

Review of Master's Thesis

Student: Martín Gago José
Title: Asynchronous Input/Output for Scientific Simulations (id 19648)
Reviewer: Vaverka Filip, Ing., UPSY FIT VUT

- 1. Assignment complexity** **average assignment**
- 2. Completeness of assignment requirements** **assignment not fulfilled**

The assignment wasn't fulfilled in its entirety as some parts are drawn up only partially and other parts are missing. To be specific, the author did some work at studying supercomputer architectures and options for distributed parallel I/O (although only MPI standard was considered).
The author further implemented a few micro-benchmarks and tested various approaches to parallel I/O and performed series of benchmarks, yet the analysis and interpretation of the results is quite lacking (for example, there are no explicit measurements of latency and throughput of the I/O - only simple wall-time measurements are present). The rest of the assignment (points 4-7) are missing (ie. asynchronous I/O for scientific application was neither designed, implemented nor evaluated).
- 3. Length of technical report** **shorter than minimum requirements**

The length of the technical report doesn't fulfill minimal requirements. About one half of the text is widely known theory about cluster architectures and MPI standard (essentially compilation). The rest of the text describes implemented micro-benchmarks and attempts to analyze obtained results. The discussion of the results is mostly superficial and no specific solutions are proposed.
- 4. Presentation level of technical report** **25 p. (F)**

The structure of the text is logical but the division into sections is not uniform (some sections are only a few sentences long, others span over multiple pages). Some of the crucial plots are removed from the text and moved into appendices, although they are necessary for understanding of the text itself.
- 5. Formal aspects of technical report** **10 p. (F)**

Linguistic quality of the work is so poor it makes understanding of its contents difficult and often leads to confusion.
There is a number of flaws in the formal aspects, the most problematic of which is the lack or non-descriptiveness of plot titles and axis labels (all plots are titled as "Times" and axis labels are missing entirely). Figures are referenced loosely and it's hard to tell what figure is the author talking about. A few images seem to be taken over from various publications but proper citations of their sources are missing.
- 6. Literature usage** **25 p. (F)**

The author doesn't seem to work with literature well. There are only 10 references listed and the majority of them are Wikipedia articles and some of them are redundant (e.g. TOP500 is referenced directly and also as a Wikipedia page). There is also an online reference to the university campus website which seems to be accessible only for enrolled students.
- 7. Implementation results** **40 p. (F)**

The student only implemented a set of micro-benchmarks (instead an actual scientific application he was supposed to implement). Yet, the code of these benchmarks is mostly readable, documented and an usage of OpenMPI is mostly correct.
- 8. Utilizability of results**

This work contains some new and potentially useful results in the area of distributed I/O behavior, yet experiments are insufficiently described. There is little to no information about the setup of benchmarks, such as MPI or filesystem settings.
- 9. Questions for defence**
 - What filesystem was used to perform benchmarks and how it was set up?

- What could be done to minimize an overhead of writing a small file from many nodes in parallel?
- How does achieved I/O throughput compare to peak throughput of the machine used to conduct benchmarks?

10. Total assessment

25 p. failed (F)

The main problem of this work is that it doesn't fulfill the assignment in its entirety and the core of the work is missing. Aforementioned flaws in formal aspects and grammar are severe and make the text hardly readable and understandable. The work contains some new information, yet its value is undermined by its presentation form and incompleteness.

In Brno 12. January 2017

.....
signature