Master Degree Program in Computer Science and Networking

High Performance Computing

4th appello – July 3, 2013

The answers can be written in English or in Italian. Please, present the work in a legible and readable form. All the answers must be properly and clearly explained.

Question 1

A computation is composed of processes P and Q interacting according to a request-reply manner.

Each requests from P to Q is an integer array B[M]. The replay from Q to P is an integer array C[M]. P sends a new request B after receiving the reply C. C is the result of a matrix-vector product C = A × B , where matrix A[M][M] is statically encapsulated in Q.

M = 2K.

Q is executed on a D-RISC architecture with scalar pipelined CPU, single buffering, 4-stage integer functional units, on-demand primary data cache with capacity 32K words and block size 8 words, secondary cache on-chip assumed with negligible fault probability.

Define and evaluate a parallel implementation of Q to be executed on a parallel architecture with $N = 128$ PEs based on the CPU described above, $T_{setup} = 1000\ \tau$, $T_{trasm} = 25\ \tau$, exclusive process mapping, zero-copy communication support, and communication processor.

Question 2

Verify if the assumed values of T_{setup} and T_{trasm} are well approximated for the execution of the parallel program of Question 1 on the following architecture:

i) all-cache SMP multiprocessor with shared main memory;

ii) 2-ary 7-fly wormhole interconnection network, with 1-word links and flits, single buffering interfaces, and link transmission latency equal to 4τ;

iii) interleaved memory macro-modules, each one with 8 modules and clock cycle equal to 30τ;

iv) CPU defined in Question 1;

v) periodic retry locking;

vi) automatic directory-based cache coherence with invalidation.

Note: students can request to shortly look up the Course Notes copy on the teacher's desk for formulas, numeric values and figures related to performance metrics.