International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)
International Conference on Information Systems and Computing (ICISC-2013), INDIA.

PEER TO PEER NODE GROUPINGS BASED ON CONTENT USING GAME THEORETIC STRATEGY IN INTER-NETWORKS

P.Kanaga[1], N.Mohan praphu[2]
[1,2] PG Scholar, Computer Science and Engineering, Shivani Engineering College,Trichy.
Kpkanagapalani3@gmail.com, Sbc_ias@gmail.com

Abstract

Peer-to-peer systems (P2P) have grown-up drastically in recent years. Peer-to-peer suggests the probable low cost sharing of information, autonomy, and privacy. In Peer to Peer network game theory will be the main tool for modeling and analyzing node behavior. In our concept the participant or nodes are linked based on their content or interest, therefore form clusters. The configurations of clustered overlays can be framed by using game theoretic strategy; the nodes determine their cluster membership with the ambition of improving the recall of their queries. Overlays can be developed together tentatively also experimentally in provisions of stability, optimality, load balance and the required operating cost. Here we also proposed an uncoordinated cluster formulation protocol. The notion of protocol is to each node plays in the vicinity and autonomously from all other nodes in the system.

I. INTRODUCTION

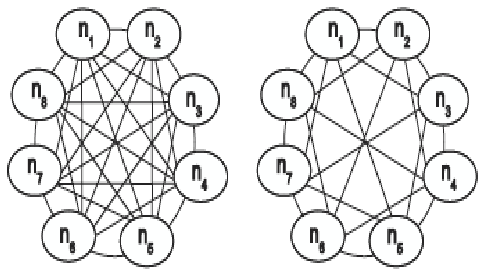
Content sharing applications has been suddenly increased such as individuals concerning social networking and peer-to-peer file sharing. Dimensions from the large-scale systems show that the interactions among their participants or nodes signify the subsistence of groups or else cluster of nodes having related content or interest. The algorithm which has been used in Bit Torrent for Peer selection was exposed for the arrangement of clusters of peers having similar content or interest. The chief motive for the development of such clusters is with the intention of clustered overlays facilitates their participant to discover and exchange data relevant to their queries with a lesser amount of effort. For ex, traces of popular p2p systems have indicated that nodes demonstrate the assets of interest-based locality, that is, if a node holds content satisfying some query of another node, then it most likely also maintains supplementary content of interest to this other node [3], [8]. We tackle the forceful creation and alteration of clustered overlays by using game theoretic approach Each node plays by selecting which clusters to join so the utility function is minimized that depends on cluster membership cost and query evaluating cost of the clusters.

II. RELATED WORK

The application of game theory to network creation has been assortment of work [7].

Nodes act as uncoordinated selfish agents in the network formation game [1] whose goal is to choose other nodes to link with. For the creation of a link the nodes pay, but the node attains gain by reducing their shortest distance to all other nodes in the system. The various versions of game have been proposed for the clustered overlay networks. The objective of our approach is to using games to mock-up clustered overlay creation toward improving recall. In our replica we consider clustered overlays instead of unstructured ones and the gain depends on the recall not based on the number of nodes. In [12], links are directed and nodes are allowed to express preferences regarding the choice of their neighbors. The recent research in p2p overlays have focused on organizing nodes in cluster [9].

III. GAME FORMATION FOR CLUSTER



a) **b)**

Fig 1.Examples of cluster topologies

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)
International Conference on Information Systems and Computing (ICISC-2013), INDIA.

In dynamic large-scale content sharing distributed system there is no possibility for a node to know and directly communicate with all other nodes in the system. As a replacement every node should create logical links with only few other nodes in the cluster. The logical links generate a logical overlay network on top of the physical one. In this manuscript we think about clustered overlays where nodes through similar content or interest structure groups called clusters. To achieve an efficient intra cluster communication nodes within each cluster are highly connected with each other (Fig.1).

Let V be the current set of nodes. The problem of cluster overlay formation can be framed as a strategic game; the strategy of node is distinct by which cluster it joins. In particular, each node n_i €V chooses which clusters to join from the set C of clusters in the system, thus defining its strategy. Let V (s_i) be the set of all nodes belonging to clusters in s_i. Our foundation is that intra cluster query estimation is very efficient. Thus, the benefit for a node n_i from choosing strategy s_i is the recall attained by evaluating the queries in Q (n_i) against the nodes in V (s_i).

IV. Cost Function

In cost function we consider the individual cost and the social cost.

Individual cost:

The individual cost for node n_i for choosing strategy s_i in a cluster configuration is given by

$$icost(ni,S) = \alpha \sum_{c_k \in s_i} \left(\frac{\boldsymbol{\theta}(|\boldsymbol{ck}|)}{|\boldsymbol{V}|}\right) + \sum_{q\ in\ Q(n_i)} \frac{\boldsymbol{num}(\boldsymbol{q},\boldsymbol{Q}(\boldsymbol{ni}))}{\boldsymbol{num}(\boldsymbol{Q}(\boldsymbol{ni}))} \sum_{n_j \in V(s_i)} r(q,nj)$$

The first term denotes the cluster membership cost, whereas the second term denotes the recall cost for obtaining outcome commencing nodes outside the selected clusters, that is, the average result loss from not participating in all clusters. The recall cost for each query is weighted by its frequency in the local workload of n_i. The factor 1/|V| is used for normalizing the cluster membership cost.

Social cost:

The social cost is the sum of all the individual cost and also the overall quality of cluster configuration is achieved by the social cost and it is given by

$$SCost(S) = \sum icost(n_i, S)$$

CASE STUDIES

Here, we focus on a number of characteristic special cases of content sharing and study for each one of them, the properties of a number of cluster configurations. The formal definition of each case as well as the derivation of our results can be found in [9].

CASE I: NO UNDERLYING CLUSTERING

Here, there is no content or query similarity among the nodes.

CASE II: SYMMETRIC SCENARIO.

The symmetric scenario corresponds to the most favorable scenario for clustering. In g(g≥ 1)content categories the nodes belong to g disjoint groups of the same size (|V|/g) such that the members of each group offer and request content from the same category.

CASE III: ASYMMETRIC SCENARIO.

In asymmetric scenario the shared content belongs to t(t>1) different categories; still each node has content that belongs to one category, but poses queries for content that belongs to a single different category.

V. Protocols For Cluster Configuration

Here we described about how the game is played by the individual nodes in the distributed system**.** To improve their recall the nodes form clusters by reconstruct their overlay network by playing the game. Cluster reformulation is essential when the recall achieved in the current cluster configuration deteriorates.

5.1 Uncoordinated Protocol

Let S_{cur} and C_{cur} be the current cluster configuration and set of (nonempty) clusters, respectively. Each node n_i considers all possible configurations S_j that differ from S_{cur} only at their i^{th} component, i.e., the strategy s_i that n_i follows. Then, n_i chooses the strategy s_{new} for which the corresponding cluster configuration S_{new} is such that

$$S_{new} = \arg\min_{S_j} icost(n_i, S_j).$$

To compute how much a node benefits from changing its strategy, we use gain defined as

$$\boldsymbol{gain}(\boldsymbol{n_i}) = \boldsymbol{icost}(\boldsymbol{n_i}, \boldsymbol{S_{cur}}) - \boldsymbol{icost}(\boldsymbol{n_i}, \boldsymbol{S_{new}})$$

If gain(n_i) >0,n_i selects a new strategy.The node re-examines its strategy selection repeatedly to cope with the system dynamics. In uncoordinated protocol the node n_i autonomously decides to play and reevaluates its gain, after the evaluation of each of its query**.**

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)
International Conference on Information Systems and Computing (ICISC-2013), INDIA.

5.2Coordinated protocol

On applying this protocol, a node in each cluster serves as representative. They are responsible for gathering and transferring information. This is triggered after global event; they will intimate their nodes to re evaluate their strategies based on optimal gain.

Partial Knowledge

The node should evaluate the cost of various strategies to choose best one using FSR.In full set routing approach each node n_i evaluates all of its queries using one of its known clusters.

Overhead control

Parameters can be presented for overhead control that can be used either alone or in combinations. The parameter values can either be the same for all nodes or be set individually by each node.

Tentative assessment:

Here we consider a set of nodes sharing documents belonging to special semantic categories. For capturing locality we use a model [2], the model uses a parameter L as a measure of interest based locality.

Newsgroup articles belonging to 10 different categories can be used as our data set. The articles were preprocessed, stop words were unconcerned, lemmatization was functional, and the consequential words were sorted by frequency of manifestation. The articles are distributed among 10,000 nodes. Inside each cluster nodes are organized in a Chord-like topology [6].

VI. SENSITIVITY ANALYSIS

First we sensitively analyze the uncoordinated protocol with respect to its tuning parameters. The basic uncoordinated protocol (unc) and the uncoordinated protocol with monitoring (mon) both are taken consideration which makes the unrealistic assumption of knowledge of all global events.

Table 2:

Parameter	Range	Default
Topology and Knowledge Degree		
Number of nodes(|V|)	-	10000
Parameter α	1-100	10
Membership cost function(θ)	Log, linear	Log
Knowledge Degree	20%-100%	FK,50%-FSR-SSR
Content-Query Distribution		
Number of categories	-	10
Interest Locality Degree(L)	-	0-1
Tuning Parameters		
Stopping condition(€)	$0.3\text{-}10^{-8}$	$10^{-2},10^{-3}$
Update probability(P_u)	0-1	0.7
Batch size in events(b)	1-100	1
Quota(ch)	1-15	∞
Quota period in events(T_q)	20	-
Granted requests %(K)	-	70%

Input Parameters

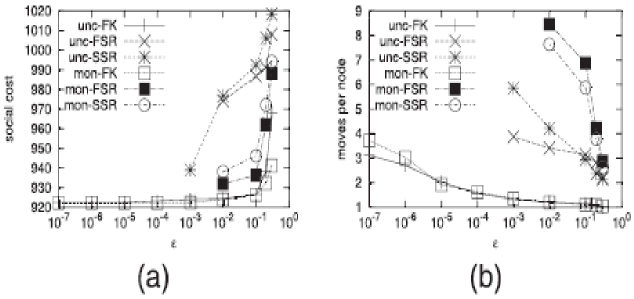
We vary the value of one of the tuning parameters and set the rest to their default values (Table 2a). The initial configuration is an overlay, where each node forms a cluster of its own. The overhead can be measured in terms of the average number of moves per node and the attained social cost. The social cost of the initial configuration is above 10,000 since all nodes have recall cost equal to 1.The social cost is mainly controlled by the stopping condition €.

IJETAE
EXPLORING RESEARCH AND INNOVATIONS

International Journal of Emerging Technology and Advanced Engineering

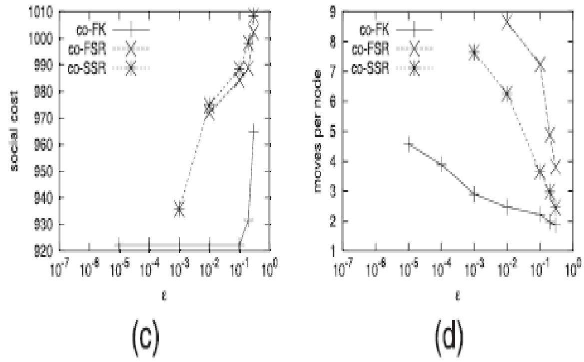
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)

International Conference on Information Systems and Computing (ICISC-2013), INDIA.



In the above fig 2 (a) and (b) are uncoordinated.

The values of € of the social cost is reduced to at most 1,050, that is more or less 1/10th of its initial value (Fig. 2a). The Uncoordinated protocol achieves the best social cost for the smaller value of € surrounded by a few moves and the uncoordinated protocol with monitoring also achieve similar social cost. The SSR or FSR protocols involving large overheads for small values of € that prevent the system from reaching a stable state. The other parameters result can be found in [9].



In the above fig (c) and (d) are coordinated protoocols with varying €.

The coordination protocol does not improve the social cost (Fig. 2c).The protocol also requires high values of € and the protocol approximately achieves the same social cost with the uncoordinated ones with similar operating cost (Fig. 2d).

Cluster configuration

Here we study the (basic) uncoordinated protocol. Mainly we focus on the three scenarios (i.e., symmetric, asymmetric, and random from the case studies) and relate the uncoordinated protocol at a number of initial configurations.

The quality of cluster configurations can be evaluated in terms of their social cost and the operating cost required for achieving number of moves (i.e., cluster changes), the number of clusters formed, the achieved balance, and the effect of the lack of global knowledge.

Let M be the number of groups as defined for configuration (C) for the symmetric and the asymmetric scenarios. We consider five initial system configurations.

i. Each node forms a cluster of its own.
ii. All nodes form a single cluster.
iii. Nodes are randomly distributed to l groups
iv. Nodes are clustered according to their content.
v. Nodes are clustered according to their queries.

Table 2b:
RESULTS REGARDING CLUSTER FORMATION

	Moves per Node			Number of Clusters			δ_S			SCost		
	FK	SSR	FSR	FK	SSR	FSR	FK	SSR	FSR	FK	SSR	FSR
Symmetric Scenario												
i	1.49	2.48	2.92	10	10.42	10.76	0.102	0.171	0.108	10.05	10.21	10.28
ii	1.46	2.07	2.75	10	10.27	10.59	0.143	0.176	0.110	10.24	10.12	10.32
iii(a)	0.91	1.89	2.47	10	10.61	10.82	0.078	0.102	0.112	10.52	11.24	11.43
iii(b)	1.05	2.42	2.92	10	11.08	10.92	0.102	0.023	0.015	11.25	11.78	12.01
iii(c)	0.82	1.81	2.61	10	11.21	10.92	0.115	0.108	0.120	10.48	11.25	11.52
iv	0	0	0	10	10	10	0	0	0	10.09	10.09	10.09
v	0	0	0	10	10	10	0	0	0	10.09	10.09	10.09
Asymmetric Scenario												
i	1.41	2.54	3.58	44.65	44.19	46.55	0.015	0.129	0.167	922.01	995.26	989.35
ii	1.52	2.38	3.27	45.45	44.86	46.76	0.035	0.112	0.109	926.15	1022.82	1027.14
iii(a)	0.88	1.87	3.05	45.25	46.12	46.75	0.078	0.109	0.120	924.21	1014.15	1016.05
iii(b)	0.78	1.79	2.92	44.45	43.27	45.23	0.109	0.092	0.097	924.15	1018.52	1027.41
iii(c)	0.71	2.02	2.85	44.71	44.78	45.27	0.092	0.150	0.087	924.24	1031.45	1041.15
iv	0.54	1.98	2.89	45.17	44.89	44.55	0.090	0.122	0.141	929.25	1056.45	1059.22
v	0.57	1.76	2.07	45.20	45.75	44.96	0.098	0.107	0.156	918.67	1032.65	1042.75
Random Scenario (No Underlying Clustering)												
i	1.74	3.45	4.02	1	1	1	0	0	0	11.33	11.33	11.33
ii	0	0	0	1	1	1	0	0	0	11.33	11.33	11.33

The results are summarized in Table 2b.

The nodes reach Nash equilibrium regardless of the number of clusters in the initial configuration. The number of clusters is dynamically determined by the protocol. When we comparing the FSR and SSR, we monitor that FSR requires more moves for reaching a stable state, but the FSR achieves a better social cost than SSR.The SSR locates relevant cluster faster, but cannot determine accurately the best among them.

In symmetric scenario, the social cost depends only on the membership cost; the recall cost is in most cases zero, while all results of the local queries for each node are located within its cluster (Table 2b, lines 3-10).

Cluster Updates:

Here we evaluate how the protocol reacts to changes. First we start from a "good" cluster configuration and apply query and content updates.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459 (Online), An ISO 9001:2008 Certified Journal, Volume 3, Special Issue 1, January 2013)
International Conference on Information Systems and Computing (ICISC-2013), INDIA.

Next we focus on a general query update scenario (WSc1), in which a data category becomes more popular;y percent of nodes selected randomly from all clusters change their query workload to queries for this grouping.

In a static overlay no measures are taken to deal with updates but our uncoordinated cluster formulation protocol reduces the social cost up to 1/3 of its value. The social cost achieved can also be compared by playing the game repeatedly with applying the protocol from scratch. Compared to our incremental reformulation protocol the reclustering from scratch can reduce the social cost up to 10 percent. In Full set routing and subset routing the difference between reclustering and reformulation are even smaller in terms of the social cost, but the overhead is reduced.

VII. 7.Conclusion

In our approach the clustered overlays can be configured with dynamic creation and adaptation and the framing strategies as a game. Here the nodes act as a Player and the cluster joining strategy is based on nodes so the utility function of the cluster membership cost and the query recall is minimized. To deal with churn, query and content updates, nodes re-evaluate their strategies resulting in dynamic reclustering.

For future work we plan to amend our protocol to handle multiple cluster membership efficiently without increasing its complexity. The unambiguous load-balance component will be constructed along with the component that will effectively hold the inter cluster communication.

References

[1] A. Fabrikant, A. Luthra, E. Maneva, C.H. Papadimitriou, and S. Shenker, "On a Network Creation Game," Proc. 22nd Ann. Symp. Principles of Distributed Computing (PODC), 2003.

[2] P. Garbacki, D.H.J. Epema, and M. van Steen, "Optimizing Peer Relationships in a Super-Peer Network," Proc. 27th Int'l Conf. Distributed Computing Systems (ICDCS), 2007.

[3] S.B. Handurukande, A.-M. Kermarrec, F.L. Fessant, L. Massoulie´, and S. Patarin, "Peer Sharing Behaviour in the eDonkey Network and Implications for the Design of Server-Less File Sharing Systems," Proc. First ACM SIGOPS/EuroSys European Conf. Computer Systems (EuroSys), 2006.

[4] G. Koloniari and E. Pitoura, "A Game Theoretic Approach to the Formation of Clustered Overlay Networks," IEEE Trans. Parallel and Distributed Systems, http://doi.ieeecomputersociety.org/ 10.1109/ TPDS.2011.155.

[5] P. Kuznetsov and S. Schmid, "Towards Network Games with Social Preferences," Proc. 17th Int'l Colloquium Structural Information and Comm. Complexity (SIROCCO), 2010.

[6] N. Laoutaris, G. Smaragdakis, A. Bestavros, and J.W. Byers, "Implications of Selfish Neighbor Selection in Overlay Networks," Proc. IEEE INFOCOM, 2007.

[7] A. Legout, N. Liogkas, E. Kohler, and L. Zhang, "Clustering and Sharing Incentives in BitTorrent Systems," Proc. ACM Int'l Conf. Measurement and Modeling of Computer Systems (SIGMETRICS),2007.

[8] K. Sripanidkulchai, B. Maggs, and H. Zhang, "Efficient Content Location Using Interest-Based Locality in Peer-to-Peer Systems," Proc. IEEE INFOCOM, 2003.

[9] I. Stoica, R. Morris, D. Karger, M.F. Kaashoek, and H. Balakrishnan, "Chord: A Scalable Peer-to-Peer Lookup Service for Internet Applications," Proc. Conf. Applications, Technologies, Architectures, and Protocols for Computer Comm. (SIGCOMM), 2001.

[10] E. Tardos and T. Wexler, "Network Formation Games," Algorithmic Game Theory, Cambridge Univ. Press, 2007.

[11] P. Triantafillou, C. Xiruhaki, M. Koubarakis, and N. Ntarmos, "Towards High Performance Peer-to-Peer Content and Resource Sharing Systems," Proc. First ACM SIGMOD/VLDB Conf. Innovative Data Systems Research (CIDR), 2003.

[12] G. Koloniari and E. Pitoura, "A Recall-Based Cluster Formation Game in Peer-to-Peer Systems," Proc. Very Large Database Endowment (VLDB), 2009.

[13] P. Kuznetsov and S. Schmid, "Towards Network Games with Social Preferences," Proc. 17th Int'l Colloquium Structural Information and Comm. Complexity (SIROCCO), 2010.