U.S. DEPARTMENT OF ENERGY

PNNL-20186

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

DnUp Routing Algorithm Description

J Carr
K Schmidt

February 2011

Pacific Northwest NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

DISCLAIMER

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service,
U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161
ph: (800) 553-6847
fax: (703) 605-6900
email: orders@ntis.fedworld.gov
online ordering: http://www.ntis.gov/ordering.htm

This document was printed on recycled paper.
(9/2003)

PNNL-20186

DnUp Routing Algorithm Description

J Carr
K Schmidt

February 2011

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99352

DnUp Routing Algorithm Description

Jared Carr, Ken Schmidt
Environmental Molecular Sciences Laboratory
Pacific Northwest National Laboratory
{jared.carr, kenneth.schmidt}@pnl.gov
2011-02-10 - PNNL-20186

Motivation

Chinook's network is not amenable to using UpDn routing algorithm, due to the unique nature of the cabling. Specifically nodes and uplinks are connected to the same switches (chips).

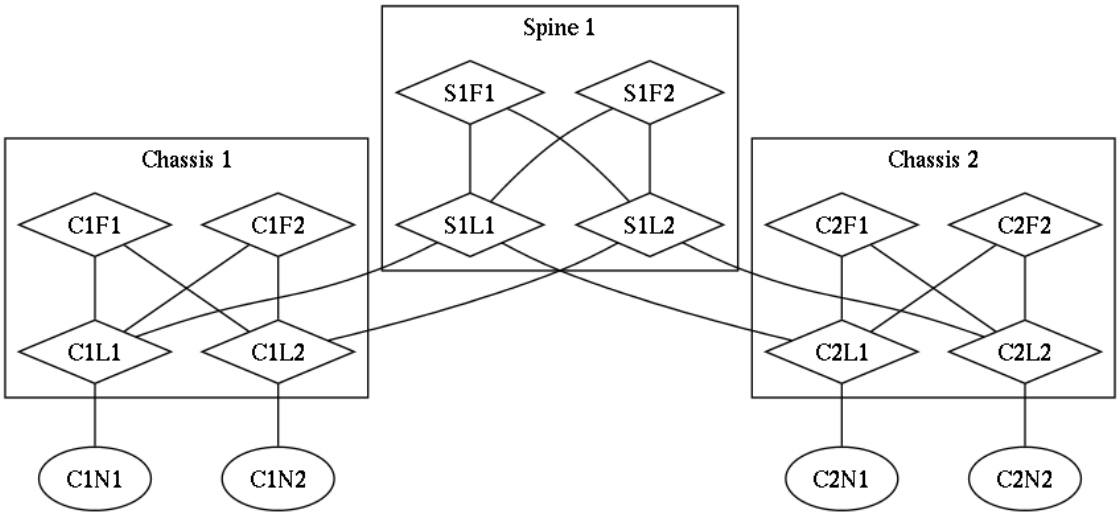

Figure 1: Physical Cable Layout.

Application of UpDn routing will cause all intra-node level chassis communication to use external ports on chassis instead of the internal fat-tree network. This will cause excessive congestion on only a few links leading to undesirable performance characteristics. Figure **2** illustrates the paths that can be generated by UpDn.

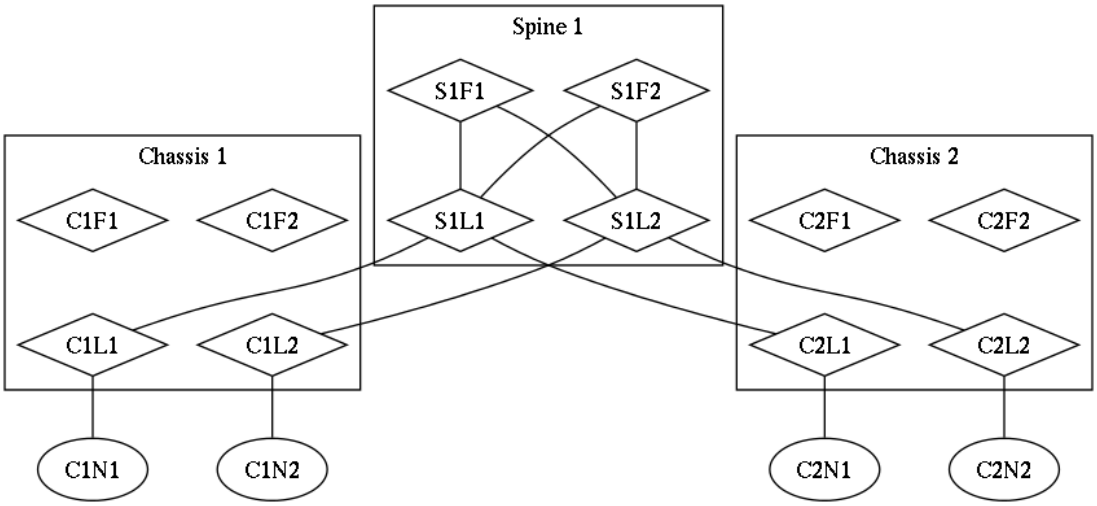

Figure 2: The possible paths using UpDown.

DnUp, The Proposed Algorithm

The purpose of DnUp is to allow intra-cu traffic to remain within a single chassis. At it's simplest DnUp reverses the ranking rules of UpDn along with the allowed transitions. Figure **3** illustrates the potentially allowed paths using DnUp. In particular nodes are capable of using their chassis fabric switches to talk to each other.

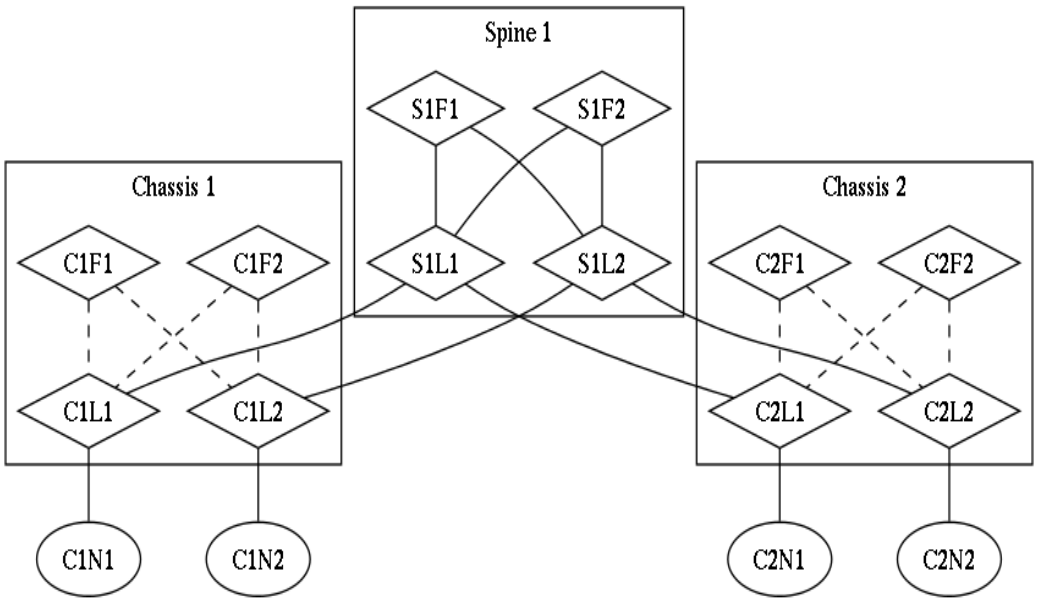

Figure 3: The possible paths using DnUp. These paths will be further restricted based on the ranking transitions described later. Dashed lines are paths explicitly added by DnUp in comparison to UpDn.

The UpDn Routing algorithm.

UpDn works by restricting routes by strict ranking rules. First a set of root nodes is determined, then nodes are ranked based on their distance from these root nodes. (Figure **4**.) Once all nodes have been ranked a filter is applied to the min-hop tables to eliminate invalid direction (rank) transitions, only a single decreasing to increasing rank transitions is allowed. After the min-hop tables are pruned the standard min-hop algorithm is used to generate routes from each node in the fabric to every lid.

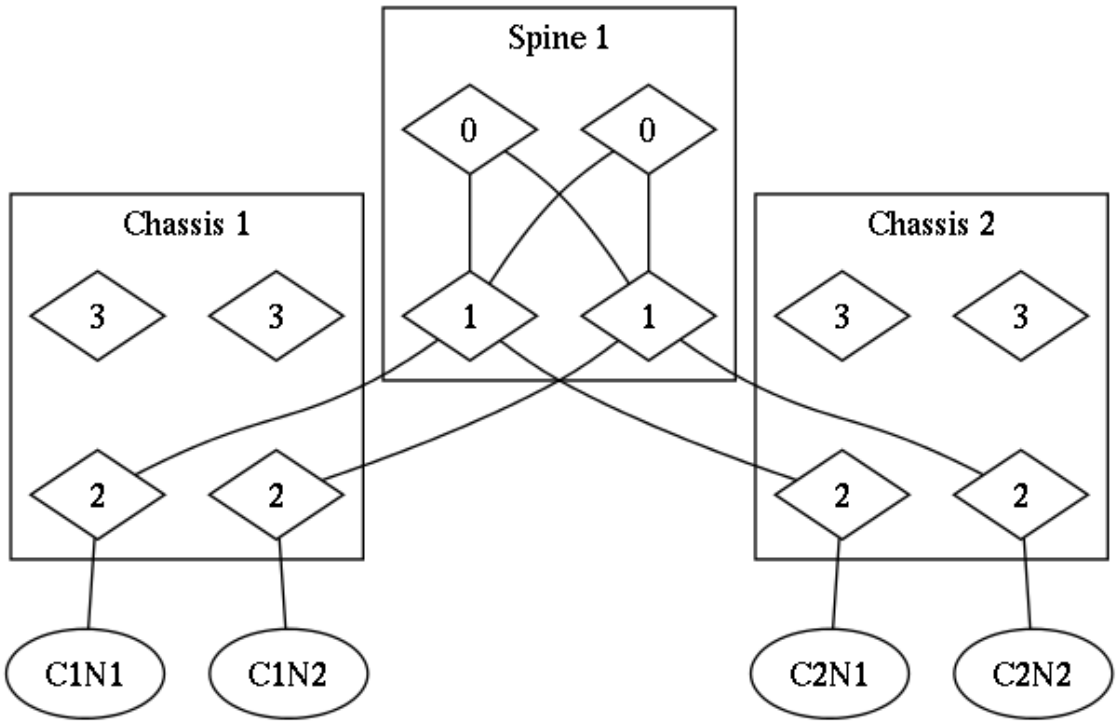

Figure 4: Ranks assigned by UpDn

The DnUp Routing Algorithm.

All switches are ranked in accordance to their relative distance from any single CA (Channel Adapter). Any switch directly connected to a CA is assigned a value of 0, any switch not connected to a CA is assigned to a value of one plus the minimum value of its neighbor switches. (Figure **5**) Next the min hop tables are pruned such that only one transition between increasing to decreasing ranks is allowed. In other words the only illegal transition is from Up to Down. After min-hop table pruning is complete the min-hop algorithm is applied to generate routes.

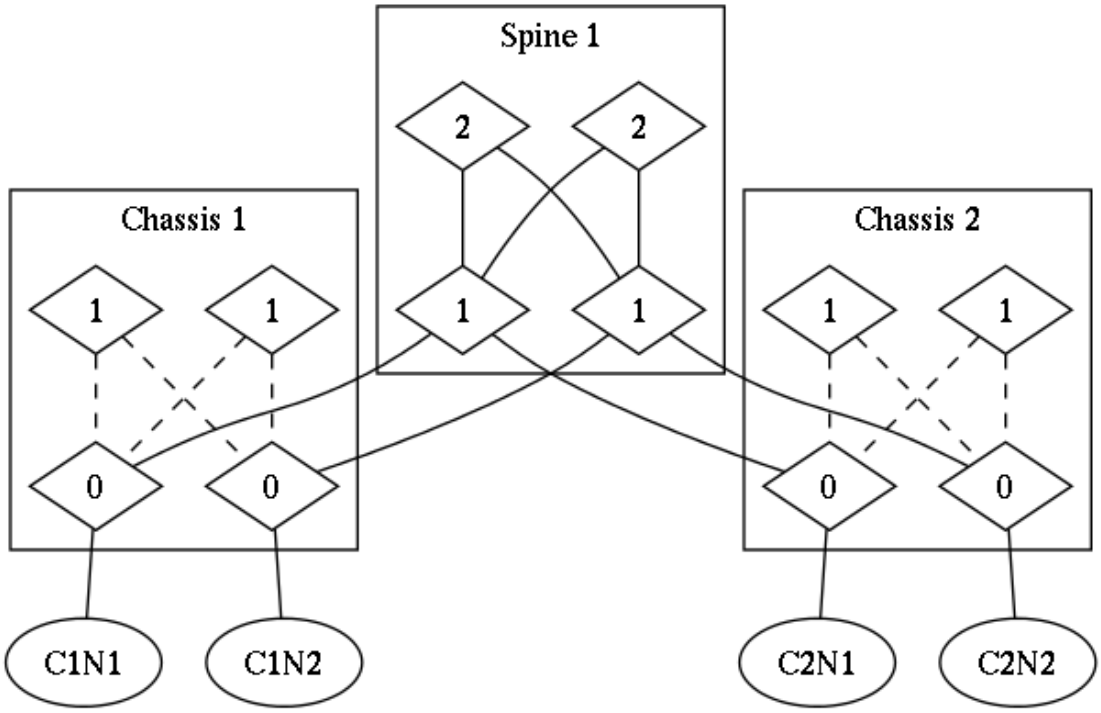

Figure 5: Ranks assigned by DnUp

Note: In some networks DnUp will not generate routes such that every node can talk to every switch in the fabric. In Figure **1** for example C2N1 will not be able to talk to C1F1.