

Hierarchical Geometric Overlay Multicast Network

Eng Keong Lua, Jon Crowcroft
{eng.keong-lua, jon.crowcroft}@cl.cam.ac.uk

Xiaoming Zhou, Piet Van Mieghem
{x.zhou, p.vanmieghem}@ewi.tudelft.nl



Motivation:

1. Overlay Multicast Networks does not require any network support except the network unicast capabilities, thus allowing service diversities, accelerated service deployment, with a greater flexibility.
2. To construct efficient overlay multicast topologies, we cannot ignore the dynamic underlying network behaviour that will incur unnecessary extensive measurements to maintain quality of connectivity causing scalability problem.

Our Proposal:

With this motivation, we attempt to use geometric overlay network to implement efficient overlay multicast and a hierarchy structure to scale the multicast:

1. Geometric position and distance information in a geometric space are assigned to overlay nodes by mapping them onto geometric plane virtually using scalable network embedding techniques to embed underlying network performance metrics, such as delay latency or available bandwidth. This simplify the construction of spanning trees and multicast routing.
2. Hierarchical structure consisting of 2-tier SuperPeers and Peer layers to construct efficient multicast delivery trees. Multicast trees' connectivity at the upper SuperPeers layer provides the high-bandwidth multicast backbone infrastructure to serve the lower Peers layer.

Bos: A Hierarchical Geometric Overlay Multicast Network built on Lightweight SuperPeers Topology (LST) geometric overlay network. It consists of 2-tier hierarchical SuperPeers (Yao-Graphs Topology Connectivity) and Peers layers to provide good multicast routing scalability, efficiency and topology manageability.

2-Tier Hierarchical Lightweight SuperPeers Topology (LST) Overlay Network

Step 1 SuperPeer Election:

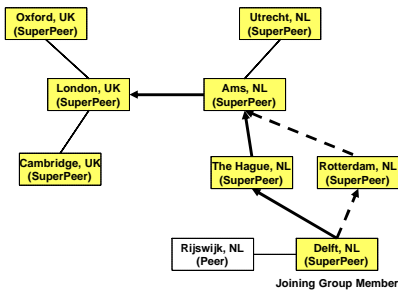
Overlay nodes are elected as SuperPeers based on resources availability, reliability and stability. SuperPeers layer acts as a high-bandwidth backbone infrastructure for the nodes in the Peers layer.

Step 2 Highways:

Highways is an overlay network control plane service that performs scalable network embedding to map overlay nodes in metric space onto geometric points in geometric space. Network superspace embedding is done to embed overlay nodes' RTT into Global geometric coordinates. Clustering is done to cluster overlay nodes, and network subspace embedding is performed within clusters for Local geometric coordinates. Clustering is done to cluster overlay nodes, and network subspace embedding is performed within clusters for Local geometric coordinates. Clustering is done to cluster overlay nodes, and network subspace embedding is performed within clusters for Local geometric coordinates. Clustering is done to cluster overlay nodes, and network subspace embedding is performed within clusters for Local geometric coordinates.

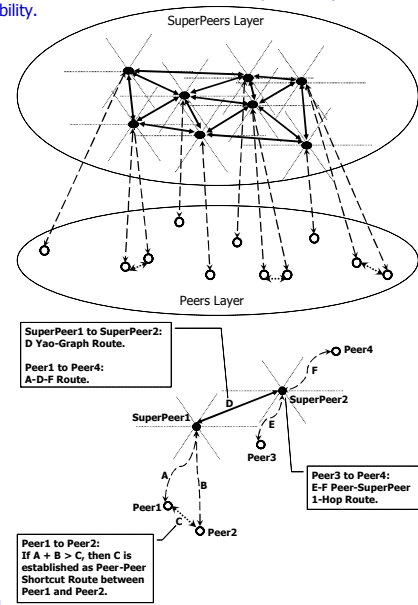
Step 3 SuperPeers and Peers Topology Construction:

In the SuperPeers layer, we use Yao-Graphs to construct the overlay network connectivity among the SuperPeers by using their geometric position information. Geometric positions are computed from Step 2. Since the geometric space around the SuperPeer is cut into six sectors with an equal angle $\theta < \pi/3$, every SuperPeer chooses the six closest SuperPeers in terms of their geometric distances to connect to. These SuperPeer-SuperPeer Yao-Graphs routes serve as the reliable high-bandwidth backbone network connectivity. In the Peers layer, Peers are directly connected to the first closest SuperPeer that is capable of serving an additional Peer. This connectivity is called the Peer-SuperPeer 1-Hop route. If there exists a shortest Peer-Peer Shortcut route between two Peers belonging to a SuperPeer, direct connectivity between these Peers are established. A combination of greedy and face routing is designed as the overlay routing protocol.



Bos: Hierarchical Geometric Overlay Multicast Network

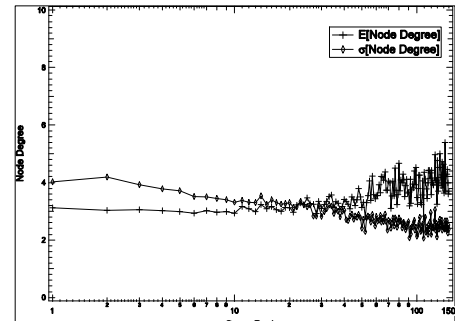
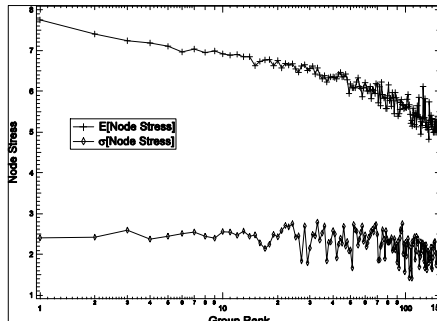
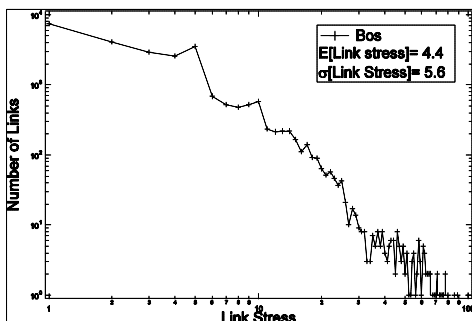
- Bos is built on LST overlay network and uses tree topology for multicasting. The multicast tree is formed by the union of source-destination overlay geometric routes.
- Bos builds geometric shortest-route multicast distribution trees among the SuperPeers in the SuperPeers layer of the LST overlay network using an adapted version of Reverse Path Forwarding (RPF) algorithm. This multicast tree construction method ensures that each SuperPeer member of multicast group receives a multicast message on the correct incoming connectivity. We build a multicast distribution tree at SuperPeers layer such that each group member has shortest geometric overlay route back to the root. The shortest-geometric-route tree is constructed on the Yao-Graphs topologies at the SuperPeers layer which could exhibit Minimum Spanning Tree characteristics.
- The multicast tree constructed at the upper SuperPeers layer provides the reliable high-bandwidth backbone multicast connectivity for the lower layer's Peers who basically connect to their SuperPeers using their direct Peer-SuperPeer 1-Hop route for overlay multicasting. If the Peer wishes to be the member of the multicast tree, the SuperPeer serving this Peer will be responsible to "Connect" to the multicast tree.



A New Multicast Group Member Joining Bos's Multicast Tree

We describe an example of the multicast group membership management operation by multicast group members (SuperPeers at SuperPeers layer) in Netherlands (NL) and United Kingdom (UK). London acts as the root of the multicast tree and a normal Peer, Rijswijk, has Peer-SuperPeer 1-Hop connectivity to the SuperPeer, Delft:

1. Delft is joining the multicast group and it sends a "Connect" message to the multicast group. Bos selects the closest parent in the multicast tree and forwards the message to that parent (The Hague). Delft's information (such as IP address and port number) are stored in its parent, The Hague's Children Table. And, the Delft's ancestor list is updated with the root to its parent, i.e. London (root) and Amsterdam (ancestor). The multicast links are shown by the solid arrows. In this way, the new group member, Delft, has joined the Bos's multicast tree with London as its root. And, its normal Peer, Rijswijk, who requires multicast service is served by its SuperPeer, Delft, in the multicast group.
2. If Delft detects that its parent, The Hague, has failed due to missing heartbeats, then it initiates the local ancestor-recovery protocol and sends a new "Connect" message to the list of closest neighboring ancestor using LST overlay geometric routing mechanism to re-join the multicast tree. On receiving the "Connect" message, Amsterdam will provide the information of its children in its Children Table to Delft. With the geometric position information of Amsterdam and its child (Rotterdam), the closest Rotterdam is chosen as the new closest parent. The new closest parent, Rotterdam, will update its Children Table of this new SuperPeer child, Delft, in the multicast group. The new multicast links are shown by the dash arrows.
3. We have illustrated in this example the efficiency and scalability of the joining multicast group member to find its parent with rich multicast resources within the same country with close proximity. If Delft is leaving the multicast group, it sends a "Disconnect" message to its parent in the multicast tree, and its information will be removed from its parents' Children Table.



Experimental Results and Performance Evaluation

We use the 10 large-scale Georgia Tech (GT) network models, each consisting of 100,000 nodes (used by Scribe and SplitStream). A Zipf-like distribution for differing group sizes is adopted. In each network model, there are 150 groups (i.e. group rank 1 to 150). In each group, there are 10% elected SuperPeers that have high capacity, exhibit stability and reliable connectivity.

Link Stress is the number of duplicate packets carried by each physical link in overlay multicast in each group.

Node Stress is the number of messages received by nodes in each group.

Average Underlay Node Degree is the average number of underlay multicast connectivity of nodes in each group.