Reliable Multicast for the Grid?

Marinho P. Barcellos[1]
marinho.barcellos@man.ac.uk

University of Manchester and BT Research Labs (UK)

[1]on leave from Unisinos (Brazil)

Demand for Multipoint Communication

- Many distributed and networked applications in GridC require multipoint communication [use cases]
- Key aspects: 1-N x N-N, data x real-time [ietf rmt wg]
- Performance is a fundamental requirement in GridC [net issues]
- GridC applications demand certain QoS from a multipoint service:
  - meet the degree of reliability required
  - achieve best goodput possible without being "unfair"
- Q: could we not achieve this multipoint service with unicast (using multiple TCP and plain UDP flows)?

maybe.

Reliable Multicast for Multipoint

On one hand, people have used TCP for multipoint

On the other, research has long claimed benefits of reliable multicast protocols

But little has been experimentally investigated, because of the challenges associated

We are experimentally comparing state-of-the-art implementations of reliable multicast, having requirements of HP Grid applications in mind

MUST project investigates reliable multicast issues in GridC (data, real-time) - RealityGrid

Evaluating One Case of Multicast in GridC

Simplest but most representative scenario: “I want to send this huge file to those machines. Right Now!”

# SuperJanet – UK academic network
  $ core and some regional nets with multicast
  $ “high” bandwidths (70/95 Mbps end-to-end, TCP/UDP)
  $ assumed generally “low” loss rates, RTTs 5-20ms

State-of-the-art protocol implementations
  $ MDP, NORM (NRL), NORM (INRIA), JRMS
  $ compared against TCP

What have we found?

problems.

...network-wise...

IP multicast infrastructure is fairly less robust than the unicast network

Bad connectivity: lack of symmetry and transitivity

Extreme session delays (frequently 2-3min) at startup

Unexpected, massive reordering of packets

Potential causes? routing pathologies, IGMP, firewalls and end-host problems

...transport-wise...

Reliable multicast implementations currently cannot deliver robust service

Network cost (bandwidth) and processing cost at end-hosts is indeed much lower with reliable multicast than using N times TCP

However, current implementations cannot meet requirements of GridC of high-performance

Three possible explanations:
  - $ infrastructural problems in the network (as noted)
  - $ design problems in the protocols (scalability)
  - $ implementation problems in current version (ongoing)

...transport-wise...

This is true even in more controlled environments (Gigabit LAN)

Performance depends heavily on the input parameters

When flow/congestion control is active, performance is very low (excessively conservative)

(Omitted plots for brevity)

Future Steps?

Application-Level Multicast: fine for low-speed replication

Group Communication protocols may help with distributed resource management

"Path-oriented" may be feasible only with multicast

Improve multicast monitoring and management technology

Research groups and IETF have focused on highly scalable reliable multicast (of bulk data)

Future Steps?

Issues: tension between “one site does not fit all” [ietf] and standardized, less complex interaction [grid]

Possibly create a new protocol instantiation of the NORM class towards GridC

Make this flexible NORM an efficient multipoint service in Globus middleware

We need protocols that work like TCP in LANs/WANs
  - $ provide abstractions similar to TCP for data transfer
  - $ and also are TCP-friendly

Discussion to be conducted within GHPN scope?

Thanks

All the people and institutions that have helped providing the necessary infrastructure

These include: Peter Coveney, Matt Harvey, Simon Clifford, Tom Crummey, Stephen Pickles, Robin Pinning, Mark Hewitt, Steven Newhouse, Jonathan Giddy, Mike Axford, Mick Russel, Mark Leese, Paul Watson

UKERNA/SuperJanet, University College London, University of Newcastle, University of Cardiff (WeSC), Southampton University, Imperial College, Cambridge University