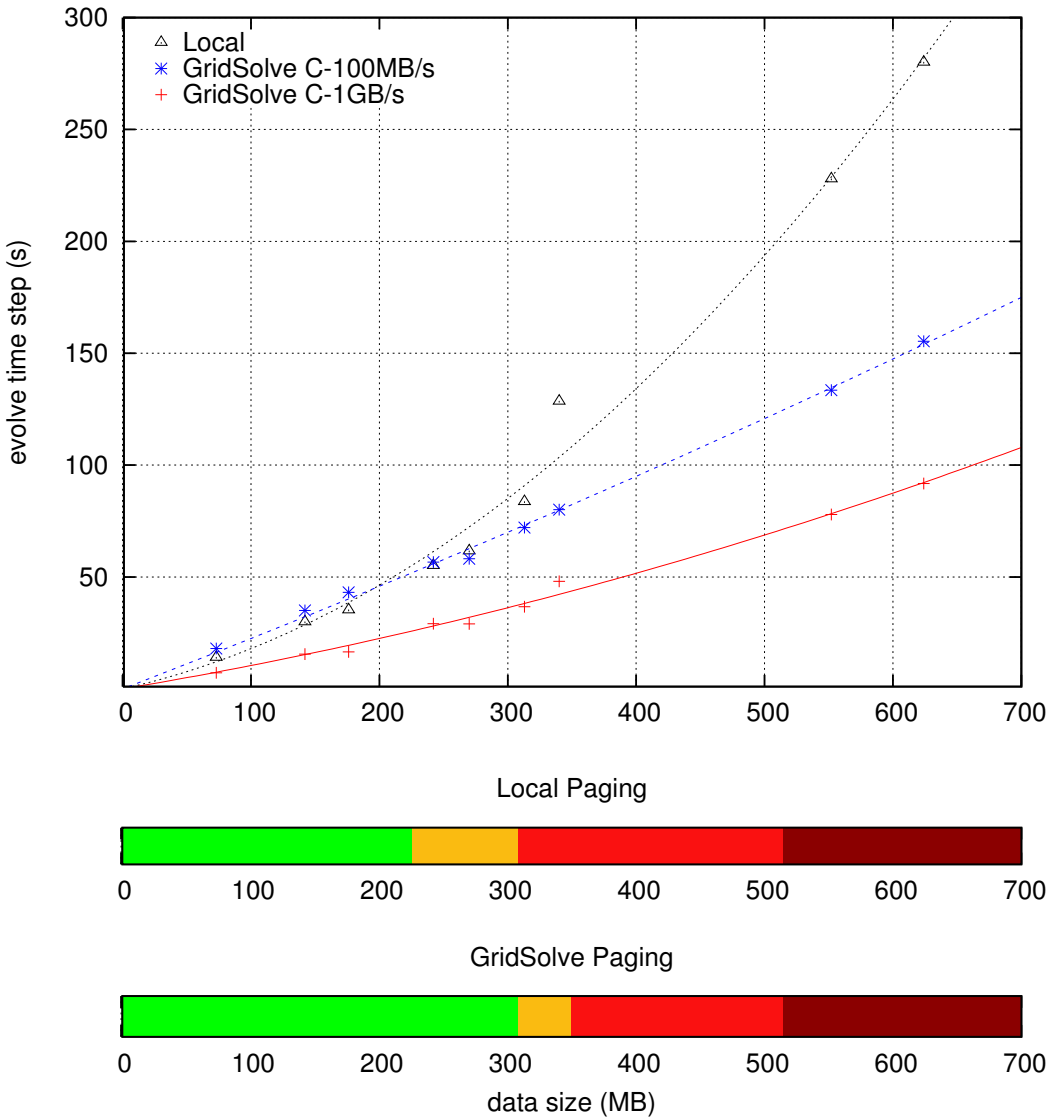
(c) GridSolve C - 100MB/s

Time Step	Paging	S_p v Loc
18.01s	No	0.78
35.02s	No	0.86
43.09s	No	0.82
55.66s	No	0.97
58.17s	No	1.06
72.50s	Light	1.15
80.05s	Yes	1.61
133.47s	Heavy	1.71
155.36s	Heavy	1.81

Sources of speed-up:

- ✔ Parallel tasks
- ✔ Powerful servers
- ✔ Less paging on the client machine
  - ✘ Save the temporary data allocation on remote tasks
  - ✘ Paging only during communication, not computation

GridSolve Results Figure



SmartGridSolve Results Table

(d) SmartGridSolve C - 1GB/s

P.ID	Time Step	Paging	S_p v Loc	S_p v GS
P1	6.99s	No	2.02	1.03
P2	14.69s	No	2.04	1.06
P3	15.52s	No	2.27	1.06
P4	27.22s	No	2.03	1.07
P5	27.13s	No	2.27	1.07
P6	27.22s	No	3.07	1.35
P7	29.13s	Light	4.41	1.65
P8	49.21s	Light	4.63	1.58
P9	50.82s	Light	5.52	1.81

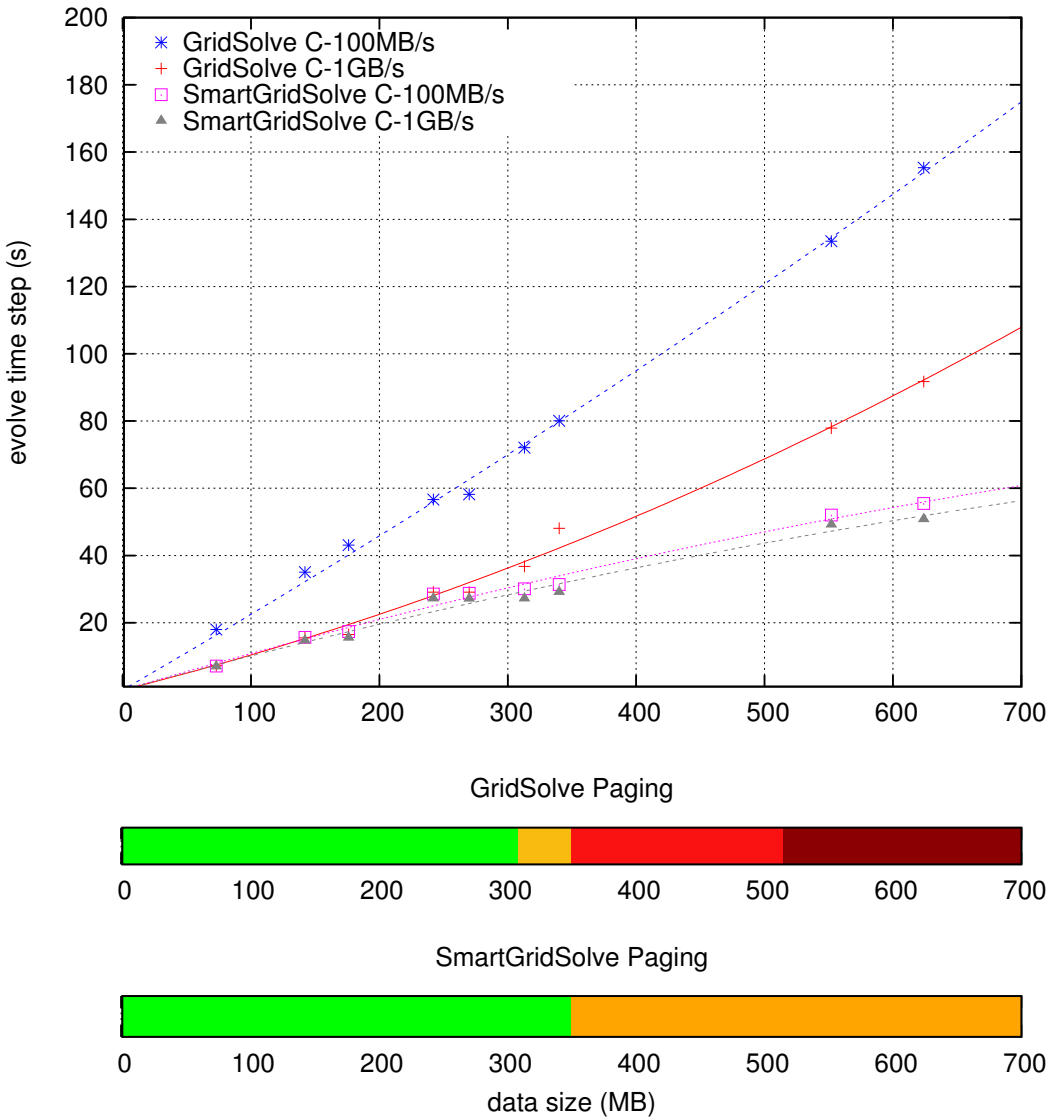
(e) SmartGridSolve C - 100MB/s

Time Step	Paging	S_p v Loc	S_p v GS
7.9s	No	1.78	2.28
15.68s	No	1.91	2.75
17.36s	No	2.03	2.48
28.56s	No	1.93	1.98
28.77s	No	2.14	2.02
30.09s	No	2.78	2.41
31.63s	Light	4.06	2.53
52.30s	Light	4.36	2.55
55.47s	Light	5.06	2.80

Sources of speed-up:

- ✔ Better mapping (group of tasks)
- ✔ Minor paging on the client machine
  - ✘ The direct server-to-server communication minimises the data on the client
  - ✘ Paging only at beginning and at end of the group of tasks
- ✔ Better communication model

SmartGridSolve Results Figure



Conclusion

Grid-enabled Hydropad is a freely available application: hcl.ucd.ie

✔ It represents a good benchmark for GridRPC-based systems

- ✘ it exemplifies typical real-life scientific applications
- ✘ it push the limits of the GridRPC model
- ✘ highlights the strengths and exposes the weaknesses of a particular middleware implementation

The experiments show:

* GridSolve version of Hydropad
  - slower for smaller problem sizes
  - faster for larger problem sizes
* SmartGridSolve version of Hydropad
  - achieves better performance than the other two versions:
    * better mapping
    * better communication model
    * less paging

Questions?