

Metrics for Grid Applicability

A Distributed Elliptic Curve Platform Assessment

ALGOS

Paulo Trezentos

Paulo.Trezentos@iscte.pt

UNIDE / ISCTE



Arlindo Oliveira

aml@inesc-id.pt

INESC-ID / IST

Parallel Processing and Applied Mathematics - PPAM 2003
Czestochowa, Poland, Sept. 7-10, 2003.

Contents

- Introduction
- Critical factors to Grid performance
- Aggregated metrics
- Drawing the Goodness Frontier
- DISTECM assessment
- DISTECM deployment & confirmation
- Future Work
- Conclusions

Introduction

The context:

- Load distribution is often needed in HPC
- Bad decisions cost time and resources
- Grid computing introduced new variables in deployment decision

This talk introduce:

- Auxiliary indicators for measuring the potential applicability of a parallel application to a Grid
- Used by our Group to assess the applicability of an ECM implementation to a Grid

Scenarios

Not all parallel implementations are adequate to all Grids.

We may have:

- Parallel application developed and build a (part of a) Grid for it
- The Grid and implement a parallel solution suitable

We were in the first case.

Critical factors

- **Minimum bandwidth required**

What is the minimum bandwidth between nodes that application requires

- **Maximum latency allowed**

What is the maximum latency / delay between nodes that application supports

- **Data transmission frequency**

How often there is a need to transmit data between nodes;

- **Amount of data to transmit**

Which amount of data is transmitted in each communication.

Aggregated metrics

Frequency of Interruptions for Communication (FIC)

$$\frac{\text{Node processing capacity (FLOPS)} * 60}{\text{Total application floating point instructions}}$$

Total Time of Data Transmission (TTDT)

$$\frac{\text{Total data to be transmitted (KB)} * 8}{\text{Bandwidth of the connection (Kbps)}}$$

Aggregated metrics II

Maximum Communication Overhead (MCO)

$$\begin{aligned} &= \frac{\text{TTDT (seconds)}}{\frac{1}{FIC} * 60} * 100 \\ &= \frac{\text{TTDT (seconds)} * FIC}{60} * 1000 \end{aligned}$$

Maximum Latency Overhead (MLO)

$$= \frac{\text{Latency/delay}}{\text{Overall time of data transfer}} * 100$$

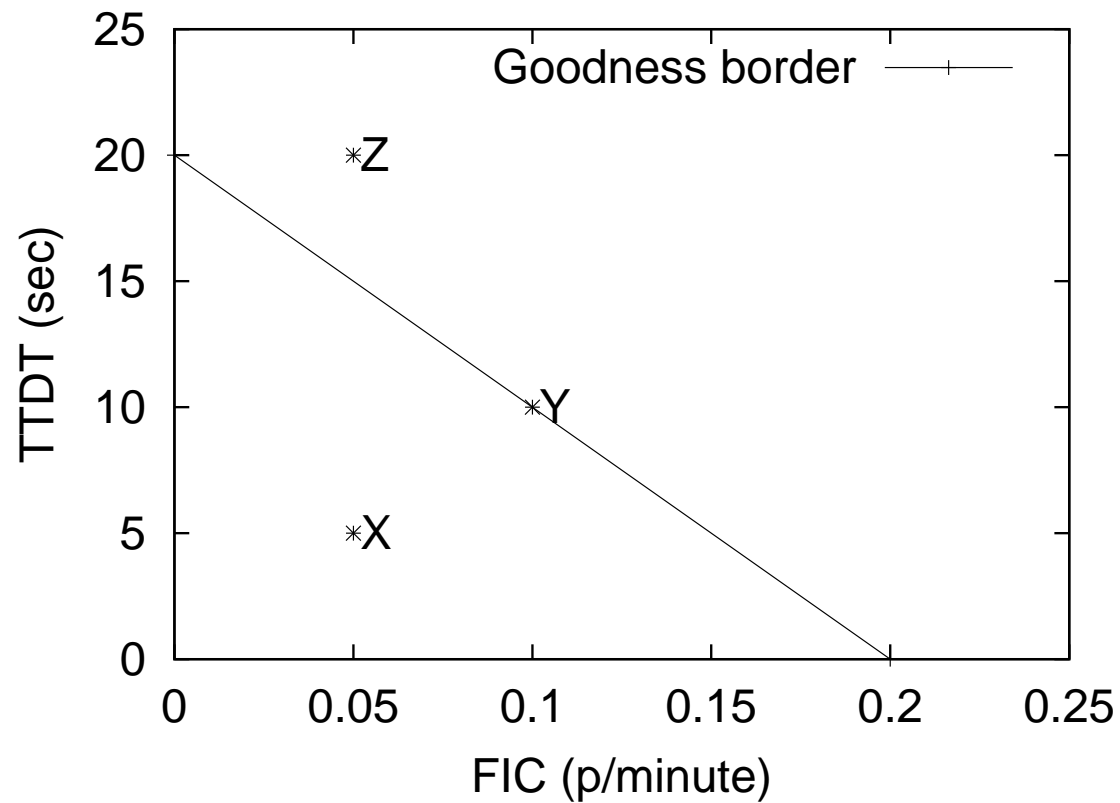
Goodness frontier

- **A** - maximum limit to TTDT above which it does not pay to use the grid, even for small values of FIC. To define its value we can take in account the latency through the MLO ratio;
- **M** - slope of the *goodness* straight line

The slope can be found through:

$$m = \frac{A}{FIC} - \frac{TTDT}{FIC}$$

Goodness application ratio



ECM over the grid (DISTECM)

An Elliptic Curve over a field K is the set of solutions that respect:

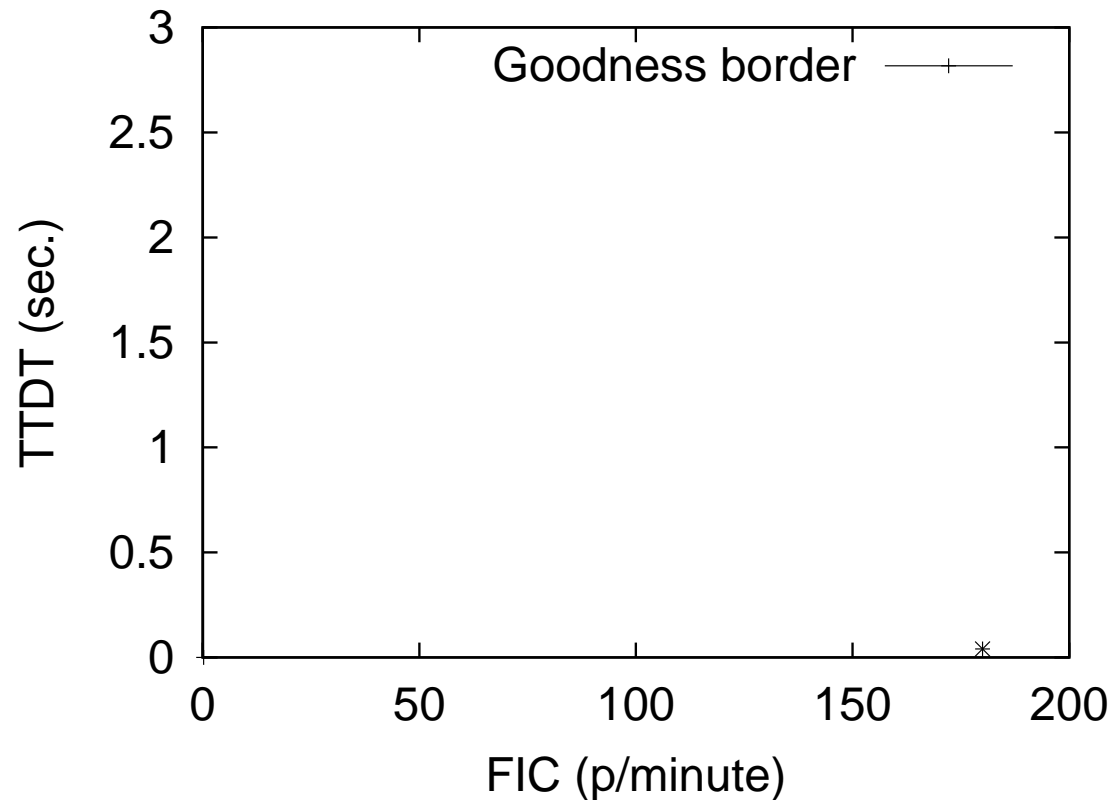
$$Y^2 = X^3 + AX^2 + BX + C \quad (1)$$

DISTECM:

- Core ECM functions provided by Paul Zimmermann ECM code
- Distributed version based:
 - Master / slave distribution paradigm
 - Globus toolkit 2
 - BSD sockets (and not PVM / MPICH)

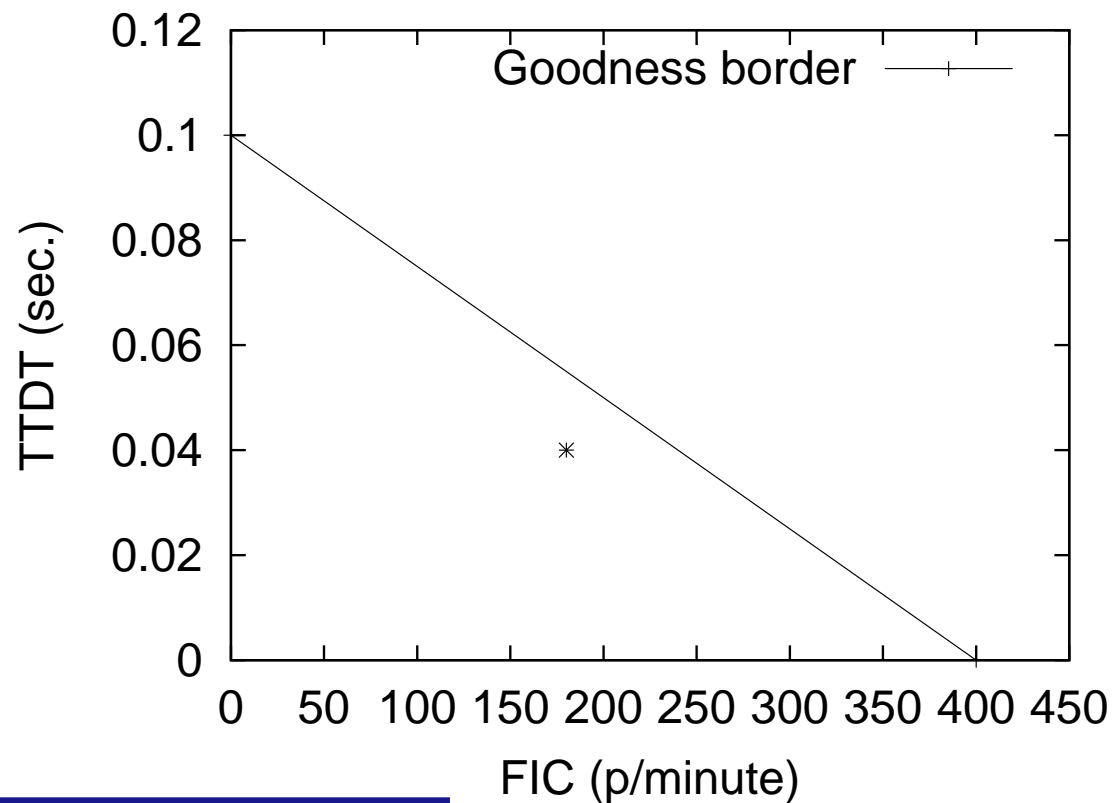
DISTECM Assessment (grid 1 - Low-end grid)

- Composed of old desktops (Pentium III 750 Mhz)
- Slow connections (ISDN 64 Kbps)



DISTECM Assessment (grid 2 - Beowulf Cluster)

- Cluster of PCs built with commodity components and off-the-shelve hardware solutions
- Connected with a 100Mbps FastEthernet



DISTECM deployment

Hardware specification:

- Cluster of Dual-Xeons
- Fast-Ethernet 100 Mbps

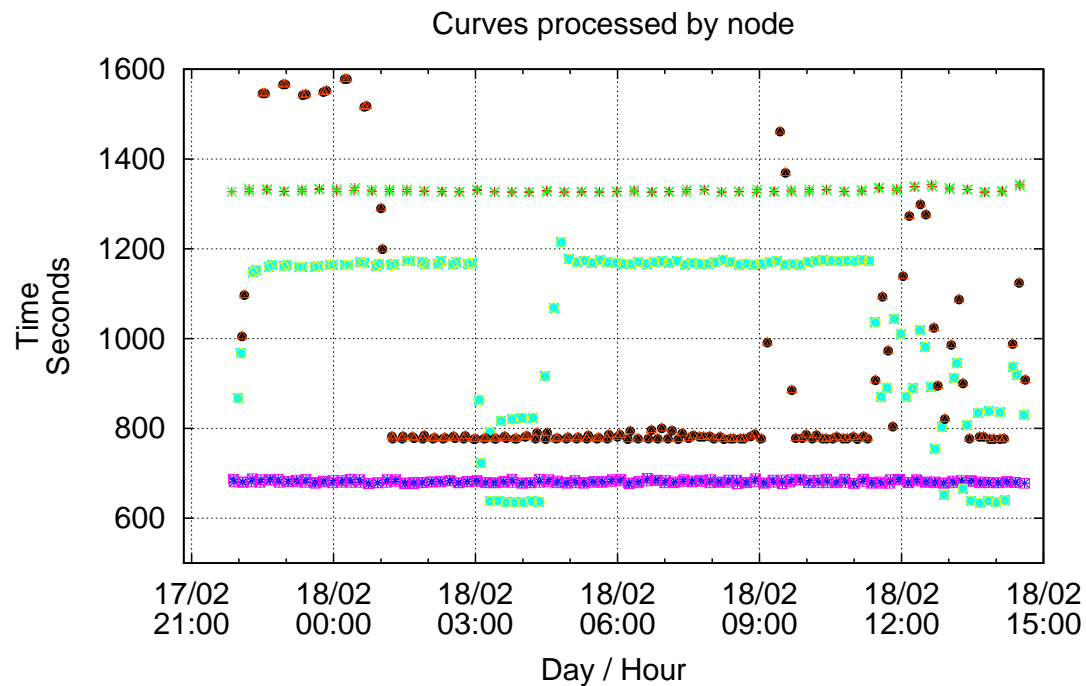
Challenge specification:

The number $P(10341)$ has 109 digits and was easily factored into

$$P(10341) = 2 * 3 * 5 * 1143831851 * C98$$

DISTECM tests

| C98 | B1=100000 Phase 1 | B1=400000 Phase 2 | B1=1500000 Phase 3 |
|------------------|----------------------|----------------------|-----------------------|
| Total Time | 3h 4' | 21h 39' | 17h e 7' |
| Processed curves | 1100 | 2900 | 518 |
| Factored | No | No | Yes |



Gandalf × Hobbit □ Isildur ○ Aragorn △ ●

DISTECM results

Factored C98 as:

$$p_{40} = 1432602103187822193769848905472938885747$$

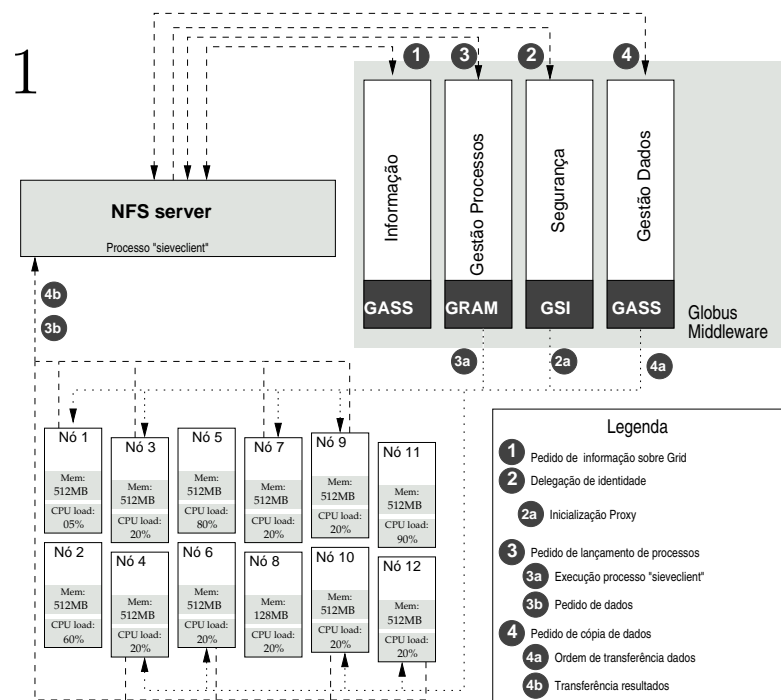
$$p_{59} = 544049964001456954474691336991327263416950\backslash$$
$$2409607474324463$$

Using 42 hours of the SMP Dual-Xeon cluster (with 8 processors).

Future work

- Enhancements of these metrics in the scope of the Global grid Forum (GGF)
- Developing distributed factorization platforms: NFS
With CWI (which was involved with RSA 512 factorization)

Already factored $11^{104} + 1$



Conclusions

Finally:

- Depicted metrics can be of help in the process of defining a formal approach to understand what applications are appropriate for use in grids
- Should not be used in a deterministic way as a “yes or no” oracle suggestion
- Useful to perform grid / application specification analysis and avoid that requirements definition is based only in “common sense experience”
- Distributed ECM (DISTECM) platform has proved to be a good approach for factoring numbers under 100 digits

•
•



Thank you. Questions ??

Paulo Trezentos

Paulo.Trezentos@iscte.pt

<http://paulo.trezentos.gul.pt/articles/>