SAND2011-5037C

Simulating Buffered Flow Control in Exascale Networks

Eric Anger, Curtis Janssen
Sandia National Laboratories, Livermore CA

MOTIVATION

- Developing the next generation of supercomputers requires cooperation between application programmers and hardware designers
- Simulations provide an accurate tool for estimating the performance of applications on specific hardware configurations
- Discrete event simulation represents the execution of an application on a machine as a chronological series of events, where the occurrence of an event modifies the state of the system
- The network used to pass data among processors accounts for a large portion of both machine performance as well as simulation runtime
- Accurately modeling the network would allow for both increased introspection into the interaction between application and network topology as well as increase the configurability and speed at which a simulation can be run

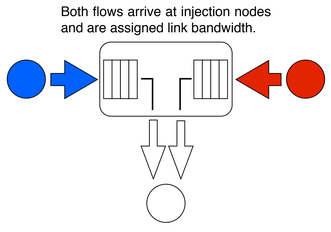
OBJECTIVES

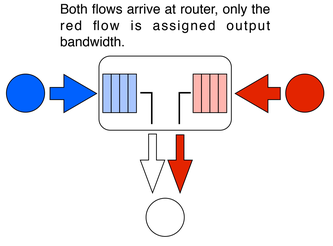
- Provide for accurate modeling of a flow-controlled, buffered network where message compete for link bandwidth
- Allow for the use of static or adaptive protocols
- Accurately represent message buffering, congestion, hot spots, and tree effects
- Enforce fair bandwidth allocation among messages competing for the same link
- Monitor progression of a message through a network, blocking on held resources
- Decrease simulation runtimes as compared to current circuit switched model

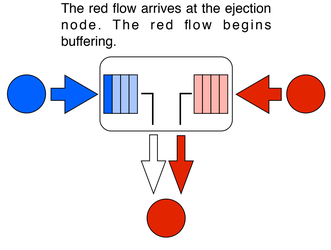
APPROACH

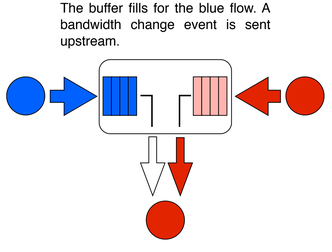
Flow progression through the network is implemented by several event types within the simulator:

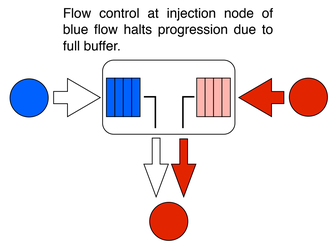
- **Arrival -** The head of a message has arrived at a vertex in the graph. It will now begin to compete for link bandwidth.
- **Departure -** The tail of a message has arrived at a vertex in the graph. A departure event will not propagate down the path of this message until all data still buffered has left. When all data has left, the flow's state is deleted and all held resources are freed.
- **Buffer Change -** A non-zero difference between input and output bandwidth assigned to a message at a router will cause the buffer to fill to capacity or drain completely, requiring the bandwidth assignments to be rebalanced.
- **Bandwidth Change -** Upstream and downstream buffer state affect the bandwidth rates assigned to flows. Flow control messages are sent in either direction as needed.

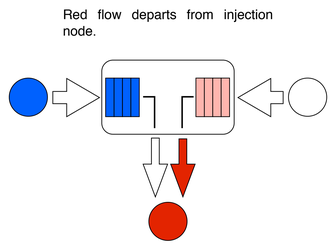


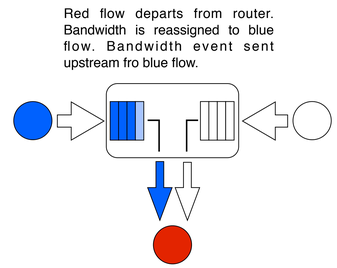


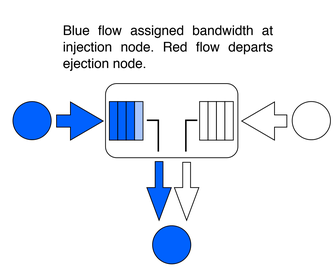


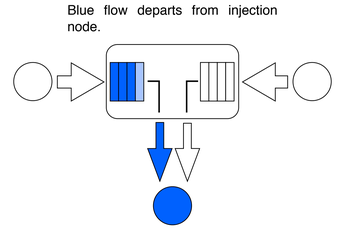


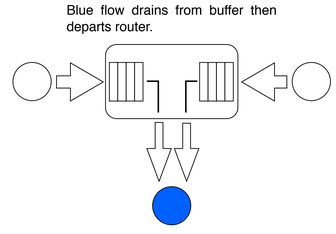












FUTURE WORK

- Performing max-min fair share bandwidth allocation for links would give a more realistic model of bandwidth sharing in the case of net demand for a link superseding supply
- Allow for configuration of number of virtual channels to limit active flows per vertex as well as facilitate deadlock-free adaptive routing protocols
- Improve flow control mechanism to real on/off or credit-based

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.