Scientific Journal of Impact Factor (SJIF): 4.72

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 4, Issue 4, April -2017

A Proximity-Aware Interest-Clustered P2P File Sharing System

[1]Ashwini Patil, [2]Prof. Panakj Agarkar

[1,2]Computer Department,Dr D Y Patil College Of Engineering And Technology,Lohgaon,Pune

Abstract - *Efficient file question is vital to the general performance of peer-to-peer (P2P) file sharing systems. Clustering peers by their common interests will considerably enhance the efficiency of file question. Clustering peers by their physical proximity also can improve file question performance. However, few current works area unit able to cluster peers supported each peer interest and physical proximity. Though structured P2Ps give higher file question efficiency than unstructured P2Ps, it's tough to understand it as a result of their strictly outlined topologies. During this work, we have a tendency to introduce a Proximity-Aware and Interest-clustered P2P file sharing System (PAIS) supported a structured P2P, that forms physically-close nodes into a cluster and more teams physically-close and common-interest nodes into a sub-cluster supported a graded topology. PAIS uses associate degree intelligent file replication rule to more enhance file question efficiency. It creates replicas of files that area unit of times requested by a gaggle of physically shut nodes in their location. Moreover, PAIS enhances the intra-sub-cluster file rummaging through many approaches. First, it more classifies the interest of a sub-cluster to variety of sub-interests, and clusters common-sub interest nodes into a gaggle for file sharing. Second, PAIS builds associate degree overlay for every cluster that connects lower capability nodes to higher capability nodes for distributed file querying whereas avoiding node overload. Third, to cut back file looking out delay, PAIS uses proactive file info assortment in order that a file requester will recognize if its requested file is in its close nodes. Fourth, to cut back the overhead of the file info assortment, PAIS uses bloom filter primarily based file info assortment and corresponding distributed file looking out. Fifth, to boost the file sharing efficiency, PAIS ranks the bloom filter leads to order. Sixth, considering that a recently visited file tends to be visited once more, the bloom filter primarily based approach is increased by solely checking the freshly more bloom filter info to cut back file looking out delay. Further, the experimental results show the high effectiveness of the intra-sub-cluster file looking out approaches in rising file looking out efficiency.*

Keywords*-MANET, P2P, Resource Sharing, PAIS, Distributed Hash Tables (DHTs).*

I. INTRODUCTION

Over the past few years, the vast quality of the net has created a big stimulant to P2P file sharing systems. There square measure 2 categories of P2P systems: unstructured and structured. Unstructured P2P networks like Gnutella and free net doesn't assign responsibility for information to specific nodes. Nodes be a part of and leave the network in keeping with some loose rules. Currently, unstructured P2P networks' file question methodology is predicated on either flooding wherever the question is propagated to any or the entire node's neighbors, or random-walkers wherever the question is forwarded to willy-nilly chosen neighbors till the file is found. However, flooding and random walkers cannot guarantee information location. Structured P2P networks i.e., Distributed Hash Tables (DHTs), will overcome the drawbacks with their options of upper potency, quantifiability, and settled information location. It strictly controlled topologies and their operation algorithms and information placement square measure exactly outlined supported a DHT organization and consistent hashing operates. The node is answerable for a key will perpetually be found despite the fact that if the system is during a continuous state of amendment. Most of the DHTs need O (log n) hops per operation request with (O log n) neighbors per node, wherever n is that the range of nodes within the system.

II. LITERATURE SURVEY

2.1 Paper Name: An efficient and scalable framework for content-based publish/subscribe systems
Authors: Yingwu Zhu · Haiying Shen
Description: Challenges for content-based publish/subscribe systems embody economical subscription management and event matching, load equalization, and economical and climbable event delivery. This paper presents Associate in Nursing economical and climbable framework for content-based publish/subscribe systems. We have a tendency to propose victimization K-D trees to dynamically partition and organize subscriptions, thereby conserving subscription neck of the woods, minimizing event matching load and guaranteeing load balance across nodes. We have a tendency to propose Associate in nursing economical event delivery mechanism that smartly exploits embedded trees in distributed hash tables to broadcast events. We have a tendency to show that the latency of event publication and delivery is low.

The event delivery mechanism will deliver events to an oversized range of subscribers at low latency and overhead, overwhelming modest information measure.

2.2 Paper Name: Hash-based proximity clustering for efficient load balancing in heterogeneous DHT networks
Authors: Haiying Shena, Cheng-Zhong Xub
Description: Distributed hash table (DHT) networks supported consistent hashing functions have Associate in nursing inherent load uneven distribution downside. The target of DHT load equalization is to balance the employment of the network nodes in proportion to their capability thus on eliminate traffic bottleneck. It's difficult owing to the dynamism, proximity and heterogeneousness natures of DHT networks and time-varying load characteristics. During this paper, we have a tendency to gift a hash-based proximity cluster approach for load equalization in heterogeneous DHTs. within the approach, DHT nodes are classified as regular nodes and super nodes per their computing and networking capacities. Regular nodes are sorted and related to super nodes via consistent hashing of their physical proximity data on the net. The super nodes kind a self-organized and churn-resilient auxiliary network for load equalization. The hierarchical data structure facilitates the planning and implementation of a locality-aware irregular (LAR) load equalization formula. The formula introduces an element of randomness within the load equalization processes in an exceedingly vary of neighborhood thus on cope with each the proximity and dynamism. Simulation results show the prevalence of the cluster approach with LAR, compared with variety of different DHT load equalization algorithms. The approach performs no worse than existing proximity-aware algorithms and exhibits sturdy resilience to the impact of churn. It additionally greatly reduces the overhead of resilient irregular load equalization attributable to the employment of proximity data.

2.3 Paper Name: Cycloid: A constant-degree and lookup-efficient P2P overlay network
Authors: Haiying Shena, Cheng-Zhong Xua, Guihai Chenb
Description: There are units several structured P2P systems that use DHT technologies to map knowledge things onto the nodes in numerous ways in which for ascendable routing and site. Most of the systems need O (log n) hops per search request with O (log n) neighbors per node, wherever n is that the network size. During this paper, we have a tendency to gift a constant-degree P2P design; specifically cycloid that emulates a cube-connected cycles (CCC) graph within the routing of search requests. It achieves a time quality of O (d) per search request by exploitation O (1) neighbors per node, wherever $n = d \times 2nd$. We have a tendency to compare Cycloid with alternative 2 constant-degree systems, Viceroy and Koorde in numerous subject aspects via simulation. Simulation results show that Cycloid has a lot of benefits for giant scale and dynamic systems that have frequent node arrivals and departures. Specifically, Cycloid delivers the next location potency within the average case and exhibits a lot of balanced distribution of keys and question masses between the nodes.

2.4 Paper Name: Pastry: Scalable, decentralized object location and routing for large-scale peer-to-peer systems
Authors: Antony Rowstron1 and Peter Druschel
Description: This paper presents the look and analysis of Pastry, a scalable, distributed object location and routing substrate for wide-area peer-to-peer applications. Pastry performs application-level routing and object location in a very doubtless very giant overlay network of nodes connected via the web. It are often wont to support a range of peer-to-peer applications, as well as world knowledge storage, knowledge sharing, cluster communication and naming. Every node within the Pastry network includes a distinctive symbol (nodeId). Once conferred with a message and a key, a Pastry node expeditiously routes the message to the node with a nodeId that's numerically nearest to the key, among all presently live Pastry nodes. Every Pastry node keeps track of its immediate neighbors within the nodeId house, and notifies applications of recent node arrivals, node failures and recoveries. Pastry takes into consideration network locality; it seeks to attenuate the space messages travel, in keeping with a scalar proximity metric just like the range of informatics routing hops. Pastry is totally decentralized, scalable, and self-organizing; it mechanically adapts to the arrival, departure and failure of nodes. Experimental results obtained with a model implementation on associate emulated network of up to one 100,000 nodes ensure Pastry's measurability and potency, its ability to self-organize and adapt to node failures, and its smart network neighborhood properties.

2.5 Paper Name: Chord: A Scalable Peer-to-Peer Lookup Protocol for Internet Applications
Authors: Ion Stoica, Robert Morris, David Liben-Nowell, David R. Karger, M. Frans Kaashoek, Frank Dabek, and Hari Balakrishnan
Description: An elementary drawback that confronts peer-to-peer applications is that the economical location of the node that stores a desired knowledge item. This paper presents Chord, a distributed operation protocol that addresses this drawback. Chord provides support for only one operation: given a key, it maps the key onto a node. Knowledge location will be simply enforced on high of Chord by associating a key with every knowledge item, and storing the key/data try at the node to that the key maps. Chord adapts expeditiously as nodes be part of and leave the system, and might answer queries even though the system is incessantly dynamical. Results from theoretical analysis and simulations show that Chord is scalable: Communication price and therefore the state maintained by every node scale logarithmically with the amount of Chord nodes.

6. Large-Scale Experiment of Co-allocation Strategies for Peer-to-Peer SuperComputing in P2P-MPI
Authors: St´ephane Genaud, Choopan Rattanapoka
Description: High Performance computing usually involves some parallel applications to be deployed on the multiples resources used for the computation. The matter of programing the applying across distributed resources is termed as co-allocation. During a grid context, co-allocation is tough since the grid middleware should face a dynamic setting. Middleware design on a Peer-to-Peer (P2P) basis are planned to tackle most limitations of centralized systems. A number of the problems self-addressed by P2P systems area unit fault tolerance, simple maintenance, and quantifiability in resource discovery. However, the dearth of worldwide information makes programing tough in P2P systems. During this paper, we tend to gift the new developments regarding neck of the woods awareness likewise as co-allocation ways on the market within the latest unharness of P2P-MPI. i) The unfold strategy tries to map processes on hosts thus on maximize the full quantity of accessible memory whereas maintaining neck of the woods of processes as a secondary objective. ii) The concentrate strategy tries to maximize neck of the woods between processes by mistreatment as several cores as hosts provide. The co-allocation theme has been devised to be easy for the user and meets the most high performance computing demand that is neck of the woods. Intensive experiments are conducted on Grid5000 with up to 600 processes on 6 sites throughout France. Results show that we tend to achieved the targeted goals in these real conditions.

***7.* Piazza: Data Management Infrastructure for Semantic Web Applications**
Authors: Alon Y. Halevy,Zachary G. Ives,Peter Mork,Igor Tatarinov
Description: The linguistics net envisions a World Wide net during which knowledge is represented with made linguistics and applications will cause complicated queries. to the current purpose, researchers have outlined new languages for specifying meanings for ideas and developed techniques for reasoning concerning them, exploitation RDF because the knowledge model. To flourish, the linguistics net has to be ready to accommodate the massive amounts of existing knowledge and also the applications operational on them. To realize this, we have a tendency to square measure two-faced with 2 issues. First, most of the world's knowledge is offered not in RDF however in XML; XML and also the applications intense it swear not solely on the domain structure of the info, however conjointly on its document structure. Hence, to produce ability between such sources, we have a tendency to should map between each their domain structures and their document structures. Second, knowledge management practitioners typically favor to exchange knowledge through native point-to-point knowledge translations, instead of mapping to common mediate schemas or ontologies. This paper describes the Piazza system that addresses these challenges. Piazza offers a language for mediating between knowledge sources on the linguistics net that maps each the domain structure and document structure. Piazza conjointly allows interoperation of XML knowledge with RDF knowledge that's among made bird of prey ontologies. Mappings in Piazza square measure provided at an area scale between little sets of nodes, and our question responsive algorithmic program is in a position to chain sets mappings along to get relevant knowledge from across the Piazza network. We have a tendency to conjointly describe Associate in nursing enforced state of affairs in Piazza and also the lessons we have a tendency to learn from it.

8. An Efficient and Trustworthy P2P and Social Network Integrated File Sharing System
Authors: Guoxin Liu, Haiying Shen,Lee Ward
Description: Efficient and trustworthy file querying is very important to the general performance of peer-to-peer (P2P) file sharing systems. Rising ways square measure starting to address this challenge by exploiting on-line social networks (OSNs). However, current OSN-based ways merely cluster common-interest nodes for prime potency or limit the interaction between social friends for prime trait that provides restricted sweetening or contradicts the open and free service goal of P2P systems. Very little analysis has been undertaken to completely and hand in glove leverage OSNs with integrated thought of proximity and interest. During this work, we tend to analyze a BitTorrent file sharing trace that proves the requirement of proximity- and interest-aware clustering. supported the trace study and OSN properties, we tend to propose a Social Network integrated P2P file sharing system with increased potency and trait (SoNet) to completely and hand in glove leverage the common-interest, proximity-close and trust properties of OSN friends. SoNet uses a graded distributed hash table (DHT) to cluster common-interest nodes, then more cluster proximity-close nodes into sub cluster, and connects the nodes in an exceedingly sub cluster with social links. Thus, once queries move trustable social links, they conjointly gain higher chance of being with success resolved by proximity shut nodes, at the same time enhancing potency and trait. The results of trace-driven experiments on the world PlanetLab testbed demonstrate the upper potency and trait of SoNet compared with alternative systems.

9. Quota: Rationing Server Resources in Peer-Assisted Online Hosting Systems
Authors: Fangming Liu, Ye Sun , Bo Li, Baochun Li
Description: Online hosting systems area unit designed to supply versatile and convenient platforms for content hosting and sharing, and have apace become a favorite among users over the web. to ensure adequate levels of service quality whereas protective prohibitory server prices, such systems area unit designed to integrate peer information measure contributions with strategic server resource provisioning in an exceedingly complementary and clear manner. Due to the

massive variety of users in real-world on-line hosting systems, it's not possible to satisfy the resource necessities of all users. This paper seeks to explore the look house of recent protocols to assign scarce server resources together with each space for storing and information measure in peer-assisted on-line hosting systems. The main focus is on the matter of resource allocation with the presence of associate progressively sizable amount of user's exploitation information measure, and also the succeeding larger variety of files exploitation server space for storing. the target is to maximize the utilization of restricted server storage and information measure resources to ensure adequate levels of service quality, with relevance file convenience and downloading performance, whereas taking full advantage of peer help. we tend to establish variety of distinctive challenges concerned in such systems, and propose our style of resource allocation protocols to handle these challenges, supported each mathematical analysis and sensible implementations. Exploitation world information sets that we've collected, we tend to measure our protocol style through intensive experimental studies from completely different views that demonstrate the effectiveness of our style and supply variety of sensible pointers.

10. Topologically-Aware overlay Construction and Server Selection
Authors: Sylivia Ratansamy,Mark Handley,Richard karp,Scott shenker
Description: In this paper, we have a tendency to gift a binning theme wherever nodes partition themselves into hints specified nodes that fall among a given hint are comparatively near each other in terms of network latency. Our binning strategy is easy (requiring stripped-down support from any activity infrastructure); scalable we have a tendency to apply this binning strategy to the 2 applications mentioned overlay network construction and server choice. We have a tendency to take a look at our binning strategy and its application exploitation simulation and net activity traces. Our results indicate that the performance of those applications will considerably improve even the rather coarse-gained information of topology offered by our binning theme.

III. PROPOSED SYSTEM

In this paper represents a proximity-aware and interest-clustered P2P file sharing System (PAIS) on a structured P2P system. It forms physically-close nodes into a cluster and any team's physically-close and common-interest nodes into a sub-cluster. It additionally places files with a similar interests along and create them accessible through the DHT Lookup () routing perform. a lot of significantly, it keeps all benefits of DHTs over unstructured P2Ps. wishing on DHT search policy instead of broadcasting, the PAIS construction consumes abundant less price in mapping nodes to clusters and mapping clusters to interest sub-clusters. PAIS uses Associate in nursing intelligent file replication formula to any enhance file search potency. It creates replicas of files that are often requested by a gaggle of physically shut nodes in their location. Moreover, PAIS enhances the intra sub-cluster file exploring through many approaches initial, it any classifies the interest of a sub-cluster to variety of sub-interests, and clusters common-sub-interest nodes into a gaggle for file sharing. Second, PAIS builds Associate in nursing overlay for every cluster that connects lower capability nodes to higher capability nodes for distributed file querying whereas avoiding node overload. Third, to scale back file looking out delay, PAIS uses proactive file data assortment in order that a file requester will grasp if its requested file is in its near nodes. Fourth, to scale back the overhead of the file data assortment, PAIS uses bloom filter based mostly file data assortment and corresponding distributed file looking out. Fifth, to boost the file sharing potency, PAIS ranks the bloom filter ends up in order. Sixth, considering that a recently visited file tends to be visited once more, the bloom filter based mostly approach is increased by solely checking the fresh further bloom filter data to scale back file looking out delay.

3.1 Advantages of Proposed System:

1. The techniques planned during this paper will profit several current applications like content delivery networks, P2P video-on-demand systems, and information sharing in on-line social networks.
2. We tend to introduce the careful style of PAIS. It's appropriate for a file sharing system wherever files are often classified to variety of interests and every interest are often classified to variety of sub-interests.
3. It teams peers supported each interest and proximity by taking advantage of a hierarchical data structure of a structured P2P.
4. PAIS uses associate intelligent file replication rule that replicates a file often requested by physically shut nodes close to their physical location to reinforce the file search potency.
5. PAIS enhances the file looking out potency among the proximity-close and customary interest nodes through variety of approaches.

IV. SYSTEM ARCHITECTURE

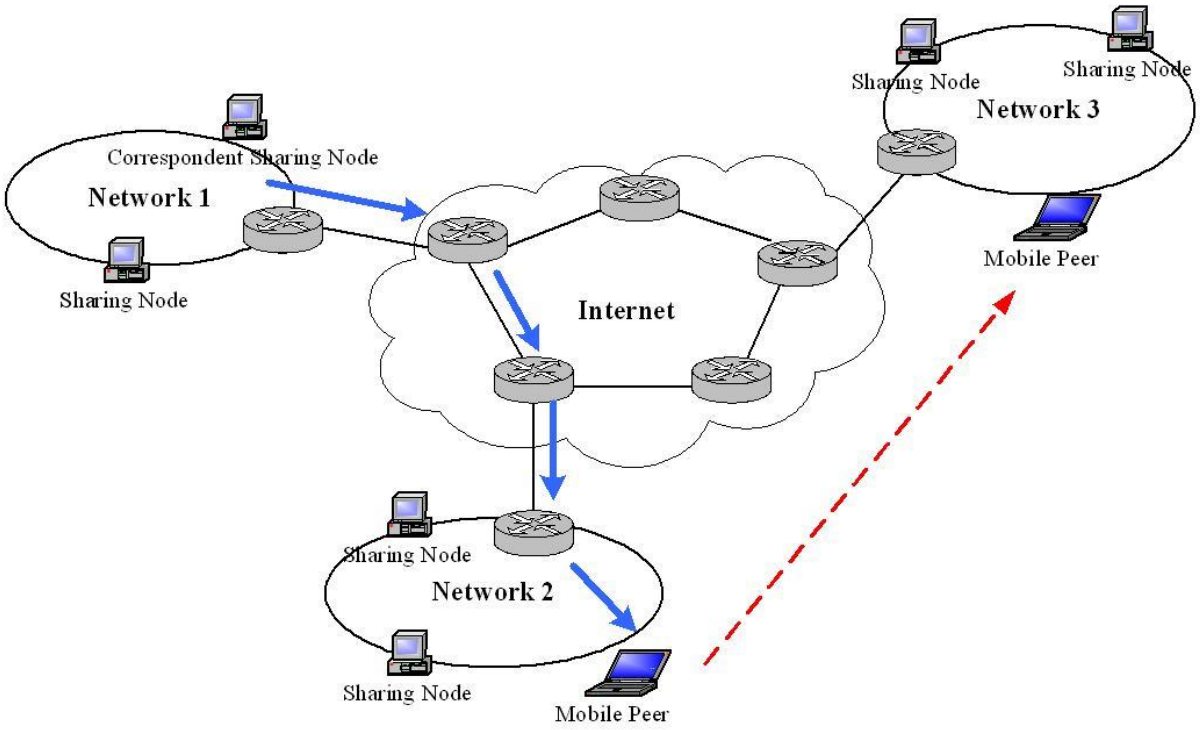

Figure 1. System Architecture of Proposed System

V. MATHEMATICAL MODEL

Input: File query in DTN.

Output: File query result in less time in DTN.

Process:

Let W be the whole system, W = {U, S, C, B, R, r, F}.
Where,

1. U is the set of number users.

   U = {U1,U2 . . . . Un}.

2. S is the system which contains the unstructured data to provide the service to user based on user request.

3. C is set of number of cluster based on user request.

   C = {C1, C2, ... Cn}.

4. B be set of bloom filter which is required to filter the user requests based on user interest.

5. F be the set of files user is requesting.

   F = {f1, f2 . . . . . fn}.

6. R be the user request for file to S.

7. r be the rank assigned to file based user request.

Step 1: User U login to the system and request for particular f1 to the system.

Step 2: The system S will process the user request R from the unstructured data. In this the bloom filter will filter the user request to check whether the same file request has come before or not if not it will rank that file.

Step 3: The system will process the user request R based on ranks assigned to files by using bloom filter.

Step 4: If same file request R is come at system more than 2 times (assigning threshold) then system will create an interest-cluster for that requested file to minimize the searching time as system will search the requested file from unstructured data.

VI. CONCLUSION

In recent years, to reinforce file location potency in P2P systems, interest-clustered super-peer networks and proximity-clustered super-peer networks are planned. Though each ways improve the performance of P2P systems, few works cluster peers supported each peer interest and physical proximity at the same time. Moreover, it's tougher to appreciate it in structured P2P systems because of their strictly outlined topologies, though they need high potency of file location than unstructured P2Ps. during this paper, we have a tendency to introduce a proximity-aware and interest-clustered P2P file sharing system supported a structured P2P. It teams peers supported each interest and proximity by taking advantage of a hierarchical data structure of a structured P2P.

REFERENCES

[1] BitTorrent. (2013) [Online]. Available: http://www.bittorrent. com/

[2] Gnutella home page. (2003) [Online]. Available: http://www.gnutella.com

[3] I. Clarke, O. Sandberg, B. Wiley, and T. W. Hong, "Freenet: A distributed anonymous information storage and retrieval system," in Proc. Int. Workshop Des. Issues Anonymity Unobservability, 2001, pp. 46–66.

[4] I. Stoica, R. Morris, D. Liben-Nowell, D. R. Karger, M. F. Kaashoek, F. Dabek, and H. Balakrishnan, "Chord: A scalable peer-to-peer lookup protocol for internet applications," IEEE/ACM Trans. Netw., vol. 11, no. 1, pp. 17–32, Feb. 2003.

[5] A. Rowstron and P. Druschel, "Pastry: Scalable, decentralized object location and routing for large-scale peer-to-peer systems," in Proc. IFIP/ACM Int. Conf. Distrib. Syst. Platforms Heidelberg, 2001, pp. 329–350.

[6] B. Y. Zhao, L. Huang, J. Stribling, S. C. Rhea, A. D. Joseph, and J. Kubiatowicz, "Tapestry: A resilient global-scale overlay for service deployment," IEEE J. Sel. Areas Commun., vol. 22, no. 1, pp. 41–53, 2004.

[7] H. Shen, C. Xu, and G. Chen, "Cycloid: A scalable constant-degree P2P overlay network," Perform. Eval., vol. 63, pp. 195–216, 2006.

[8] Z. Li, G. Xie, and Z. Li, "Efficient and scalable consistency maintenance for heterogeneous peer-to-peer systems," IEEE Trans. Parallel Distrib. Syst., vol. 19, no. 12, pp. 1695–1708, Dec. 2008.

[9] H. Shen and C.-Z. Xu, "Hash-based proximity clustering for efficient load balancing in heterogeneous DHT networks," J. Parallel Distrib. Comput., vol. 68, pp. 686–702, 2008.

[10] FastTrack. (2003) [Online]. Available: http://www.fasttrack.nu/ index_int.html

[11] S. Ratnasamy, M. Handley, R. Karp, and S. Shenker, "Topologically-aware overlay construction and server selection," in Proc. IEEE INFOCOM, 2002, pp. 1190–1199.

[12] M. Waldvogel and R. Rinaldi, "Efficient topology-aware overlay network," in Proc. ACM Workshop Hot Topics Netw., 2002, pp. 101–106.

[13] Y. Zhu and H. Shen, "An efficient and scalable framework for content-based publish/subscribe systems," Peer-to-Per Netw. Appl., vol. 1, pp. 3–17, 2008.

[14] C. Hang and K. C. Sia, "Peer clustering and firework query model," in Proc. Int. World Wide Web Conf., 2002.

[15] A. Crespo and H. Garcia-Molina, "Routing indices for peer-to-peer systems," in Proc. 22nd Int. Conf. Distrib. Comput. Syst., 2002, pp. 23–32.

[16] W. Nejdl, M. Wolpers, W. Siberski, C. Schmitz, M. Schlosser, I. Brunkhorst, and A. Loeser, "Super-peer-based routing and clustering strategies for RDF-based peer-to-peer networks," J. Web Semantics, vol. 1, no. 2, pp. 137–240, 2004.

[17] P. A. Bernstein, F. Giunchiglia, A. Kementsietsidis, J. Mylopoulos, L. Serafini, and I. Zaihrayeu, "Data management for peer-to-peer computing: A vision," in Proc. 5th Int. Workshop World Wide Web Databases, 2002, pp. 89–94.

