Scalable Session Messages in SRM

talk by Sally Floyd
http://www-nrg.ee.lbl.gov/floyd/

RM Research Group, September 19-20, 1997

Research by Puneet Sharma with Deborah Estrin,
Sally Floyd, and Lixia Zhang.

The role of session messages:

1. Information about state (e.g., the highest sequence number seen from each sender).

2. Used for estimating round-trip times between members.
This roundtrip time estimate is used for
• setting request timers
(based on estimated distance to sender),
• setting repair timers
(based on estimated distance to sender of request, and
• setting ”ignore requests” timers
(after sending or receiving a repair).

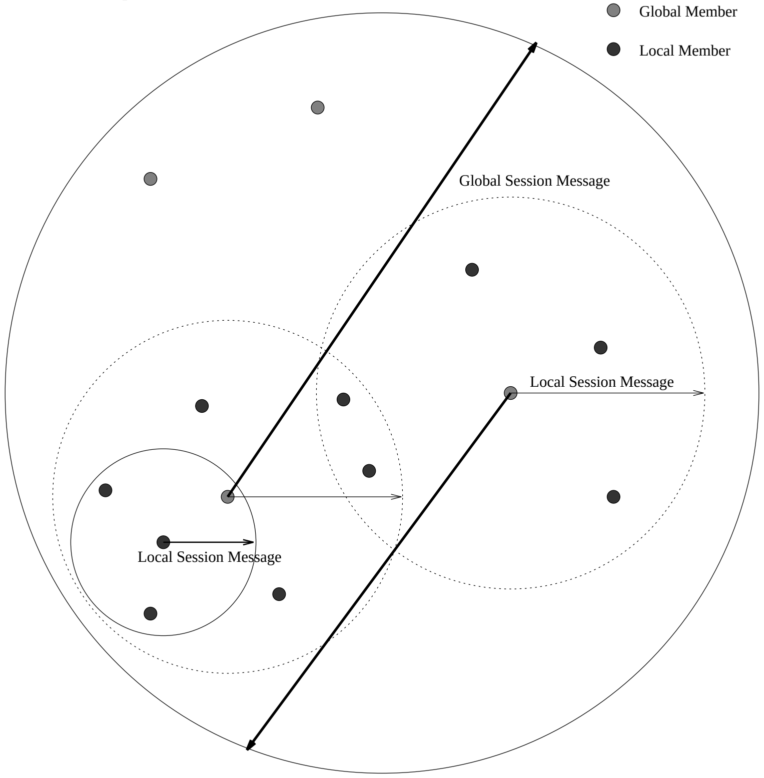
Limitations of global session messages:

1. For very large sessions, a member can't keep the estimated roundtrip time to every other member in the group.

2. Because the overall bandwidth used by session messages is limited, in large groups each member would be sending very large global session messages VERY infrequently. This could cause delays in delivering state information.

Possible approaches to scalable session messages:

1. Global representatives and local members. Self-configuring algorithms for choosing the global representatives.



2. Overlapping pools of local session messages. Problems: longer delays, difficulties in calculating RTTs.

Global representatives:

1. Between N and 2N members send global session messages (for N large). These members also send local session messages with sufficient scope to reach all of their local members.

2. Other members just send local session messages, with sufficient scope to reach their global representative.

How it would work for estimating RTTs:

1. Global representatives include in local session messages their estimated RTT to other global representatives.

2. Senders of data and senders of requests and repairs identify their global representative.

3. To first order, estimate distances by the distance between your global representative and their global representative.

Self-configuration of hierarchy:

The self-configuration algorithm is based on random timers and on self-evaluated appropriateness to change status.

1. New nodes listen to see if there are N members sending global session messages. If so, then they decide to be a local member.

2. Deciding to be a local member: Pick the nearest global representative, send local session messages with sufficient scope to reach that member. (You know that you have chosen sufficient scope if you see your local session message “reflected back” in the global representative's local session message.)

Self-configuration of hierarchy, cont.:

3. More than 2N global representatives: some global representative has to independently decide to be local. Use randomized timers based on self-evaluated appropriateness for becoming a local member.

4. Less than N global representatives: some local member independently decides to become a global representative. Use randomized timers based on self-evaluated appropriateness for becoming a global member.

Self-configuration of hierarchy, cont.:

5. Local members moving from one global representative to another.

6. Changing the global representative in a local group.

7. Deciding to change global/local status when there are between N and 2N global representatives.

Report on Puneet's simulations.