T-Man: Gossip-based Overlay Topology Management

Bernardo Gonzalez Riede Xin Ren

I. Motivation

The author clearly stated the motivation of the paper. In large, dynamic, fully distributed systems, such as peer-to-peer (P2P) networks, overlay topology forms the basis for, or has a major impact on many functions. It is well known that functions such as searching, routing, information dissemination, data aggregation, etc, need special topologies for good performance and high efficiency. Furthermore, solutions to other problems including sorting and clustering can be readily expressed as topologies. So it is necessary to have efficient and robust algorithms to create, maintain and optimize the topology, especially when there is high churn rate in the system.

II. Contributions

The contributions of the paper is also quite clear. A generic protocol, T-Man, for topology management is proposed which is robust, scalable, flexible, simple, general and fast. It is simple to implement, debug and understand. It is flexible enough to allow for changing the managed topology at run time on demand, without having to develop a new protocol for each possible topology from scratch, and also supports quickly changing topologies. It is general enough to support a wide range of different topologies and be applied as an off-the-shelf component for prototyping or even as a production solution that could be implemented even before the final desired topology is known. Finally, with logarithmic convergence times it is fast so that it is possible to construct a topology quickly from scratch (recovery from massive failures or bootstrapping other protocols on demand) or where topology maintenance is in fact equivalent to the continuous re-creation of the topology (for example, due to massive churn).

III. Solution

The proposed solution is suited for large systems for it is based on gossip communication. Moreover, it's impact and need is inversely proportional to the desired amount of neighbours c which is uniform. On a high level, T-Man works by applying a *ranking* function to a changing set of known nodes, thereafter selecting the first c nodes as neighbours. This ranking and rewiring happens periodically which allows for the introduction of *cycles* as a notion for measurement of convergence.

When a node q initiates an exchange, it communicates with its first neighbour p according to the ranking function to be able to generate a set of know nodes composed of:

- Set of neighbours of q.
- Set of neighbours of p.
- Set of random nodes known by p through a *peer sampling service*[1]

This selection of a new set of neighbours can happen as a result of an *active* or *passive* exchange. After a configured time T a neighbour starts an exchange actively while it reacts to some passively. A cycle is therefore defined as T/2, each node participating in one active and one passive exchange on average in each cycle.

The ranking function is deeply tied to the desired outcome and needs information, i.e. a profile, about each node. The profile needed for ranking has to be included in the information exchange by the nodes. T-Man being flexible, the entity deploying the topology has to carefully develop its necessary ranking. A common ranking used is distance based ranking, but can create problems if no refinement is done when using highly clustered nodes. Depending on the size of c, it may happen that the nodes connect only to other nodes in the cluster thus not knowing about the rest of the network. To be able to cope with churn, the generic component is extended to provide self-healing. Self-healing takes place by including an age field in the profile which has to be considered while ranking the known nodes. Finally, H last (oldest) nodes are dropped as set of known nodes. The experimental results indicate a value of 1 for H to be often the optimal one.

IV. Strong Points

1) References are well listed and noted.
2) Detailed description on the problem makes it easy to understand.
3) When viewing the paper digitally, the graphs in Fig. 3 are intuitive through the use of patterns to group related graphs.
4) Treating the very important clustering issue and providing a solution to it.

V. Weak Points

1) It would be more convincing if there is data or graph provided for the optimizations of the solution.
2) On printed paper, the graphs in Fig. 3 are difficult to distinguish.
3) Self-healing reads like an optimization and should be mentioned in section 3.2.

REFERENCES

[1] Márk Jelasity, Rachid Guerraoui, Anne-Marie Kermarrec, and Maarten van Steen. The peer sampling service: Experimental evaluation of unstructured gossip-based implementations. In Hans-Arno Jacobsen, editor, *Middleware 2004*, pages 79–98, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.