

snow: Simple Network of Workstations

Luke Tierney

Department of Statistics & Actuarial Science
University of Iowa

September 13, 2007





Motivation and Background

- Objective: shared memory parallel computing using R.
- Several tools are available, including
 - raw socket (`socketConnection`, `serialize`, `unserialize`)
 - rpvm package
 - Rmpi package
- Also available: parallel random number generators, including
 - rsprng package
 - rlecuyer package
 - rstreams package
- PVM and MPI are very powerful but also complex.
- Want higher level facilities that
 - make it easy to do simple scatter-compute-gather computations
 - can transparently use different communication back ends
 - simplify handling of random number generation



The snow Package

- `snow` is a parallel computing package for R
- `snow` is motivated by the CoW package from Scientific Python.
- `snow` uses a master/worker model:
 - The user starts an ordinary R session as the master process.
 - This session creates (or connects to) a set of worker processes.
 - Jobs are sent to the worker processes and results are returned.
- The underlying message passing can be based on
 - raw sockets (no additional packages/software needed)
 - PVM (uses `rpvm` and PVM)
 - MPI (uses `Rmpi` and LAM-MPI; other MPIs may also work)
- Which communication mechanism is used only matters at startup.