Interest-based Self Organization in Group-Structured P2P Networks

Sardar Kashif Ashraf Khan [#1], Laurissa N Tokarchuk[#2]

[#]*Queen Mary, University of London, School of Electronic Engineering and Computer Science,,*

[1]ee07m094@elec.qmul.ac.uk

[2]laurissa.tokarchuk@elec.qmul.ac.uk

***Abstract*— Peer-to-Peer (P2P) Networks are becoming very popular these days. The scattered environment in P2P networks makes it different from the traditional networks. Each member in P2P network can act as a client as well as a server, which is different from its traditional counterpart such as internet, where a central server is required to control the network.**

Finding resources in P2P systems is a major issue. Extensive search may be needed to resolve queries in the network. In traditional P2P search algorithms such as in Gnutella protocol V0.4 search is via query flooding with significant exploration cost.

In this paper, a group-structured P2P network is proposed. The group-based search provides well-organized searching technique by targeting particular group or community. We propose a new protocol for building and repairing of overlay topologies based on the formation of interest-based superpeers. An interest-based superpeer algorithm creates groups or societies that have common interests. Simulations indicate that the proposed protocol is highly capable and powerful even in the dynamic nature of P2P networks, where nodes are continually joining and leaving the network and the protocol supports restoration of the network even after the catastrophic failure of interest-based superpeers.

***Keywords*— Interest-based superpeers, Capacity-based superpeers, Interest, Uniform, Uniform-interest, Power Law, Power Law-interest, neighbourhood relationship.**

I. INTRODUCTION

The idea of superpeer is emerging in P2P systems. For example, the popular P2P application KaZaA [1] implements superpeer architecture. The superpeer architecture is based on a two level hierarchical infrastructure. The nodes that are more powerful and trustworthy among a set of neighbouring nodes become a server and provide services to set of clients. Clients send their queries to their respective superpeer and receive replies from them. Superpeers are responsible for resolving queries on behalf of their clients by communicating with the other superpeers present in the network. Each superpeer maintains the list of contents shared by its clients.

In this paper, we present a new protocol to build and administer overlay topologies that is founded on an interest-based superpeer paradigm. An interest-based superpeer paradigm can be achieved by combining the concepts presented in [2, 3, 4] and [5]. The rationale behind deploying an interest-based superpeer structure is to create an overlay topology in such a way that relationships between nodes are based on the common interests. The term *Interest* can be considered in generic way such as movies, music etc. Here in this paper we assume that each node entering the network declares its particular interest. Nodes that have a common interest will form a neighbourhood relationship with the nodes of same interest and elect superpeers among themselves. The superpeers take server-like responsibilities for that particular subset of nodes and resolve queries on behalf of those clients. The resultant topology will create different groups based on interests over the overlay network. The benefit of creating interest based groups is to gather nodes of similar interests, which will not only ease the searching of contents but also require less number of queries to locate the contents.

The work presented in this paper proposes a modification to capacity-based algorithm proposed by [5], in order to create a superpeer overlay topology using the notion of interests. The protocol is based on the concept of gossiping [5, 6], in which each node starts information trading at regular intervals of time with a randomly selected peer node. During the trade, the nodes exchange information regarding their interests, current position: either superpeer or client and maximum capacity i.e. number of clients it can handle. On the basis of this information, the role of a node can be exchanged; a client could elect as a superpeer and start taking the job of other superpeers to ease the load. The superpeer may decide to become a client if it finds a more powerful client (in terms of bandwidth, processing power and memory) in its area and transfer its entire load to newly elected superpeer.

The resulting protocol is highly capable and powerful even in the dynamic nature of P2P networks where nodes are continually joining and leaving the network and restore the network even after the catastrophic failure of interest-based superpeers. Due to the continual gossiping of topology information, nodes are continually updating their database and when a new node enters in the network the nodes may learn about the newcomer by asking its identifier during information exchange. On the other hand the nodes that stop participating in the network or have crashed due to some reason will be gradually eliminated from the network.

The paper organised as follows: Section II provides the related work. Section III introduces the concept of an interest-based overlay topology and presents the proposed algorithm. Section IV provides the experimental results and discussion. Section V discusses the conclusion and future work.

II. RELATED WORK

The recent focus of researchers in P2P is to reduce the number of query search messages in order to effectively locate particular content in the network. For this purpose many algorithms such as [2, 3, 4] propose the concept of groups or clusters in P2P networks. A network is divided into different groups; each group declares a particular interest, thus forming communities of nodes having similar interest. The formation of communities increase the query success ratio as information among nodes in a particular community broadcasts more efficiently and thus reduces the overall network traffic.

The protocol presented in this paper is an extension of Montresor's work. In his work, the nodes with more powerful resources (in term of processing power, memory and bandwidth) are the suitable candidates for the role of server, whereas, less powerful nodes become clients. In any given P2P network there are different types of nodes in terms of processing power, storage and bandwidth, however, in his work, the superpeer selection is based on capacity i.e. number of clients a node can handle, so in this paper the term capacity-based superpeer is used to refer to his work. In capacity-based superpeer topology, the nodes exclusively play either the role of superpeer or client. The current role of a node is not permanent. If a node with larger capacity enters into the network; the present superpeer withdraws from its current status and becomes the client of newly elected superpeer. The neighbourhood relationship is based on the current job (either superpeer or client) of a node, if a node is acting as a superpeer then it forms the relationship with a random number of other superpeers and also makes the relationship with the set of clients for which it is responsible; whereas, the client is responsible for precisely one superpeer.

III. OUR WORK

In this section we present our interest-based superpeer overlay model design and the algorithm used to implement the model.

A. The Algorithm

For the formation of interest-based superpeer overlay topology, the essential points which must be considered are:

- Each node must declare its interest. For simulation, we represent the general interest such as movies, music etc with the whole numbers (0, 1, 2).
- Nodes must participate either as a superpeer or as a client, however the current role of the node is not static, it may change as higher capacity (cache size) nodes enter in the network.
- The node that acts as a superpeer, when finds a new node with higher capacity (cache size) in the network, the role changes: new node promotes into superpeer and old one demotes into client.

In interest-based superpeer overlay topology, the relationship between the superpeer and its clients are based on common interest, each client will register with exactly one superpeer, thus forming a group that contains nodes with similar interest. However, the neighbourhood relationship among interest-based superpeers is totally random and not necessarily based on common interest.

1) Building Blocks for Interest-based Superpeer Topology

The approach that is used to construct an interest-based superpeer overlay topology, based on the combination of two levels hierarchy approach:

- Underlying Topology
- Overlaying Topology

In underlying topology, all nodes present in the network will communicate with each other and share their interests, thus forming the groups in which neighbourhood relationship is based on common interests. The resultant topology is depicted in the figure 1 (b) below.

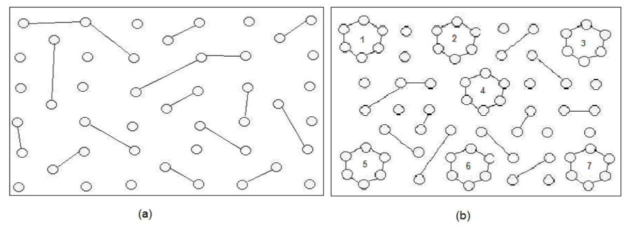

Fig. 1 (a) No. of nodes entering the network and formation of random connection (b) Underlying topology, nodes are forming interest-based groups.

The interest-based superpeer overlay topology shown in figure 2 works on the top of underlying topology, continually extract information of high capacity node from the underlying topology through information exchange and elect an interest-based superpeer among set of neighbouring nodes in underlying network. The thick lines depicted in the figure below show the relation between interest-based superpeers, whereas, thin lines show the relation between interest-based superpeers and their clients.

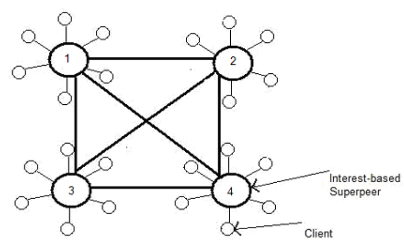

Fig. 2 Interest-based superpeer overlay topology built on underlying topology.

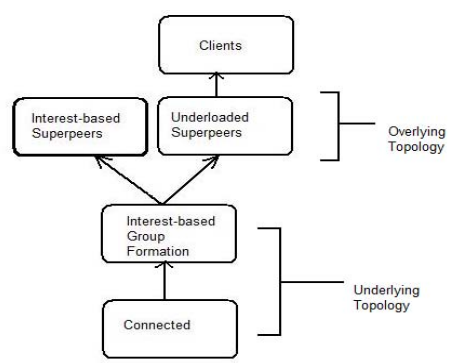

Fig. 3 The structural design of a node participating in the interest-based superpeer protocol.

Keeping in mind the underlying and overlying topology, the design of algorithm is divided into five sets of neighbourhood relationships that a node could have within the network, as shown in figure 3.

2) The Gossip Protocol

The gossip protocol which we used in the simulation is the modification of protocol already proposed by [5]. The figure 4 shows a generic gossip protocol [5] that works at every level shown in figure 3 but with little modification (as per requirement of the set).

```
Loop (forever)
Wait (time units)
R ← RANDSELECPEER( )
if (interest==R . interest) {
POSTSTATUS(R)
Rs ← receive (R)
UPDATE (Rs, R)}
else continue;
```

(a) **Initiator thread**

```
Loop (forever)
Rs ← receive (*)
POSTBACKSTATUS(R)
UPDATE(R, sender (Rs))
```

(b) **Listener thread**

Fig. 4 The pseudo code for modified gossip-based protocol. Notations: R is remote peer.

3) Underlying topology based on NEWSCAST

In NEWSCAST, each node consists of a set of *peer descriptors* of fixed size known as *partial view*. A peer descriptor includes the identification of node and *timestamp*. The purpose of timestamp is to maintain the timing information of descriptors i.e. time when they are created. The timestamp makes it possible to record the fresh information of peers all the time and keeps the network up to date. [5]

The *RANDSELECPEER ( )* method returns the address of a node, which is selected randomly from the current partial view. During the information exchange among nodes the *UPDATE* method combines the partial views, therefore generates the fresh partial view per information exchange among participating nodes.

The fresh information propagates in the network when nodes share the information of their current partial views by using methods *POSTSTATUS ( )* and *POSTBACKSTATUS ( )*. In this way, the nodes update its partial views with fresh descriptors. Due to this continual updating of partial views, the nodes that are no longer participating in the network or have crashed will be removed automatically, since crashed nodes are not available for exchanging partial view information.

4) Interest-based Superpeer and Client Relationship

The algorithm, which is used to build the relationship between Interest-based superpeer and clients, based on modified gossip protocol, is shown in the Figure 5.

The logic behind the use of RANDSELECPEER method is that all interest-based superpeers continuously look for those nodes with more capacity (in terms of handling nodes) than the current ones and when they find one, push their entire load to the new one. For this purpose, the function looks for those interest-based superpeers who are underloaded i.e. have equal or more capacity than this node, and have similar interest. The process carries on until the particular node with highest capacity is located or no further node is available to make connection with.

```
RANDSELECPEER ()
    S <= {(R|cR >= cP) && (R belongs-to UNDERLOADED) && (P.interest ==
    R.interest)}
    T <= null
    While (S != {} && T = null)
        R <= <select a random node from S>
        S = S – {R}
        lR <= <request load from R>
        if (lR < cR && (cP < cR || lR > lp))  //Found
        T <= R
    return T
POSTSTATUS (T)
    C <= <choose min (cT – lT , lP) local clients>
    send C to T
POSTBACKSTATUS (T)
    send {} to T
UPDATE (C , T)
    CLIENTS <= CLIENTS ∪ C
    if (lT ==0 && lP < cP)
    CLIENTS <= CLIENTS ∪ {T}
    <T becomes a client>
    else if (∃ R belongs-to CLIENTS : cR > cq)
    <transfer clients of T to R>
    CLIENTS <= CLIENTS ∪ {T} - {R}
    <T becomes a client, R becomes a server>
```

Fig. 5 The interest-based superpeer selection algorithm. Notations: P is local peer; T and R are identifiers of remote one; l_R, l_P and l_T denote the load in terms of number of clients peers R, P and T respectively; C_R, C_P and C_T denotes the capacity of peers R, P and T respectively.

After the successful selection of interest-based superpeer, the method POSTSTATUS is used to get the information from the selected superpeer that the amount of loads it ready to accept. The POSTBACKSTATUS method returns nothing i.e. empty set, because client transfer is one way process.

The method UPDATE is used for adding clients in the CLIENTS set of that node, whose capacity is higher and selected as a superpeer (local node). If the CLIENTS set of remote superpeer node is an empty set, it joins the local node as a client. If the capacity of the client present in CLIENTS set of local node is higher than the remote superpeer, the position of that client and remote superpeer swaps .

IV. RESULTS AND DISCUSSION

This section shows the behaviour of our interest-based superpeer protocol. The simulator used for the simulation is PeerSim [7].

In these experiments, we check the behaviour of our protocol by comparing it with the work done by [5]. Two different approaches used for the capacity parameter: uniform and power law. In uniform distribution, the range of the capacity is chosen randomly i.e. from 1 to maximum capacity

[1, C_{max}], the intention to use this distribution is to provide the fact that results produced are not limited to a particular distribution. Whereas, in power law distribution, a node with capacity has probability of *$P[c_n=x]$*, where *n* is the node and *x* is the capacity with bounded range *[1, C_{max}]*, which is equivalent to the notation $x^{-\alpha}$, where α is the distribution component. $\alpha=2$ will be used in the experiments which provides reasonable approximation. The reason behind the power law distribution approach is to use it on the existing P2P networks where nodes are not capable of handling many clients because they do have not enough capacity.

The terms uniform-interest and power law-interest represents our work. The parameters used in these experiments are shown in the table below.

TABLE I
PARAMETERS USED IN EXPERIMENTS

Description	Value
Maximum no. of rounds an experiment can run	200
Size of partial view used in NEWSCAST	30
Max. capacity of a node	500
Total no. of nodes in a network	5×10^4, 1×10^4
Random interest	10

Unless stated otherwise, values used in all experiments are the same as given in the table.

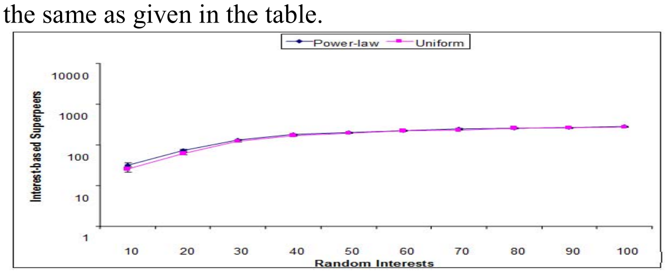

Fig. 6 Total no. of interest-based superpeers in the network as no. of random interest increases.

Figure 6 shows the behaviour of interest-based superpeers in the network as number of random interest increases. In this simulation we used whole numbers to represent the interest of nodes. Nodes entering the network are assigned random interests.

Figure 6, clearly shows that as the number of nodes increase in the network, the number of superpeers also increases in the network. This proves the assumption that more interest-based superpeers will be formed as number of nodes increase in the network.

A question arises, why the increase in interest-based superpeer is not linear? This is due to the fact that the formation of the superpeers is based on random interests; for instance, consider the case where the total number of random interests in the network is 10 this means that it is not necessary that always 10 different interest will enter the network, it is possible that same interest many enters multiple times into the network, thus multiple superpeers with similar interest may exist in the network. So curve cannot be linear, due to random behaviour.

Figure 7 shows the behaviour of the protocol with respect to the variation in the nodes with different interests entering in the network. It is quite evident from the graph that when the numbers of nodes with different interests increase in the network, the number of superpeer will also increase, which will ultimately decrease the capacity utilization of superpeers. This also proves that there is a direct relation between numbers of nodes interest in the network and network size.

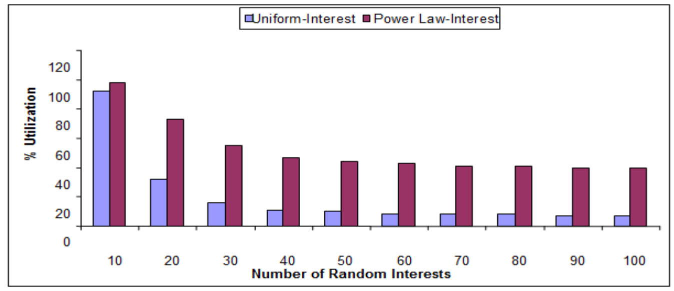

Fig. 7 Capacity utilization of interest-based superpeers at given number of random interests.

Another aspect of new protocol is also evident from the graph, the smaller the number of interests in the network the quicker the formation of interest-based superpeers and the capacity utilisation of superpeers reach up to 90%.

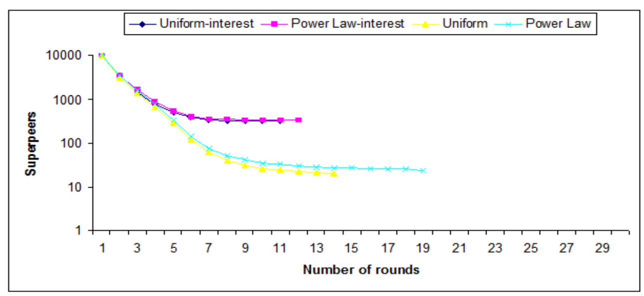

Fig. 8 Total no. of superpeers in the network after the execution of the specified no. of rounds.

The curves in Figure 8 show the number of superpeers in the network after execution of different number of rounds. The results show that the interest-based superpeer algorithm proved to be extremely fast, the resulting topologies have achieved the level after less than 13 rounds after that no more interest-based superpeers are formed. The capacity-based superpeers achieved the level after less than 21 rounds.

Since, the new protocol uses the notion of interest and the maximum number of random interests used in this experiment is 500, the curve stays at around 500 superpeers. This validates that the formation of superpeers in new algorithm is based on interests. Each superpeer is responsible for the set of clients whose interests are same thus forming a group or cluster of common interest.

Figure 9 shows the behaviour of the protocol with respect to the change in maximum capacity (cache size of node), only number of rounds required to achieve the 90% threshold are shown. The results obtained are as expected. As capacity increases the ratio of interest-based superpeers decreases, but with the expense of increase in number of rounds, shown in Figure 9.

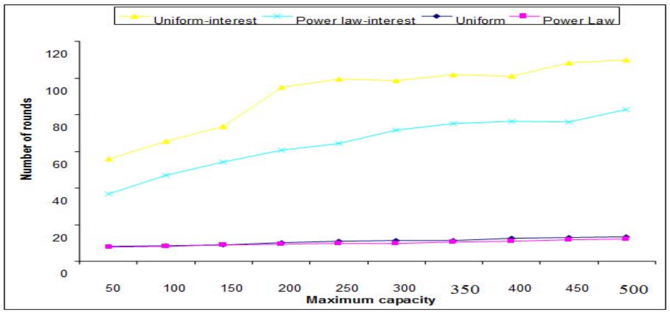

Fig. 9 No. of rounds needed to reach the 90% utilization threshold, at given maximum capacity.

For the capacity-based protocol, the results show a gradual increase in number of rounds required with respect to capacity. The number of rounds required peaks at 14 rounds where capacity is equal to 500. While the new interest-based protocol exhibits similar behaviours, the number of rounds required is far higher than the capacity-based one. The reason for the higher number of rounds is due to the interest factor. For the interest factor, a longer time is required for the formation of interest-based superpeers and to achieve the 90% utilisation, thus accounting for the increase in number of rounds.

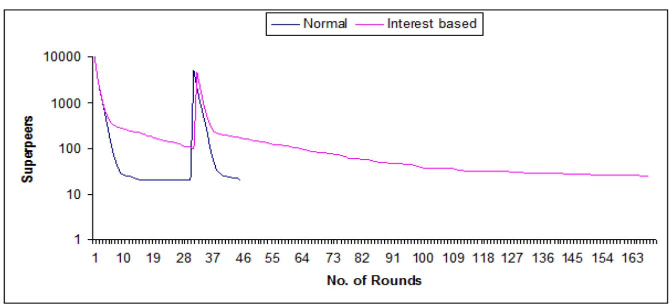

Fig. 10 Sudden failure states, 50% of the superpeers are removed from the network at round 30.

The Figure 10 shows the robustness of new protocol in sudden failure situations, where 50% of the superpeers are removed from the network at round 30.

The results are quite similar to the existing results, after the 50% crash of the interest-based superpeers; the new protocol starts repairing the network by selecting the new interest-based superpeers among the remaining nodes.

V. Conclusion and Further work

Different test have been performed to analyse the behaviour of the new protocol and compared it with the existing protocol [5]. The tests showed the satisfactory results, the new proposed interest-based protocol is effectively working in the dynamic nature of P2P networks and can handle the abrupt leaving and joining of nodes in the network. The interest-based superpeer protocol also seems to be very robust in the sudden failure situation, where a number of interest-based superpeers are removed from the network; the protocol effectively starts repairing the network with the remaining nodes available.

However, the convergence time in the existing protocol seems to be faster than the new proposed interest-based superpeer protocol, the reason behind this behaviour is due to the fact that the new protocol is based on the interest paradigm. In capacity-based superpeer protocol the selection of superpeer among set of neighbouring nodes is only based on capacity, so convergence time is low. Whereas, in interest-based superpeer protocol, the neighbourhood relationship of nodes is based on interest, so selection of superpeer is based on capacity as well as interest. This process naturally requires more time for the selection of interest-based superpeer, thus convergence time is high.

Future work will focus in the several areas which have not been taken into account in this paper: The first is how to minimise the number of multiple interest-based superpeers with similar interests. Although superpeers with similar interest create redundancy in the network but still requires limiting them at an extent, thus required management in this area. Second, is the issue of how to deal with the situation where a single node has multiple interests. Our modelling of the interest-factor currently makes very broad assumptions as to the classification of a peer within an interest group. We plan to use learning techniques to help create superpeers based on a multi-interest taxonomy. Finally we will investigate the impact of further self organisation on interest-based superpeers.

References

[1] The Kazaa website. [Online]. Available: https://www.kazaa.com

[2] Kojima K., "Grouped peer-to-peer networks and self-organization algorithm", IEEE international conference on Systems, Mans and Cybernetics, vol. 3, pp. 2970-2976, Oct. 2003.

[3] Gatani, L.; Lo Re, G.; Gaglio, S., "An Adaptive Routing Protocol for AdHoc Peer-to-Peer Networks", Sixth IEEE International Symposium on a World of Wireless Mobile and Multimedia Networks, pp. 44 – 50, June 2005.

[4] Khambatti, M., Ryu K. D., & Dasgupta P. (2004). "Structuring Peer-to-Peer Networks Using Interest-Based Communities". The SpringerLink website. [online]. Available: http://www.springerlink.com/content/ja62jfgqu6kxa1xr/fulltext.pdf

[5] Montersor A., "A Robust Protocol for building Superpeer Overlay Topologies", Univ. of Bologna, Bologna (Italy), Tech. Rep. UBLCS-2004-8, 2004.

[6] Eugster P.T., Guerraoui R., Kermarrec A.-M., & Massouli´e L., "From Epidemics to Distributed Computing", IEEE Computer, vol. 37, pp. 60-67, 2004.

[7] The PEERSIM website.[Online]. Available: http://peersim.sourceforge.net/