# MessageWay: Extensions, Implementation

Anthony Skjellum, Thomas McMahon
Integrated Concurrent & Distributed
Computation Research Laboratory
Mississippi State University

IETF Meeting, Los Angeles
March 4, 1996

## Overview of Talk

- Proposed Extensions
  - Discussion
  - Motivation
- Implementation Status and Directions
  - UDP-based implementation
  - Follow ons
- Summary and Conclusions

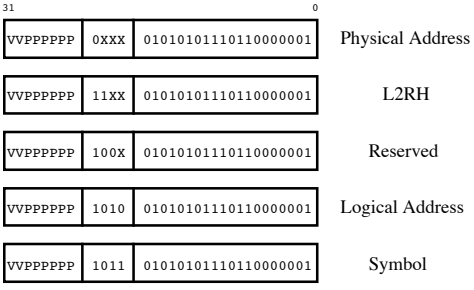
## Proposed Extensions

- Extended Addressing
- Optional Hierarchical Addressing
- Symbols
- Change to enhance priority support
- Proposed Extensions work in combination
- Designed to make MessageWay more flexible/ useful
- EEP header is still only 16-bytes in length!
- Supercedes recent transmissions on mailing list

## Extended Addressing

- Larger SAN conglomerates (32K addresses not enough!)
- Finer-grain addressing (*e.g.,* endpoints can be port-like)
- 24-bit addresses, embedded in LSb's of 32-bit word
- Version/Priority are in MSb's of 32-bit word
- RRP will need significant updates to conform to this scheme
- Makes other extensions, discussed next, possible

## Address/Routing Structure




## Optional Hierarchical Addressing

- Routing Tables contain "single matching" Node-Structures
- Augment to support mask: support hierarchical naming
- Can alternatively support address ranges too (why not both!?!)
- Probably will pick SAN ID's as high order bits not all zero, LSb's zero
- Allows routing tables to be smaller
- Recognizes that L2RH conversion in final HR of SAN can do cheap translation of L2RH to local SR

## Capabilities — Exceptions and Usual Case

- In Routing Table, Capability Exceptions (e.g., disk nodes) listed first
- More general capabilities listed second, encompassing earlier nodes
- Order of hits in table significant
  - Disk nodes are 1000-1003
  - Video nodes are 1004-1010
  - General Purpose Nodes are 1000-1999

## Symbols

- Routing configuration information (*e.g.,* twin-HR-to-twin-HR)
- Extensibility of node and HR behavior (*e.g.,* allow add on trailers to be indicated)
- Interleave with L2RH's to allow multi-level "instructions" to pass between HR's
- In-prefix, serve to "instruct" recipients
- In EEP-suffix, act as TLV options (discussed separately)
- Want to experiment with symbols in implementation before formalizing request for inclusion (except options, which we propose to include as of now)

## Options (type-length-value)

- One bit of the EEP header is reserved, indicating options (f=1)
- With f=1, options are permitted
- A count option allows two or more to be supported, as an escape
- An option is at least 64-bits long, promote continued 64-bit alignment
- Useful for providing "out-of-band" info to higher layers (*e.g.,* datatype info to MPI)

## Change to enhance priority support

- Version and Priority appear in EEP, L2RH, symbols
- Priority is early and ubiquitous
- 6 priority bits in L2RH's provide for early decision-making
- This will be useful in the realtime MessageWay implementation

## Implementation Status and Directions

- MSU Reference Implementation Underway
- Base implementation to build on UDP
- Second implementation target is heterogeneous SAN network
- Realtime experiments to follow

## Heterogeneous SAN network @ MSU

- Myrinet
- ATM
- Two Myrinet SAN's connect through ATM
- First UDP, later "native"

## MSU MsgWay API's

- Tier-1 API to mimic standard closely (unreliable, lots of inquiry)
- Tier-2 API to provide end-to-end reliable transfer, plus additional functionalty
- Tier-3 API's (Security, Realtime) to be explored
- Not part of protocol or standard (ideally, to be FYIs)

## Summary and Conclusions

- Certain extensions greatly expand MsgWay functionality
- Interdependent extensions are being discussed and proposed
- Enhancements to RRP as well as EEP are included
- The only controversial issue is symbols, which will be fleshed out in course of implementation effort, and discussed online at next meetings
- MSU reference implementation underway