THANOS (Thread & memory migration Algorithms for NUMA Optimised Scheduling)

Hopefully, Marvel will take this as a fan reference and not like a copyright infringement.

This repo holds the migration tool for NUMA systems developed by Ruben Laso Rodriguez for his PhD thesis as a continuation of Oscar's García Lorenzo PhD thesis.

Description

This migration tool uses hardware counters to decide whether a thread or a memory page needs a migration and its destination in a NUMA system.
Multiple metrics, like instructions per seconds, floating-point operations and mean memory latency are consulted to evaluate both performance and the likelihood of a thread to be migrated. These metrics are obtained using Linux `perf_event` interface and Intel PEBS (Precise Event-Based Sampling).

Currently, the thread migration strategies are based on tickets or scores, where several characteristics of a thread and its possible destinations are evaluated, and scores are assigned.
For the algorithms working with tickets, after every migration is evaluated (tickets are assigned), a weighted random process is done to select the migration to be performed. The higher the number of tickets, the higher the chances of a migration to be selected.
For the algorithms working with scores, the migrations with higher scores are those to be performed.

For the memory migration algorithms, metrics like average latency are taken into account to move pages across the system. Generally, pages are migrated to their favourite node (the node which makes more accesses to that page), but other consideration might be made to avoid phenomena like memory congestion.

For further details on how the algorithms work, please read the papers noted below.
If a particular algorithm is not listed in those papers, it is because the paper in which is included is still under review or it was simply not good enough.

Requirements

This is the list of requirements:

- Linux platform (should still work with old kernels).
- A C++ compiler supporting C++20 standard.
- Intel based processors in a NUMA system.
- Currently, this tool uses more than four general-purpose hardware counters per core, so hyperthreading should be disabled to raise the maximum available counter up to eight.
- The content of the file `/proc/sys/kernel/perf_event_paranoid` should be 0 or lower.

Compile

Use `cmake` to build `thanos`. Recommended to build in `Release` mode for optimal performance.
For example:

```
mkdir build
cd build
cmake .. -DCMAKE_BUILD_TYPE=Release
make
```

Further compilation options are shown with:

```
cmake <path to CMakeLists.txt> -LAH
```

Let me know if you have any trouble with the compilation of the code.

Execution

Once you have the compiled executable, the execution of this tool is very similar to `mpirun`. Just execute:

```
$ ./thanos [options] your_program
```

Note that this tool will only migrate its child process, leaving the rest of the system untouched to avoid, as much as possible, collisions between OS scheduler and the migration tool.

If you wish to execute several programs to be managed and scheduled by the tool, gather them in a script and execute:

```
$ ./thanos [options] your_script
```

The list of options is available through the command

```
$ ./thanos -h
```

Contact

For any further doubt in the compilation or execution process, algorithmic of the tool, etc., or any kind of suggestion, do not hesitate to contact me at r.laso@usc.es or ruben.laso.rguez@gmail.com.

Finally, if you use this tool for research, please contact me for obtaining the citation of the last published paper related to this tool.

Relevant publications

Paper explaining CIMAR, NIMAR and LMMA:

- Laso, R., Lorenzo, O. G., Cabaleiro, J. C., Pena, T. F., Lorenzo, J. Á., & Rivera, F. F. (2022). CIMAR, NIMAR, and LMMA: Novel algorithms for thread and memory migrations in user space on NUMA systems using hardware counters. Future Generation Computer Systems, 129, 18-32. https://doi.org/10.1016/j.future.2021.11.008

```
@article{LASO202218,
    title = {CIMAR, NIMAR, and LMMA: Novel algorithms for thread and memory migrations in user space on NUMA
    journal = {Future Generation Computer Systems},
    volume = {129},
    pages = {18-32},
    year = {2022},
    issn = {0167-739X},
    doi = {https://doi.org/10.1016/j.future.2021.11.008},
    url = {https://www.sciencedirect.com/science/article/pii/S0167739X21004374},
    author = {Ruben Laso and Oscar G. Lorenzo and José C. Cabaleiro and Tomás F. Pena and Juan Ángel Lorenzo
    keywords = {NUMA, Scheduling, Thread migration, Memory migration, Hardware counters}
}
```

Papers explaining LBMA and IMAR^2:

- Laso, R, Lorenzo, OG, Rivera, FF, Cabaleiro, C, Pena, TF, Lorenzo, JA. LBMA and IMAR2: Weighted lottery based migration strategies for NUMA multiprocessing servers. Concurrency Computat Pract Exper. 2021; 33:e5950. https://doi.org/10.1002/cpe.5950

```
@article{https://doi.org/10.1002/cpe.5950,
    author = {Laso, R. and Lorenzo, O. G. and Rivera, F. F. and Cabaleiro, J. C. and Pena, T. F. and Lorenzo,
    title = {LBMA and IMAR2: Weighted lottery based migration strategies for NUMA multiprocessing servers},
    journal = {Concurrency and Computation: Practice and Experience},
    volume = {33},
    number = {11},
    pages = {e5950},
    keywords = {hardware counters, performance, roofline model, thread migration},
    doi = {https://doi.org/10.1002/cpe.5950},
    url = {https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.5950},
    eprint = {https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpe.5950},
    abstract = {Summary Multicore NUMA systems present on-board memory hierarchies and communication networks
    year = {2021}
}
```

- García Lorenzo O., Laso Rodríguez R., Fernández Pena T., Cabaleiro Domínguez J.C., Fernández Rivera F., Lorenzo del Castillo J.Á. (2020) A New Hardware Counters Based Thread Migration Strategy for NUMA Systems. In: Wyrzykowski R., Deelman E., Dongarra J., Karczewski K. (eds) Parallel Processing and Applied Mathematics. PPAM 2019. Lecture Notes in Computer Science, vol 12044. Springer, Cham. https://doi.org/10.1007/978-3-030-43222-5_18

```
@InProceedings{10.1007/978-3-030-43222-5_18,
    author="Garc{\'i}a Lorenzo, Oscar
    and Laso Rodr{\'i}guez, Ruben
    and Fern{\'a}ndez Pena, Tom{\'a}s
    and Cabaleiro Dom{\'i}nguez, Jose Carlos
    and Fern{\'a}ndez Rivera, Francisco
    and Lorenzo del Castillo, Juan {\'A}ngel",
    editor="Wyrzykowski, Roman
    and Deelman, Ewa
    and Dongarra, Jack
    and Karczewski, Konrad",
    title="A New Hardware Counters Based Thread Migration Strategy for NUMA Systems",
    booktitle="Parallel Processing and Applied Mathematics",
    year="2020",
    publisher="Springer International Publishing",
    address="Cham",
    pages="205--216",
    abstract="Multicore NUMA systems present on-board memory hierarchies and communication networks that infl
    isbn="978-3-030-43222-5"
}
```

License

- Ruben Laso.