Simulation possibilities of Semantic Class-based P2P in PeerSim environment

Ilya Rudomilov

Dept. of Computer Science, Czech Technical University, Technická 13, 121 35 Praha, Czech Republic

rudomily@fel.cvut.cz

Abstract. *This paper discusses the simulation possibilities of unstructured P2P semantic among PeerSim, which is one of the most popular envoironment recently, thanks to its modularity and ease of configuration. This is useful tool to evaluate theoretical model of our project "Semantic P2P Search engine": a heterogeneous network of independent nodes with clusters with text documents. Despite the fact that earlier PeerSim widely used in the modeling of structured P2P networks mainly (Chord, BitTorrent, Pastry, etc.), we used it in the unstructured context. The aim of project is to investigate the possibilities of protocols PeerSim to create nodes and basic communication between them into unstructured P2P network.*

Keywords

Class-based P2P, Simulation P2P, PeerSim, Semantic P2P.

1. Introduction

In recent years there has been a significant growth of interest in Peer-to-Peer architecture, especially for file distribution and VoIP-communication. Its benefits are obvious: scalability, high reliability, lack of need for each node of the global information network. PeerSim environment allows to model hundreds of thousands of independent nodes on a normal desktop PC with just 1 to 2 GB of RAM.

1.1. P2P background

Although usually the term "P2P" in a commercial made to represent fully decentralized, ad hoc network, in reality this is not the case, there are several types of P2P networks, and only a small portion is completely decentralized. As were defined in paper[8], types of P2P architecture are:

- Centralized P2P — the core of the network represented by the server, the network and coordinating with global information about all of its nodes. Examples of this type may be BitTorrent.
- Semi-decentralized (or hybrid) P2P — having multiple servers fixed at the center to distribute the load and ensure reliability. For example, the edutella network.
- Decentralized P2P — all nodes are independent and do not single out one or more central servers. An example of such a network can be called network Gnutella, our project is based on its ideas.

At the same time, the availability of information on other P2P sites can be classified as the following types:

- Structured P2P — has information about the other nodes according to a well-defined requirements. For example, this network Pastry[7], CAN, Chord[10].
- Unstructured P2P — the information does not have and generating search within the network, usually with the help of search. So random looking data on the network Gnutella.

Structured approaches are sufficient for IP-telephony and file-sharing, because there are sufficient a little data about objects (files) to identify them (name and surname of the author, date of publication, the file size). In this regard, researchers are working on semantic solutions for P2P: from networks with pre-defined onlologies to solutions with semantic descriptors.

2. Our project and its requirements

Our project [9] is about experimental class of P2P networks — semantic P2P search — search in text files, distributed amonng independent nodes. Bring them to the same format and can not identify unambiguously describe the content impossible manually specifying keywords, the author is not enough for efficient text search. Projects like Edutella is demanding clear requirements for storage of files, but we are close to the real needs of users and thus the restriction on the format of the stored data do not take into account.

Ultimately, our project is formed on the basis of the following techniques:

- Vector Space Model (VSM) organizes contains text documents in the form of matrices;
- Online spcherical k-means clustering (OSKM) groups the documents on the nodes in thematic clusters based on analysis VSM matrices;
- Gia defines handshake-protocol for independent nodes;
- Gnutella Efficient Search (GES) shared between semantically (thematically) similar nodes and different nodes;
- Class-based Semantic Search (CSS) introduces contrast GES group level (classes) within a single node.

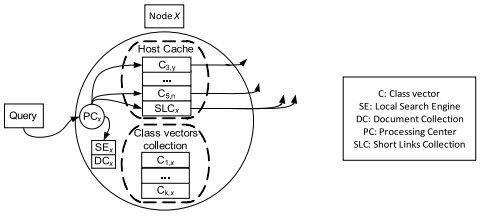

Fig. 1. Component of single node in proposed theoretical model of Semantic P2P Search engine

Accordingly, we can distinguish main functional requirements of our project to the environment simulation. Envoironment should support:

- unstructured P2P;
- heterogeneous objects;
- clustering;
- churn (continuously joining and leaving P2P networks by nodes).

PeerSim provides possibilities to implements all these requirements.

3. PeerSim

PeerSim[6] is lightweight environment, that was created by Mrk Jelasity, Alberto Montresor, Gian Paolo Jesi and Spyros Voulgaris from The University of Bologna in 2005. Since first years it has become popular with researchers from University of Bologna and Trento, Italy and later it was released under LGPL open source license.

As was defined in paper[3], the philosophy of PeerSim is to use a modular approach, as the preferred way of coding with it is to re-use existing modules. These modules can be of different kinds, for example there are modules which can construct and initialize the underlying network, modules which can handle the different protocols, modules to control and modify the network. PeerSim offers a lot of these modules in its sources, which ease greatly the coding of new applications.

We propose PeerSim to evaluate unstructured P2P models because of its several main characteristics:

- modularity — protocols are represented by additional packages;
- cross-platform — PeerSim is written in Java and requires it to support backend implementation of the protocol;
- easy configuration with simple ASCII files;
- extremely high scalability (up to 10^6 nodes in cycle-based mode).
- support of both P2P architecture (structured and unstructured);
- support of clustering (due to a list of "neighbors" inside Newcast object)
- capability to implement heterogenous implemented protocols, which is proven by several implementations:
  - Heterogeneous Cluster-based Newscast Protocol[4];
  - Heterogeneous Cluster-based Newscast Protocol Handling Multiple Resources[1].

At the same time, the environment has a number of disadvantages:

- the lack of a graphical user interface (GUI);
- the complexity of parsing of retrieved data.

The detailed review of several tools for P2P simulation can be found in paper[2].

3.1. Modes

PeerSim has 2 modes of simulation: *cycle-based* and *event-based*. Cycle-based mode is basic for PeerSim and very clear: all iteration are processed in cycle and engine are checking state of all nodes on every iteration. It is very fast and scalable solution (up to 10^6 nodes), but without any transport level interpretation, i.e. not so realistic. Event-based mode has more complex time management with communication support and rely on messages, i.e. communication with messages can be implemented. Undoubtedly, event-based mode is more realistic and realistic projects should be evaluated in event-based mode because of evaluation churn index.

3.2. Structure

The structure of PeerSim nodes are very close to real nodes in P2P networks. PeerSim has been used to evaluate structured models, but there are possibilities to adopt it for unstructured, even semantic approaches. PeerSim defines main classes:

- Network is represented by a list of nodes
  - Node has list of neighbours with their IP-addresses, resp. data about their content
- Initializers are executed before the simulation
- Controls are executed during the simulation to modify or monitor other components (create, destroy nodes, calculate clustering index, etc)

Generic structure of PeerSim node under paper[6] is shown on Fig. 2.

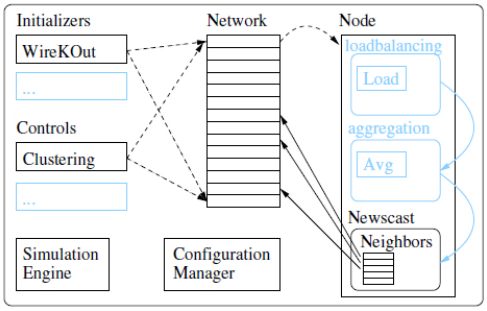

Fig. 2. PeerSim components, additional components are shown in light color

There are simillars with our project and other semantic-based P2P in aspect of using list of neighbours. Certainly, it is not container with semantic descriptors, but it would be used to store descriptors, e.g. in projects [4], [1].

3.3. Configuration

Configuration of PeerSim is provided by ASCII text file, which contents pair of options and their values. For instance, configuration file can be represented as:

```
random.seed 123456789
simulation.cycles 5
simulation shuffle

ovelay size 500
overlay.maxsize 3500
protocol.0 example.SimpleNewcast
protocol.0.cache 30

init.0 peersim.dynamics.WireRegularRandom
init.0.protocol 0
init.0.degree 30
```

4. Conclusion

PeerSim is modern and effective simulation tool for P2P networks. Wide range of researches use it due to relatively easy implementation of new protocols, high persofmance and good scalability. We pay main attention to projects [5], [4], [1] and their results indicates good results in semantic P2P area.

Hoewever, it is only simulation testbed completely non-hierarchical peer to peer networks and does not provide possibilities to implement network in details.

Acknowledgements

Research described in the paper was supervised by Assoc. Prof. Ivan Jelínek, FEE CTU in Prague and supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS11/129/OHK3/2T/13.

References

[1] S. Aslam, I. Kazmi, M. Javed, and M. Anwar. Cluster based peers configuration with multiple physical parameters using HCNP in peer-to-peer overlay networks. In *2010 International Conference on Computer Applications and Industrial Electronics (ICCAIE)*, pages 142–147, Dec. 2010.

[2] R. Bhardwaj, V. S. Dixit, and A. K. Upadhyay. An overview on tools for peer to peer network simulation. *International Journal of Computer Applications*, 1(19):74–81, 2010.

[3] B. Defude. P2P simulation with peersim. Feb. 2007.

[4] I. Kazmi, S. Aslam, and M. Javed. Cluster based peers configuration using HCNP in peer-to-peer overlay networks. pages 158–163. IEEE, July 2010.

[5] I. Kazmi and S. F. Y. Bukhari. PeerSim: an efficient & scalable testbed for heterogeneous cluster-based P2P network protocols. *International Conference on Computer Modeling and Simulation*, 0:420–425, 2011.

[6] A. Montresor and M. Jelasity. PeerSim: a scalable P2P simulator. In *Proc. of the 9th Int. Conference on Peer-to-Peer (P2P'09)*, pages 99 –100, Seattle, USA, Sept. 2009.

[7] A. Rowstron and P. Druschel. Pastry: Scalable, decentralized object location, and routing for large-scale peer-to-peer systems. In *Middleware 2001*, pages 329–350, 2001.

[8] I. Rudomilov and I. Jelinek. Semantic P2P search engine. In *Federated Conference on Computer Science and Information Systems 2011 (FedCSIS'11)*, page 991995, Szczecin, Poland, Sept. 2011.

[9] I. Rudomilov and I. Jelinek. Class-based approach in semantic P2P information retrieval. In *Proc. of the Federated Conference on Computer Science and Information Systems 2012 (FedCSIS'12)*, pages 257–261, Wroclaw, Poland, Sept. 2012.

[10] I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan. Chord: A scalable peer-to-peer lookup service for internet applications. *ACM SIGCOMM Computer Communication Review*, 31(4):149–160, 2001.

About Authors...

Ilya Rudomilov was born in Petropavlovsk-Kamchatsky, Russia. He took his master degree in Computer Science at the Irkutsk State Technical University, Russia. Actually he is a PhD student with project "Semantic P2P Search engine" under the supervision of Assoc. Prof. Ivan Jelínek at FEE CTU, Czech Republic.