T-Man: Gossip-based Overlay Topology Management

Rakesh Gopinath Nirmala Arunakumari Yedurupaka

I. Motivation

One of the important considerations to build large scale robust networks is to design a better topology that impacts various network functions. In large and dynamic peer-to-peer networks, overlay topologies play a vital role in providing high-performing, reliable services. With churn and huge network sizes, P2P networks require efficient and robust algorithms to construct, maintain and optimize the topology. Many specialized topologies and protocols are proposed to achieve higher performances with respect to various functions such as entities discovery, routing, information dissemination, data aggregation etc. However, none of these are generic enough and support on-demand topology changes. T-MAN is one such generic protocol to cherish the aforementioned needs.

II. Contributions

The primary contribution of the authors is a faster and more generic topology management protocol specifically for large scale, dynamic systems that supports changing the underlying topology at run time thus enabling automatic topology evolution. The other striking feature is it uses a simple and general ranking function combined with gossip communication of local state information making it more adaptable to networks of different dimensions and supporting a wide range of topologies.

III. Solution

The proposal is to construct a desired target topology efficiently based on a ranking function wherein every node is allowed to sort any subset of nodes (potential neighbors) according to a preference to be selected as their neighbor. The ranking function could be a simple distance-based metric or any relevant measure. Each node in a network is identified by an address and maintains a set of other nodes using partial views (or views). Views are sets of c node descriptors and node addresses. The node descriptor represents the profile (information) of the node such as ID, physical location, semantic description of content stored at the node etc. The links between the nodes are defined by the addresses and the parameter c in view defines the degree of the node. The degree is uniform for all nodes.

The protocol implementation includes two threads responsible for carrying major operations. One active thread for initiating the communication with other nodes and one passive thread waiting for the inbound messages. The same protocol runs on all the nodes. Using the ranking function and the node addresses in the current view, first c elements are periodically exchanged and thus neighbors are chosen in every round. For this, T-MAN uses an extremely robust peer sampling service called NEWSCAST protocol wherein each nodes exchanges their random views and update their local views at regular intervals of time thus forming a new random sample. This randomness is particularly important in large scale networks that adds long range links thereby reducing the time taken for network convergence.

Further, to optimize the protocol, connection limits and node hunting in each cycle are implemented. To avoid exchanging entire information, only first c items are exchanged after ranking and sorting. Moreover, to operate efficiently in presence of churn, a certain number of old descriptors are removed in every cycle of operation. This is called Self-healing implemented using the age field in the node descriptors. The topology gradually stabilizes in every cycle learning through the views of its current neighbors in each cycle. It is to be noted that during a cycle, each node is updated once on an average.

IV. Strong Points

1) The protocol has greater potential due to the generality in its approach and the flexibility it offers to change the topologies at run time on demand making it more adaptable for diverse networking environments.
2) Also, the protocol is highly useful for large scale, dynamic networks due to its capacity to converge faster - convergence time increases as the logarithm of network size - thus achieving the target topology in lesser number of cycles (time period).
3) The protocol is highly scalable, robust and simple with the general ranking function and results in greater performance with proper healing chosen as evident from the evaluation results and simulations.

V. Weak Points

1) On whole, the paper does not discuss on ways for choosing an optimal ranking function among many possibilities for a given problem or an environment.
2) It also does not mention any ways for optimally choosing a better c value for the exchange of number of node items for different feasible topologies.
3) One more open ended question is it does not specify what is the dependency of the convergence and convergence time on various topologies.