Title: Profiling interprocess communications on multicore servers

Advisors:

- Vivien Quéma, Grenoble INP
- Renaud Lachaize, Université Grenoble 1

Context:

The context of this internship is a freshly started research project focused on the design and implementation of a new operating system (OS) optimized for emerging hardware architectures and modern workloads. The internship is tightly integrated with the project and offers opportunities to carry on the research as a Ph.D student.

Modern machines are very difficult to program because of their rich and fast evolving topologies, which include complex network-on-chips interconnecting many processing cores, deep memory hierarchies with non-uniform access times (NUMA), and high-speed I/O (storage and networking) devices. Besides, modern workloads (e.g., Cloud-based services and Big data applications) are outstanding because they have very demanding requirements on all the hardware resources (CPUs, memory, storage and networking devices). In addition, these workloads also stress the logical facilities of the operating system, for instance the memory management and interprocess communication subsystems.

We believe that current operating systems like Linux are a poor match for such a context. In particular, they suffer from a lack of correlation between chains of resource dependencies. Indeed, current operating systems make task scheduling decisions based on a very narrow view of the logical resources. For instance, a task may be elected to run on a core and almost immediately get blocked because it attempts to acquire an unavailable lock or to write in a shared buffer with no space left. Such a sequence of events is inefficient as it exacerbates the overhead of the OS (e.g., increasing the frequency of context switches) and, when generalized at the scale of thousands of threads, can significantly hurt performance.

Internship:

The goal of the internship is to design and implement a profiling tool for programmers and OS designers. The purpose of this tool is to trace the chains of resource dependencies in order to allow pinpointing inefficient interactions between multiple tasks (i.e., processes and threads). The first version of the tool will target the Linux kernel and focus on interactions via explicit communication channels, such as pipes and Unix domain sockets. The key research problems to be addressed are:

1) The design of a low-overhead and accurate mechanism for tracing interactions between execution flows;
2) The design of algorithms for post-mortem processing, aimed at automatic detection of inefficient scheduling/interaction patterns;
3) Suggestions for adapting the algorithms to on-line processing.

The internship will involve the following phases:

1) Study of the related work on existing profiling tools and basic building blocks for efficient instrumentation of operating system kernels;
2) Design and implementation of the profiling tool on Linux;
3) Experimental validation of the tool (accuracy, overhead, insight) on synthetic and realistic workloads.

Keywords: Operating systems; Multicore; Linux kernel; Profiling.

Two selected publications made by the research team on the internship topic:

- *Traffic Management: A Holistic Approach to Memory Placement on NUMA Systems*. Mohammad Dashti, Alexandra Fedorova, Justin Funston, Fabien Gaud, Renaud Lachaize, Baptiste Lepers, Vivien Quéma, and Mark Roth. In Proceedings of the International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS), Houston, USA, March 2013.

- *MemProf: A Memory Profiler for NUMA Multicore Systems*. Renaud Lachaize, Baptiste Lepers, and Vivien Quéma. In Proceedings of the USENIX Annual Technical Conference (USENIX ATC), Boston, USA, June 2012.