PAMP Predictable Parallel Protocol Processing

A project proposal for PAMP/ARTES

Per Gunningberg and Mats Björkman
Uppsala University

15th May 1998

1 Short summary

We propose a project within PAMP that addresses multiprocessor communication issues. The focus of the Uppsala work will be on support for parallel protocol implementations on SMPs when predictable delay is of importance, and on operating system/run time environment resource management support for real-time protocols written in high level languages such as Erlang. The industrial partner is Ericsson SARC and the contact persons are Håkan Millroth/Bjarne Däcker.

2 Problem statement

With the rapid growth of networking and networked applications, the demands put on the communication subsystem have increased tremendously the past years and will continue to increase. Since the communication performance bottleneck in today's high-speed networks most often is found in the nodes rather than in the physical links themselves, methods that better utilize the available systems are needed. We foresee a growing number of SMP platforms both as end-node clients and as network servers. Many servers will handle distributed real-time applications such as video-on-demand, distributed multi-player games and video conferences. Communication subsystem resource management must therefore be able to meet QoS demands. Hence, there is a need for methods and tools for efficient implementation of communication protocols and resource management on SMP systems.

The proposed work will be divided into two subtasks, one on network interfaces and one on real-time support for communication.

2.1 Subtask 1: Parallel Network Subsystems

Within the area of parallel network subsystems, the objective is to get the right data to/from the network from/to the right process at the right time, as efficient as possible. There are a couple of interconnected issues here. The first issue relates to the network architecture that should be designed to match the multiprocessor architecture and its internal interconnection architecture. Problems that arise here are: Where should the interface(s) be located? It may be one interface shared by all processors or attached to one of them, or there may be multiple interfaces attached to separate processors which all may need to be synchronized. What are the real-time characteristics and performance tradeoffs in these architectural design choices?

A second issue is how to make parallel protocol implementations that have predictable performance while still achieving high utilization.

For many applications with QoS demands, a predictable *delay* is the most important factor. While it is true that a predictable minimum throughput (steady-state over a finite time) automatically gives a bound on the maximum delay (because of the bounded delay jitter and some known minimum delay), most applications with demands on delay have much lower tolerances than those implied by throughput constraints. Hence, predictable delay in parallel communication protocol implementations is an important field of study.

In our previous efforts on performance of multiprocessor network interfaces, which we intend to build on, we have mainly studied throughput optimizations. The proposed work will thus complement our previous work.

2.2 Subtask 2: Real-time Support for Communication

The Internet of the future will be very heterogeneous, ranging from optical based high capacity networks to wireless networks with low bandwidth, high error rate, and varying connectivity. Higher layer protocols and distributed applications are likely to be adaptive to compensate for this heterogenity. Another trend is the use of "fat" network servers as proxies or agents to carry out parts of the tasks of "thin" clients with low capacity or poor network connectivity. Many applications of the future will require real-time and predictive services. Examples of such applications include video conferences, virtual meeting rooms, distributed interactive simulations and plain old telephony. In order to meet the Quality of Services (QoS) required by these applications, resources in the network, in network servers, as well as in the end systems, must be reserved and scheduled according to end-to-end QoS parameters.

We are doing research on operating systems that supports real-time applications and higher layer protocols. The operating system must manage and schedule its resources to achieve high utilization while still providing predictable performance. The dominant resources in a *uni*processor are the CPU, the primary memory and the bus/network interface. The objective with this proposed research is to extend this to a multiprocessor environment, where multiple processors and memory can be shared or reserved and where there can be contention for shared interconnects (e.g. buses). To achieve its objective, the operating system must schedule its resources in a concerted way, also in interaction with the reservation schemes of the network (such as ATM or IETF traffic classes).

3 Main ideas

3.1 Subtask 1: Parallel Network Subsystems

We propose the study of multiprocessor communication subsystems with emphasis on how to achieve predictable delay while still retaining high performance.

In our approach we intend to study simulation models or emulations of different network architectures that are detailed enough to get predictive performance results. The work builds on our previous research on ATM/SCI interfaces, ATM/TCP interaction, the design of efficient network APIs (sockets) and parallel protocol implementations.

3.2 Subtask 2: Real-time Support for Communication

In this subtask we will study how the operating system can support adaptive protocols and real-time applications written in high level languages, such as Erlang. It is an extension to the scheduling work at Chalmers with respect to communication. We will investigate into operating system mechanisms and abstractions for the programmers to control the resources of the machine. Current general purpose operating systems, such as Unix, do not offer sufficient control, or have inefficient mechanisms for real-time distributed applications.

4 Expected results and impacts

Expected results are software modules, tools and methods for efficient and predictable parallel protocol implementations on SMPs.

We expect a number of academic and industrial results regarding protocol processing and network interfaces to SMPs. The main expected academic results are licentiate and doctoral theses within the following areas:

- Methods for predictable delay characterization of high performance parallel communication sub-systems for SMPs.
- Resource reservation principles for real-time protocols in SMP servers.

The expected industrial results include:

- Tools and software modules for protocol implementations (written in e.g. Erlang) with a fine grained control of host and network resources. This result is of interest for Ericsson SARC/ Erlang Systems.

- Partitioning and parallelizing strategies for predictable delay in multiprocessor protocol implementations. These are anticipated to be of interest for Ericsson.

5 Project plan

5.1 Subtask 1: Parallel Network Subsystems

Phase 1 of this subtask will include the characterization of those aspects of multiprocessor networking that must be modelled in order to correctly capture the behaviour of the system with respect to predictable delay.

Phase 2 of this subtask will include the development of simulation models and implementations that capture the aspects characterized in phase 1 of the subtask.

Phase 3 of this subtask will be an analysis of the predictive capabilities of the developed models and the applicability to networked applications.

Milestones and deliverables:
Deliverable (1 year after start): Report on state-of-the-art.
Deliverable (2.5 years after start): Licentiate thesis.
Deliverable (5 years after start): Final report. Doctoral thesis.

5.2 Subtask 2: Real-time Support for Communication

Phase 1 of this subtask includes the identification of the resources that must be controlled by the operating system in order to be able to give Quality-of-Service guarantees.

Phase 2 of this subtask includes the development of software modules, mechanisms and policies for communication sub-system resource management in an SMP environment.

Phase 3 of this subtask includes the application of resource management as developed in phase 2 of this subtask to applications with Quality-of-Service demands.

Milestones and deliverables:
Deliverable (1 year after start): Report on state-of-the-art.
Deliverable (2.5 years after start): Licentiate thesis.
Deliverable (5 years after start): Final report. Doctoral thesis.

6 Preliminary budget

The project will involve two Ph. D. students and two senior advisors. The activity level of these will be 80% and 20% respectively. Equipment depreciation cost is based on three workstations with high performance network interfaces, software and regular maintenance costs, at a cost level of 100KSEK each. Travelling cost estimate is based on internal Swedish travel trips motivated by the close collaboration with the other PAMP/ARTES partners and normal travelling for international conferences, workshops etc.

Item	Cost
Researcher 20 %, Per Gunningberg, Docent	72 kkr
Researcher 20 %, Mats Björkman, PhD	72 kkr
PhD Student 80 %	163 kkr
PhD Student 80 %	163 kkr
Social costs (45.5 % of rows 1 – 4)	214 kkr
Dept. OH (13.64 % of rows 1 – 5)	93 kkr
Equipment depreciation	100 kkr
Travel	105 kkr
Office cost (10 % of rows 1 – 8)	98 kkr
University OH (13.64 % of rows 1 – 8)	134 kkr
Högskolemoms (8.7 % of rows 1 – 10)	121 kkr
Total	1335 kkr

7 Related research

The proposer, the Communication and Distributed Systems group at the Department of Computer Systems, Uppsala University, has a research competence within multiprocessor protocol implementations. The group has been working in the data communication area for more than a decade with a focus on end system (host) issues. It has produced prototypes of multiprocessor implementations of protocols, efficient network interfaces and measurement tools, including the development of the parallel x-kernel. Members of the group has also contributed to Ericsson's switch control protocols.

There are several relevant projects ongoing at Uppsala University of which PAMP will benefit from. They include the Esprit Long Term Research project HIPPARCH, on new communication protocol architectures, and a project for Ericsson SARC on operating systems support for protocols written in Erlang. Other relevant work include multiprocessor implementations of the Mach operating system and Desk Area Networks for multi computer systems.

Nationally, the academic partners behind the PAMP proposal perform important related research. Internationally, contacts with researchers performing related research includes Jonathan Smith at UPenn, involved in parallel networking as well as Kurose/Towsley at UMass. For QoS support, the EU Nemesis project and the Campbell et al (Columbia) QoS Architecture are important related reseach projects.

8 Relation to the profile

The main objective of PAMP is to develop methods and tools for exploiting symmetric multiprocessors for real-time applications. Network servers for real-time distributed applications are expected to be SMPs in the future. Uppsala University will address the communication, parallel protocol processing and network interface issues in the design of SMPs for applications with soft real-time demands.

9 Industrial relevance

The proposed Uppsala work addresses predictability and performance of the interconnection of high performance multiprocessors. Hence, any industry writing distributed real-time applications running on interconnected SMP systems that require QoS will benefit from the Uppsala research.

Our research on QoS support for communication is of direct interest to Ericsson SARC as explained above.

10 Relation to other SSF programmes

There are two research programmes supported by SSF that are to some extent related to our project proposal. PCC addresses communication, but mainly radio and protocol design issues. "Vetenskapliga beräkningar" addresses multiprocessor issues, but only from the viewpoint of applications for scientific calculations.

11 Context

11.1 The research group

Our Uppsala research group is headed by Docent Per Gunningberg and consists today of two senior researchers (Per Gunningberg and Mats Björkman), six PhD students (Kristina Lundqvist (lic), Björn Knutsson and Bob Melander at Uppsala, Thiemo Voigt (externally financed) and Gordon Beaton and Catrin Hansson (both inactive)) as well as one associated researcher (Bengt Ahlgren from SICS).

11.2 Complementary activities and funding

Current funding:

Type	Source	MSEK/year	Expires	Comments
Long Term Research	EC	1.1	Sept 98	
Industry	Ericsson	0.5	June 98	May be renewed
Basic	TFR	0.3	Dec 98	

Funding for 1999:

Type	Source	MSEK/year	Status	Comments
Industry	Ericsson	0.5	Pending	Decision pending
Basic	KK	0.8	Planned	Planned submission
Basic	TFR	0.8	Proposed	Submitted

11.3 Research cooperation

Apart from the research cooperation that relates to the PAMP project as described above, we have a close research cooperation with the Computer and Network Architecture (CNA) group at the Swedish Institute of Computer Science (SICS). We are also part of a cluster within NUTEK's Complex Systems program.

Our foremost international collaborations are within the Esprit LTR HIPPARCH project and with Telenor Research Norway. Other relevant international collaborations are those with the network research group at the University of Arizona and with the University of Karlsruhe.

The HIPPARCH project is of strategic importance to us. It has high visibility and excellent partners. They are UCL London, INRIA Nice, SICS, Dassault Electronique and University Technology Sydney (UTS) Australia. The project has generated direct research contacts between UTS, Ericsson Radio and Uppsala University.

We have an extensive informal contact network in Europe as well as in the United States thanks to significant PC assignments and long/short-term visits. For example, during last year, our group at DoCS had six internationally highly recognized short-term visitors.

11.4 Industrial cooperation

Our industrial partner in the proposed project is Ericsson SARC.

Contact person: Håkan Millroth. Email: hakanm@erix.ericsson.se

Phone: 08 - 719 86 74. Address: Ericsson SARC, Älvsjö.