DDC: A Dynamic and Distributed Clustering Algorithm for Networked Virtual Environments based on P2P Networks

Moritz Steiner* and Ernst W. Biersack
Institut Eurécom
Sophia-Antipolis, France
{steiner,erbi}@eurecom.fr

ABSTRACT

This paper presents a dynamic and distributed algorithm for the clustering of peers in a Networked Virtual Environment (NVE) based on a fully distributed peer-to-peer (P2P) network.

The main idea is to classify connections in short intra-cluster connections and long inter-cluster connections. The insertion of new peers or the deletion of existing peers can result in the merging of two clusters or the split of one cluster in two parts.

Categories and Subject Descriptors: C.2.1 [Computer-Communication Networks]: Network Architecture and Design — *Network topology, Distributed networks*

General Terms: Algorithms, Performance

Keywords: NVE, P2P, social networks, clustering

1. INTRODUCTION

Networked Virtual Environments (NVEs) are computer generated, synthetic worlds that allow simultaneous interactions of multiple participants. Especially with the boom of Massively Multiplayer Online Games (MMOGs) NVEs are becoming popular nowadays. However, to create a large scale NVE among others the scalability problem of the commonly used client-server model must be solved.

Recently some peer-to-peer (P2P) architectures [4, 6] have been proposed to solve the scalability issue. Yet, using a P2P solution to build a large scale virtual environment brings up classical P2P problems. Typically in P2P-based NVEs each peer only knows its direct neighbors and some other peers in its Awareness Area. The only way to handle a query in such a scenario is to forward it to all neighbors, that is to flood the entire virtual world until a hit to the query is found. The flooding model requires a lot of bandwith and hence does not prove to be very scalable.

Our approach to cope with this problem is based on the properties of the social network of people who are using the system. A social network is a set of people with some interactions between them [5]. In recent years a number of properties such as degree distribution, the small world effect, the search ability and clustering have been identified. The assumption is that people interact in the virtual world as in the real world described by the social networks, and thus tend to form clusters as well. Hence, the peers have to have the ability to freely choose the coordinates in the virtual world. Consequently the probability that a request can be satisfied inside the same cluster is very high.

The idea is to identify these clusters and to limit the scope of queries to the respective cluster, and not to flood the entire network. Therefore connections between peers are classified in short intra-cluster connections and long inter-cluster connections. The insertion of new peers or the deletion of existing peers can result in the merging of two clusters or the split of one cluster in two parts.

This poster abstract proposes a distributed and dynamic algorithm for the clustering of peers (DDC) in a NVE based on a fully distributed P2P network. DDC is not only applicable to NVEs, but to all proximity graphs with a clustered repartition of vertices where only local knowledge is available.

*Moritz Steiner is also affiliated with the Department of Mathematics and Computer Science, University of Mannheim, Germany.

CoNEXT'05, October 24–27, 2005, Toulouse, France.
ACM 1-59593-035-3/05/0010.

2. RELATED WORK

Cluster analysis has been a research topic for decades. However, most of the algorithms only deal with static data and are centralized. For our problem, we need a distribued clustering algorithm that can handle join and leave of peers. Also, each peer that executes the algorithm only knows its direct neighbors and not all peers in the system. Eldershaw and Hegland [1] have proposed a clustering criterion using Delauny triangulation that can be used for distributed clustering: Based on the neighbors in the Delaunay triangulation, the neighbors closer than a distance p belong to the same cluster. However, fixing a universally valid distance p is not trivial.

Estivill-Castro et al. [2] suggested a more elaborated criterion: the *long-short* criterion. For a peer p_i, let $\mathrm{Mean}(p_i)$ be the mean length of connections from p_i to its neighbors in the Delaunay triangulation, and let $\mathrm{Dev}(p_i)$ be the standard deviation of the length of these links. Let p_i and p_j be two peers. Let $d(p_i, p_j)$ be the Euclidean distance of their connecting link. $p_i p_j$ is a "short" intra-cluster link if

$$d(p_i, p_j) < \mathrm{Mean}(p_i) + w \cdot \mathrm{Dev}(p_i)$$

else it is a "long" inter-cluster link.

The parameter w scales the granularity of the clusters globally. It does not depend on local conditions. To get satisfactory results w must be adapted to the overall node distribution, which is not known locally.

3. DYNAMIC DISTRIBUTED CLUSTERING

In this section we describe our dynamic and distributed clustering algorithm (DDC).

The *long-short* criterion has to be adapted to avoid errors in the cluster determination. In the version presented in [2] two Delaunay neighbors p_1 and p_2 may classify their link $p_1 p_2$ in a different way.

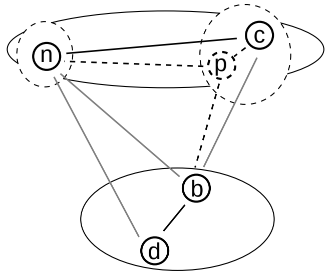

Figure 1: The solid lines show the links and the clusters before the arriving of p, the dashed lines afterwards.

p_1 may come to the result that it is a intra-cluster link, p_2 that it is a inter-cluster link.

Instead, DDC uses a weighted average of the local criteria of both peers: The link from p_1 to p_2 is a intra-cluster link if

$$d(p_1, p_2) < \frac{\text{Mean}(p_1) + w \cdot \text{Dev}(p_1) + \text{Mean}(p_2) + w \cdot \text{Dev}(p_2)}{2}$$

else it is a inter-cluster link. For the implementation this criterion is used. This criterion is bijective and therefore more stable than the original version.

The first and second peer in the NVE each form a cluster of their own. Each peer p maintains two lists: a list of its Delaunay neighbors $D(p)$ and a second one of its cluster neighbors $C(p)$. Moreover it stores the identifier of the cluster it belongs to $cid(p)$. Note that $C(p) \subseteq D(p)$ and that $C(p) \subseteq$ cluster $cid(p)$.

Every newly arriving peer p checks if it is near enough to its closest Delaunay neighbor c to join its cluster according to the *long-short* criterion. The other neighbors do not have to be taken into consideration. If p is close enough to c it joins the cluster of c, if not p forms a cluster of its own.

Due to the insertion of p, Mean(c) and Dev(c) change and hence c's cluster has to be eventually split up or merged with another cluster. c therefore performs a test with the *long-short* criterion on all its Delaunay neighbors $D(c)$, and checks if the result matches with the classification of $D(c)$ and $C(c)$. If it does not match, the lists are updated and the concerned peers check recursively their respective neighbors. Two cases must be distinguished, we focus on peer $n \in D(c)$:

1) $n \in D(c)$ and $n \in C(c)$ but according to the *long-short* criterion $n \notin C(c)$: n and c are too far away to be connected via an intra-cluster link. In this case it seems that the cluster $cid(c) = cid(n)$ has to be split (Figure 1). n randomly chooses an new cluster identifier $cid(n)$ and starts the recursive *split* call to $D(n)$. We now consider peer $a \in D(n)$ on the reception of the *split* message from n. a updates its cluster identifier to $c(n)$ and checks if a peer $a_i \in C(a)$ passes the *long-short* criterion. For the peers that passes the recursive call goes on. The peers that fail keep their old cluster number. If the cluster really needed to be split some of the cluster peers get the new cluster identifier $cid(n)$, the other keep the old one $cid(c)$. It might happen that c and n are too far away to be connected by an intra-cluster link but nevertheless are part of the same cluster because there exists a path of intra-cluster links connecting them. In this case the recursive *split* message find its way to c and all peers of the cluster have the new cluster number $cid(n)$, the cluster keeps its form an is not split up.

2) $n \in D(c)$ and $n \notin C(c)$ but according to the *long-short* criterion $n \in C(c)$: n and c are close enough to be connected via an intra-cluster link, but they are not part of the same cluster yet.

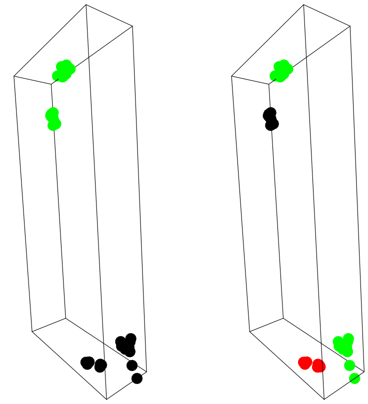

Figure 2: On the left side the clustering result with $w = 0$ (two clusters) on the right with $w = -0.6$ (four clusters).

n changes its cluster number $cid(n)$ to $cid(c)$ and tells every peer $n_i \in C(n)$ to propagate the new cluster number $cid(n) = cid(c)$ to their respective cluster neighbors $C(n_i)$.

4. SIMULATION RESULTS

To simulate DDC, first clusters of peers must be generated. For this purpose, the so called Lévy Flight was used. It generates distances between peers with the Lévy distribution [3]. The resulting coordinates result in a highly clustered peer distribution.

The first goal of the simulation was to gain experience with the *long-short* criterion and to find an appropriate value for w. Since no global view of the NVE exist, w value cannot be derived from the peer distribution but has to be chosen in advance. Values between 0 and -1 are reasonable for w. The simulations results show that for $w = -0,6$ the cluster size is of about 100 peers, for great numbers of total peers. Figure 2 shows an example with only 50 peers. That means the clusters are not to large to be flooded and not so small that no answer to a query can be expected.

5. CONCLUSION

This poster proposal presents DDC, a dynamic and distributed algorithm to cluster nodes in a P2P-based NVE. The motivation for the clustering is to define a natural limit for the scope of a query, in order to avoid flooding the entire network of peers.

Future work includes dealing in an efficient way with peer movements in the virtual worlds. Moreover DDC will be tested in an existing NVE like Solipsis to validate the basic assumptions on the behavior of peers.

6. REFERENCES

[1] C. Eldershaw and M. Hegland. Cluster analysis using triangulation. In *Computational Techniques and Applications*, pages 201–208. World Scientifi c, 1997.

[2] V. Estivill-Castro, I. Lee, and A. T. Murray. Criteria on proximity graphs for boundary extraction and spatial clustering. In *Proc. of the 5th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 348–357, 2001.

[3] M. Gutowski. L´evy Flights as an Underlying Mechanism for Global Optimization Algorithms. *ArXiv Mathematical Physics e-prints*, June 2001.

[4] J. Keller and G. Simon. Solipsis: A massively multi-participant virtual world. In *Int. Conf. on Parallel and Distributed Techniques and Applications*, pages 262–268, 2003.

[5] M. E. J. Newman, D. J. Watts, and S. H. Strogatz. Random graph models of social networks. In *PNAS*, volume 99, pages 2566–2572, 2002.

[6] M. Steiner and E. W. Biersack. A fully distributed peer to peer structure based on 3d delaunay triangulation. In *Algotel*, pages 93–96, May 2005.