

# Complexity Analysis of Pipeline Mapping Problems in Distributed Heterogeneous Networks

YI GU<sup>1</sup>, QISHI WU<sup>1</sup>, MENGXIA ZHU<sup>2</sup>,  
and NAGESWARA S. V. RAO<sup>3</sup>

<sup>1</sup>Dept. of Computer Science, University of Memphis, Memphis, TN, USA

<sup>2</sup>Dept. of Computer Science, Southern Illinois University, Carbondale, IL, USA

<sup>3</sup>Computer Science & Math Div, Oak Ridge National Laboratory, Oak Ridge, TN, USA

*Largescale scientific applications require using various system resources to execute complex computing pipelines in distributed networks to support collaborative research. System resources are typically shared in the Internet or over dedicated connections based on their location, availability, capability, and capacity. Optimizing the network performance of computing pipelines in such distributed environments is critical to the success of these applications.*

*We consider two types of largescale distributed applications:*

- (i) *interactive applications where a single dataset is sequentially processed along a pipeline; and*
- (ii) *streaming applications where a series of datasets continuously flow through a pipeline.*

*The computing pipelines of these applications consist of a number of modules executed in a linear order in network environments with heterogeneous resources under different constraints. Our goal is to find an efficient mapping scheme that allocates the modules of a pipeline to network nodes for minimum endtoend delay or maximum frame rate.*

*We formulate the pipeline mappings in distributed environments as optimization problems and categorize them into six classes with different optimization goals and mapping constraints:*

- (i) *Minimum Endtoend Delay with No Node Reuse (MEDNNR),*
- (ii) *Minimum Endtoend Delay with Contiguous Node Reuse (MEDCNR),*
- (iii) *Minimum Endtoend Delay with Arbitrary Node Reuse (MEDANR),*
- (iv) *Maximum Frame Rate with No Node Reuse or Share (MFRNNRS),*
- (v) *Maximum Frame Rate with Contiguous Node Reuse and Share (MFRCNRS), and*
- (vi) *Maximum Frame Rate with Arbitrary Node Reuse and Share (MFRANRS).*

*Here, “contiguous node reuse” means that multiple contiguous modules along the pipeline may run on the same node and “arbitrary node reuse” imposes no restriction on node reuse. Note that in interactive applications, a node can be reused but its resource is not shared.*

*We prove that MEDANR is polynomially solvable and the rest are NP-complete. MEDANR, where either contiguous or noncontiguous modules in the pipeline can be mapped onto the same node, is essentially the Maximum n-hop Shortest Path problem, and can be solved using a dynamic programming method. In MEDNNR and MFRNNRS, any network node can be used only once, which requires selecting the same number of nodes for onetoone onto mapping. We show its NP-completeness by reducing from the Hamiltonian Path problem. Node reuse is allowed in MEDCNR, MFRCNRS and MFRANRS, which are similar to the Maximum n-hop Shortest Path problem that considers resource sharing. We prove their NP-completeness by reducing from the Disjoint-Connecting-Path Problem and Widest path with the Linear Capacity Constraints problem, respectively.*

**Keywords** Computational complexity; NP-complete; Distributed computing; Pipeline mapping

Address correspondence to Qishi Wu, Dept. of Computer Science, University of Memphis, Memphis, TN 38152, USA. E-mail: qishiwu@memphis.edu

