Cluster Entries for Semantic Organization of Peer-to-Peer Network*

Nicolas Lumineau[1], Anne Doucet[1], and Bruno Defude[2]

[1] LIP6 - 8, rue du Capitaine Scott 75 015 Paris, France.
{Nicolas.Lumineau,Anne.Doucet}@lip6.fr
[2] GET-INT - 9, rue Charles Fourier 91011 EVRY cedex, France
Bruno.Defude@int-evry.fr

1 Context

Techniques of pure flooding queries through a peer-to-peer network[1] reach their limits for data localization which are not replicated and stored through a high number of nodes. To improve implicitly the query routing towards relevant nodes in order to smartly recover all the relevant data, we propose to organize a peer-to-peer network in which the logical neighborhood of each node is built according to the semantic of its content. Indeed, the data we target, are semantically rich, easily classifiable with a field of interest, or a theme (e.g. "rock", "jazz", … in music, or "hydrology", "oceanography",… in sciences of environment). The motivation of our approach is based on the fact users/providers query the network on the theme of the data they store. Thus, if a node contains many data about the theme "Hydrology", the query propagation is more efficient, if the neighbors of this node contain data about "Hydrology". In the prototype VENISE (serVicE for Node Insertion in Semantic clustErs), we propose a protocol of node insertion based on cluster entries to cluster a peer-to-peer network.

2 Semantic Representations

To build a network organized according to the semantic content of nodes, we define

- a set of *theme* statically established and used as common referential to semantically compare nodes.
- the semantic representation of each node, named *Thematic Vector.* The values of this vector contain the proportion of data stored on the node according to each theme.
- the semantic representation of each cluster, named *Aggregation Vector.* The values of this vector are learnt by a Machine Learning technique to take into account the aggregation of Thematic Vectors of nodes contained in the cluster and the semantic distance between clusters. Thus, a neural network is used as in the approach of self organizing maps[2] to classify nodes in cluster.

* This research is done in the context of the PADOUE project (http://www-poleia.lip6.fr/padoue) financed by ACI GRID of French Research Minister (htp://www-sop.inria.fr/aci/grid/public).

M. Bouzeghoub et al. (Eds.): ICSNW 2004, LNCS 3226, pp. 319–320, 2004.

– An abstract entity to point on the node in charge of the physical insertion of a new node in the network, named *Cluster Entries*.

3 Protocol of Node Insertion

The process of node insertion is performed in three main steps:
– The relevant cluster entry selection: the thematic vector of the new node is compared (according to the least square method) with the aggregation vector of each cluster entry. This allows selecting the most relevant cluster entry.
– The physical node insertion: the node pointed by the selected cluster entry handle the physical insertion of the new node in the network (i.e. it is used as entry in the network).
– The cluster entries updating: the insertion of a new node in a cluster changes the content of a cluster. Thus, according to an algorithm of machine learning, the aggregation vectors of all cluster entries are updated.

The poster shows how machine learning technique can be used to organize a peer-to-peer network. Our main goal is to exploit the most relevant semantic in each level in a peer-to-peer network. To improve implicitly query propagation process, the semantic of node content is used to organize the network in order to reduce the number of hops necessary to resolve queries. In this way, to further improve the algorithm of query propagation we can use another semantic as experiences of user communities [3]. This allows to rout queries straightforwardly into relevant nodes with query solutions.

References

[1] Aberer, K., Hauswirth, M., Peer-to-Peer Information Systems: Concepts and Models, state-of-the-art, and Future Systems, Tutorial IEEE ICDE, 2002.
[2] Kohonen, T.,Exploration of very large databases by self-organizing maps. In Proceedings of ICNN'97, International Conference on Neural Networks, pages PL1-PL6. IEEE Service Center, Piscataway, NJ.
[3] Lumineau Nicolas, Doucet Anne, S*haring Communities Experiences for Query Propagation in Peer-to-Peer Systems,* In 8th International Database Engineering and Applications Symposium (IDEAS'04), 7-9 Juillet 2004, Coimbra, Portugal.