A Reliable Multicast Framework for Light-weight Sessions and Application Level Framing

Paper review by: Ehsun Darody
810181050

SRM (Scalable Reliable Multicast) is a reliable multicast framework for light-weight sessions and application level framing. It scales well to both very large networks and very large sessions.

SRM's design follows ALF protocol model. SRM is also heavily based on the group delivery model that is the center-piece of the IP multicast protocol. It requires only the basic IP delivery model and in a fashion similar to TCP adjusts its control parameters based on the observed performance within a session. Fate-sharing coupling of sender and receiver in unicast transmissions doesn't generalize to multicast. Thus, we use receiver-based reliability. Also, using shared communication state to name data doesn't work well in the multicast. So, we use naming in application data units (ADUs) model.

Each member of the multicast group is individually responsible for detecting loss, generally by detecting gap in the sequence space. Each member generates low-rate periodic session messages to detect last packet loss. When a host detects a loss, it schedules a repair request for a random time in the future. When the timer expires, host multicasts a request to the group. The timer is chosen from the uniform distribution [C1*d, (C1+C2)*d] seconds, where d is the host's estimate of one way delay to the source. C1 and C2 are parameters. When some host receives a request that is capable of answering schedules and sends repair message in the same manner (uniform distribution [D1*d, (D1+D2)*d]).

The simplest congestion control mechanism for SRM would be for all members of the multicast group to assume a fixed bandwidth constraint over the aggregate session.

SRM's reliable delivery can be used in a protocol layer such as one depicted in Figure 1. Note that the SRM, itself doesn't maintain messages ordering. A higher layer protocol would be implemented to do this task, if required.

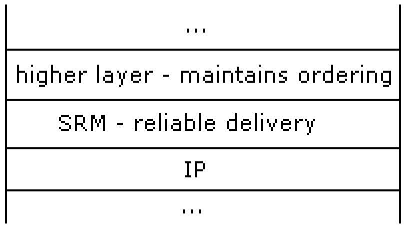

Figure 1

To test performance of the request-repair algorithm we simulate it on different network topologies and session sizes with constant parameters (C1, C2, D1, and D2).

In a chain topology C1 and D1 are set to 1, and C2 and D2 are set to 0 that leads to deterministic request timers as a function of distance to the source. Due to deterministic suppression there will be only one request and one repair messages.

For a star topology, all nodes detect a loss at exactly the same time. The essential feature of the loss recovery algorithm in a star is the randomization used to reduce implosion that is probabilistic suppression. C1 and D1 are set to 0 and C2 and D2 are set to a value greater that C1 and D1.

The loss recovery performance in a tree topology uses both the deterministic and probabilistic suppression. It depends principally on the distance of the sender from the congested link and on the ratio between the C2 and C1.

To extend the basic request/repair algorithm, we use an adaptive algorithm that adjusts the timer parameters as a function of both the delay and of the number of duplicate requests and repairs in recent loss recovery exchanges. In this fashion the algorithm can adapt the timer parameters not only to fit the generally fixed underlying topology, but also to fit a changing session membership and pattern of congestion.

Local recovery is used where the neighborhood affected by the loss is small. Thus, the bandwidth costs of the loss recovery algorithm is reduced due to multicast requests and repairs to a limited area. Available mechanisms for local recovery are "Administrative scoping", "Separate multicast groups" and "TTL-based scoping".