



## Super Peer File Sharing System Using Proximity Aware Interest Clustering.

Pratiksha Gaikwad<sup>1</sup>, Azim Shaikh<sup>2</sup>, Abhijeet Sasulkar<sup>3</sup>, Rajesh Gaud<sup>4</sup>

<sup>1,2,3,4</sup> Department of Computer Engineering Ambi, Pune.

**Abstract** — Efficient file question is very important to the general performance of super peer file sharing systems. Clustering peers by their common interests will considerably enhance the potency of file question. Clustering peers by their physical proximity may improve file question performance. However, few current works area unit ready to cluster peers supported each peer interest and physical proximity. Though structured super peer offer higher file question potency than unstructured super peer, it's tough to appreciate it because of their strictly outlined topologies. During this work, we have a tendency to introduce a Proximity-Aware and Interest-clustered super peer file sharing System (PAIS) supported a structured super peer, that forms physically-close nodes into a cluster and any teams physically-close and common-interest nodes into a sub-cluster supported a hierarchic topology. PAIS uses associate degree intelligent file replication rule to any enhance file question potency. It creates replicas of files that area unit oftentimes requested by a gaggle of physically shut nodes in their location. Moreover, PAIS enhances the intra-sub-cluster file ransacking through many approaches. First, 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 degree 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 info assortment so a file requester will apprehend if its requested file is in its close nodes. Fourth, to scale 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 potency, 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 recently further bloom filter info to scale back file looking out delay. Further, the experimental results show the high effectiveness of the intra-sub-cluster file looking out approaches in up file looking out potency.

**Keywords-** PAIS, Super peer Proximity-Aware and Interest-clustered, Bloom Filter.

### I. INTRODUCTION

Mobile Ad-hoc Networks (MANETs) and super peer overlay networks communicate several very important aspects like organization, decentralization; economical resource sharing and that they even undertake the essential challenge of giving affiliation in a much suburbanized, energetic atmosphere. Though, once observing ad-hoc networks created up by mobile appliances like sensible phones, we won't deem the constant end-to-end path regarding peers as for basic web super peer methods; rather, we have a tendency to possess to manage with low node incidence that produces mobile disconnected networks. This method presents economical file looking for DTNs. It exploits the properties of social networks to boost file looking potency on the premise of file requests. Over the past few years, the huge quality of the net has created a major stimulation to super peer file sharing systems. There are 2 categories of super peer systems: unstructured and structured. Unstructured super peer networks like Gnutella and Freenet doesn't assign responsibility for information to specific nodes. Nodes be a part of and leave the network in step with some loose rules. Currently, unstructured super peer networks' file question methodology is predicated on either flooding wherever the question is propagated to all or any the node's neighbours, or random-walkers wherever the question is forwarded to haphazardly chosen neighbours till the file is found. However, flooding and random walkers cannot guarantee information location. Structured super peer 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 are exactly outlined supported a DHT organisation and consistent hashing operate. The node is answerable for a key will continuously be found despite the fact that if the system is during a continuous state of modification. Most of the DHTs need  $O(\log n)$  hops per operation request with  $(O \log n)$  neighbours per node, wherever  $n$  is that the variety of nodes within the system. Through the study of a true trace, we have a tendency to found that the interests (content) cluster of every node will facilitate guide file looking. We have a tendency to additionally realize that the movement patterns of mobile nodes will a lot of accurately predict the encountering of nodes holding the requested files.

## II. LITERATURE REVIEW

### 1. Paper Name: 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 programming the appliance across distributed resources is termed as co-allocation. During a grid context, co-allocation is troublesome since the grid middleware should face a dynamic setting. Middleware design on a Peer-to-Peer (P2P) basis is projected 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 shortage of worldwide data makes programming troublesome in P2P systems. During this paper, we have a tendency to gift the new developments regarding neighbourhood awareness similarly as co-allocation methods on the market within the latest unleash of P2P-MPI.

- i) The unfold strategy tries to map processes on hosts thus on maximize the whole quantity of accessible memory whereas maintaining neighbourhood of processes as a secondary objective.
- ii) The concentrate strategy tries to maximise neighbourhood between processes by victimisation as several cores as hosts provide. The co-allocation theme has been devised to be easy for the user and meets the main high performance computing demand that is neighbourhood. Intensive experiments are conducted on Grid5000 with up to vi00 processes on 6 sites throughout France. Results show that we have a tendency to achieve the targeted goals in these real conditions.

### 2. Paper Name: 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 information is delineated with wealthy linguistics and applications will create 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 information model. To flourish, the linguistics net has to be able to accommodate the large amounts of existing information and also the applications operative on them. To attain this, we tend to square measure baby-faced with 2 issues. First, most of the world's information is offered not in RDF however in XML; XML and also the applications overwhelming it swear not solely on the domain structure of the information, however conjointly on its document structure. Hence, to produce ability between such sources, we tend to should map between each their domain structures and their document structures. Second, information management practitioners usually favour to exchange information through native point-to-point information translations, instead of mapping to common mediate schemas or ontologies. This paper describes the square system that addresses these challenges. Square offers a language for mediating between information sources on the linguistics net that maps each the domain structure and document structure. Square conjointly allows interoperation of XML information with RDF information that's in the middle of wealthy raptor ontologies. Mappings in square measure provided at an area scale between tiny sets of nodes, and our question responsive algorithmic rule is ready to chain sets mappings along to get relevant information from across the square network. We tend to conjointly describe associate degree enforced situation in square and also the lessons we tend to learned from it.

### 3. Paper Name: 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 vital to the general performance of peer-to-peer (P2P) file sharing systems. Rising ways area unit getting down addressing 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 trustiness that provides restricted sweetening or contradicts the open and free service goal of P2P systems. Very little analysis has been undertaken to totally and hand and glove leverage OSNs with integrated thought of proximity and interest. During this work, we have a tendency to analyse a BitTorrent file sharing trace that proves the requirement of proximity- and interest-aware clustering. supported the trace study and OSN properties, we have a tendency to propose a Social Network integrated P2P file sharing system with increased potency and trustiness (SoNet) to totally and hand and glove leverage the common-interest, proximity-close and trust properties of OSN friends. SoNet uses a class-conscious distributed hash table (DHT) to cluster common-interest nodes, then additional cluster proximity-close nodes into sub-cluster, and connects the nodes during a sub cluster with social links. Thus, once queries follow trustable social links, they conjointly gain higher chance of being with success resolved by proximity shut nodes, at the same time enhancing potency and trustiness. The results of trace-driven experiments on the \$64000 world PlanetLab testbed demonstrate the upper potency and trustiness of SoNet compared with alternative systems.

#### **4. Paper Name: Quota: Rationing Server Resources in Peer-Assisted Online Hosting Systems**

**Authors:** Fangming Liu, Ye Sun, Bo Li, Baochun Li

**Description:** Online hosting systems are designed to produce versatile and convenient platforms for content hosting and sharing, and have become a favourite among users over the web. To ensure adequate levels of service quality whereas preserving preventative server prices, such systems are designed to integrate peer information measure contributions with strategic server resource provisioning during a complementary and clear manner. Attributable to the big range of users in real-world on-line hosting systems, it's not possible to satisfy the resource needs of all users. This paper seeks to explore the planning house of recent protocols to assign scarce server resources as well as each cupboard space and information measure in peer-assisted on-line hosting systems. The main focus is on the matter of resource allocation with the presence of more and more sizable amount of user's exploitation information measure, and therefore the succeeding larger range of files exploitation server cupboard space. The target is to maximise the employment of restricted server storage and information measure resources to ensure adequate levels of service quality, with reference to file convenience and downloading performance, whereas taking full advantage of peer help. We have a tendency to determine 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 knowledge sets that we've collected, we have a tendency to judge our protocol style through intensive experimental studies from totally different views that demonstrate the effectiveness of our style and supply variety of sensible pointers.

#### **5. Paper Name: Topologically-Aware overlay Construction And Server Selection**

**Authors:** Sylvia Ratansamy, Mark Handley, Richard karp, Scott shenker

**Description:** In this paper, we tend to gift a binning theme wherever nodes partition themselves into hints such nodes that fall among a given hint square measure comparatively near each other in terms of network latency. Our binning strategy is easy (requiring nominal support from any activity infrastructure), scalable we tend to apply this binning strategy to the 2 applications mentioned overlay network construction and server choice. We tend to take a look at our binning strategy and its application mistreatment simulation and net activity traces. Our results indicate that the performance of those applications will considerably improve even the rather coarse-gained data of topology offered by our binning theme.

### **III. EXISTING SYSTEM**

A key criterion to evaluate a P2P file sharing system is its file location potency. To enhance this potency, various strategies are planned. One technique uses a brilliant peer topology that consists of super-nodes with quick connections and regular nodes with slower connections. A super-node connects with alternative super-nodes and a few regular nodes, and an everyday node connects with a super-node. During this super-peer topology, the nodes at the middle of the network area unit quicker and thus manufacture a lot of reliable and stable backbone. This enables a lot of messages to be routed than a slower backbone and, therefore, permits larger quantifiability. Super-peer networks occupy the middle-ground between centralized and completely parallel P2P networks, and have the potential to mix the advantages of each centralized and distributed searches.

Another category of strategies to enhance file location potency is thru a proximity-aware structure.

The third category of strategies to enhance file location potency is to cluster nodes with similar interests that scale back the file location latency.

#### **3.1 Disadvantages of Existing System:**

- 1) Although varied proximity-based and interest-based super-peer topologies are projected with totally different options, few strategies are unit able to cluster peers in line with each proximity and interest.
- 2) In addition, most of those strategies are unit on unstructured P2P systems that don't have any strict policy for topology construction.
- 3) They cannot be directly applied to general DHTs in spite of their higher file location potency.

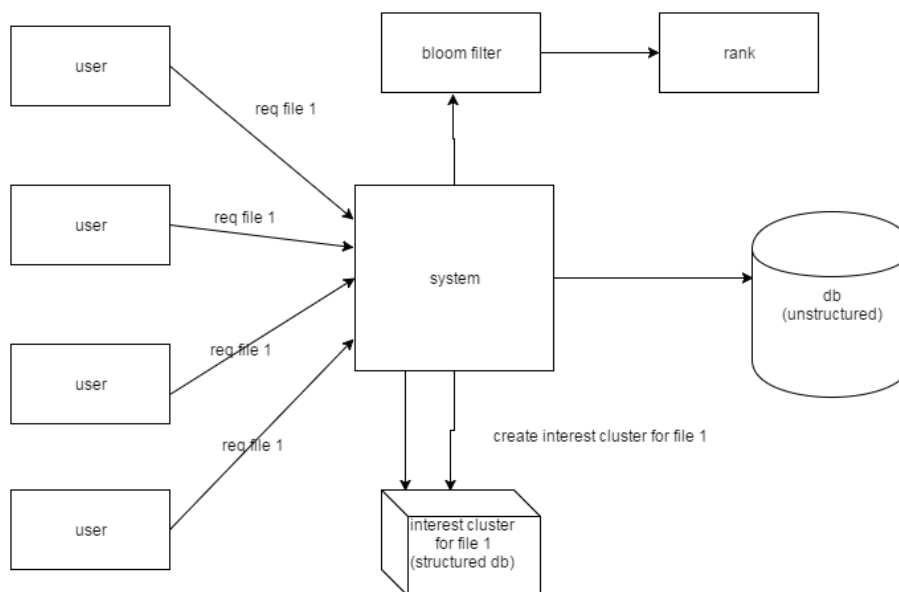
#### IV. PROPOSED SYSTEM

This paper represents a proximity-aware and interest-clustered super peer file sharing System (PAIS) on a structured super peer system. It forms physically-close nodes into a cluster and any team's physically-close and common-interest nodes into a sub-cluster. It places files with constant interests along and build them accessible through the DHT Lookup () routing perform. A lot of significantly, it keeps all advantages of DHTs over unstructured super peer. Hoping on DHT search policy instead of broadcasting; the PAIS construction consumes abundant fewer prices in mapping nodes to clusters and mapping clusters to interest sub-clusters. PAIS uses an intelligent file replication rule to increase the file search ability. It creates duplicates of files that area unit requested by a gaggle of physically shut nodes in their location. Moreover, PAIS enhances the intra sub-cluster file ransacking through many approaches 1st, it divides 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 an overlay for every cluster that connects lower capability nodes to higher capability nodes for distributed file querying for avoiding node overhead. Third, to cut back file looking out delay, PAIS uses proactive file info assortment in order that a file requester will apprehend if its requested file is in its close nodes. Fourth, to cut back the overhead of the file info assortment, PAIS uses bloom filter based mostly file info assortment and corresponding distributed file looking out. Fifth, to enhance 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 freshly additional bloom filter info to cut back file looking out delay.

##### 4.1 Advantages of Proposed System:

- 1) The techniques planned during this paper will profit several current applications like content delivery networks, super peer video-on-demand systems, and information sharing in on-line social networks.
- 2) We have a tendency to introduce the elaborate style of PAIS. it's appropriate for a file sharing system wherever files is classified to variety of interests and every interest is 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 super peer.
- 4) PAIS uses an intelligent file replication algorithmic program that replicates a file often requested by physically shut nodes close to their physical location to boost the file search potency.
- 5) PAIS enhances the file looking out potency among the proximity-close and customary interest nodes through variety of approaches.

#### V. SYSTEM ARCHITECTURE



#### VI. MATHEMATICAL MODEL

**Input:** File query in DTN.

**Output:** File query result in less time in DTN.

**Process:** Let W is the Whole System Consists:

$W = \{U, S, C, B, R, r, F\}$ .

Where,

1.  $U$  is the set of number users.

$U = \{U_1, U_2, \dots, U_n\}$ .

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.

$C = \{C_1, C_2, \dots, C_n\}$ .

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.

$F = \{f_1, f_2, \dots, f_n\}$ .

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  $f_1$  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.

**Output:** Minimize the file searching efficiency

## VII. RESULT ANALYSIS

ScreenShot 1



Screenshot 2

SUPER PEER FILE SHARING SYSTEM USING PROXIMUS AWARE INTEREST CLUSTERING

HOME ADMIN USER LOGIN USER REGISTRATION CONTACT

### User Login

User Name :

Password :

Login

Screenshot 3



Search Reset

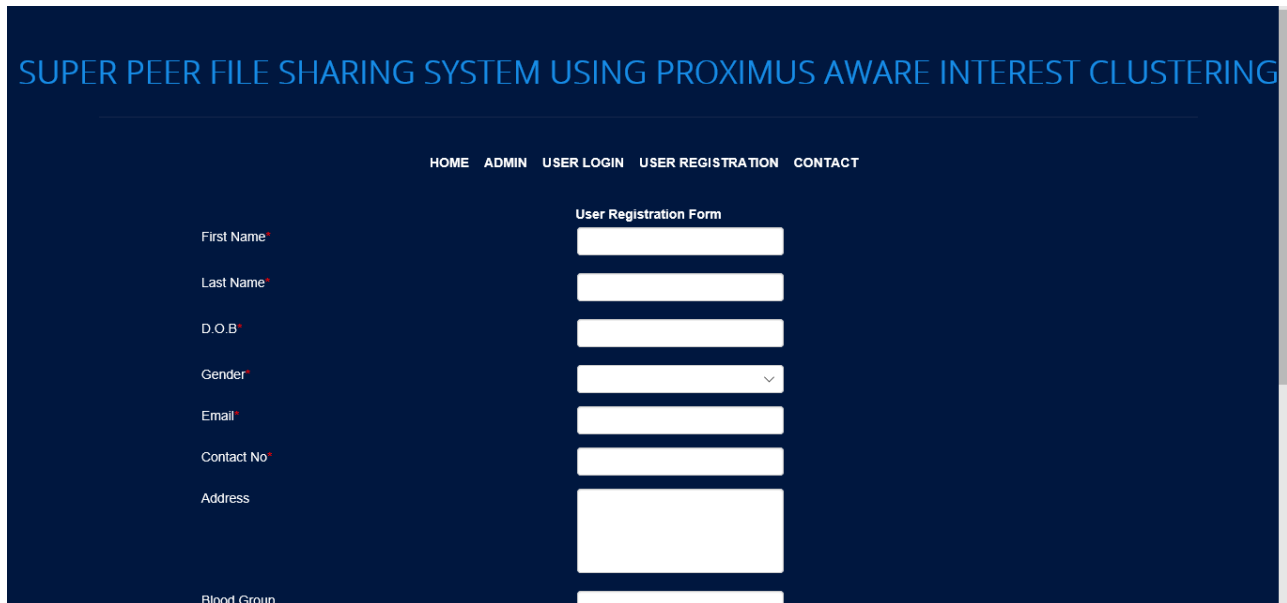
Screenshot 4

SUPER PEER FILE SHARING SYSTEM USING PROXIMUS AWARE INTEREST CLUSTERING

HOME LOGOUT

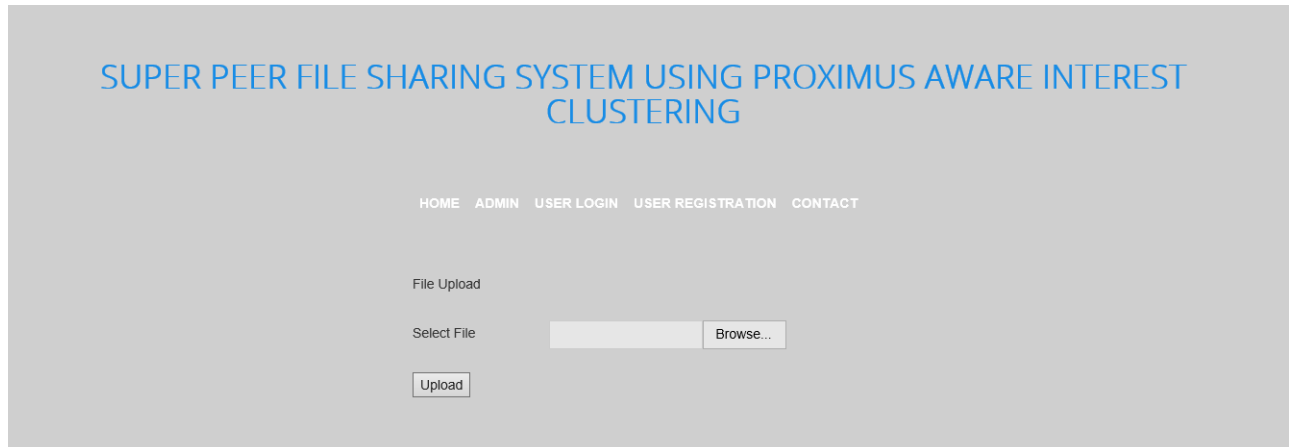
| Files Related To Search Query |                    |
|-------------------------------|--------------------|
| Filename                      | Download           |
| java&dotnet2.txt              | <a href="#">Do</a> |
| java&dotnet.txt               | <a href="#">Do</a> |
| android4.txt                  | <a href="#">Do</a> |
| dotnet&java.txt               | <a href="#">Do</a> |
| java2.txt                     | <a href="#">Do</a> |

**Screenshot 5**



The screenshot displays a web application titled "SUPER PEER FILE SHARING SYSTEM USING PROXIMUS AWARE INTEREST CLUSTERING". The navigation bar includes links for HOME, ADMIN, USER LOGIN, USER REGISTRATION, and CONTACT. The "User Registration Form" is the central focus, featuring input fields for First Name, Last Name, D.O.B, Gender (a dropdown menu), Email, Contact No, and Address. A "Blood Group" field is partially visible at the bottom left.

**Screenshot 6**



The screenshot shows the same web application with the title "SUPER PEER FILE SHARING SYSTEM USING PROXIMUS AWARE INTEREST CLUSTERING". The navigation bar remains the same. The "File Upload" section is active, showing a "Select File" button next to a text input field, a "Browse..." button, and an "Upload" button below them.

## **VIII. ACKNOWLEDGMENT**

Authors want to acknowledge Principal, Head of department and guide of their project for all the support and help rendered. To express profound feeling of appreciation to their regarded guardians for giving the motivation required to the finishing of paper.

## **IX. CONCLUSION**

In recent years, to reinforce file location potency in super peer systems, interest-clustered super-peer networks and proximity-clustered super-peer networks are planned. Though each method improves the performance of super peer systems, few works cluster peers supported each peer interest and physical proximity at the same time. Moreover, it's more durable to comprehend it in structured super peer systems as a result of their strictly outlined topologies, though they need high potency of file location than unstructured super peer. During this paper, we tend to introduce a proximity-aware and interest-clustered super peer file sharing system supported a structured super peer. It teams peers supported each interest and proximity by taking advantage of a hierarchical data structure of a structured super peer.



## REFERENCES

- [1] Genaid Stéphanie and Choonan Rattananoka "Large-scale experiment of co-allocation strategies for peer-to-peer supercomputing in P2P-MPI." *Parallel and Distributed Processing, 2008. IPDPS 2008. IEEE International Symposium on*. IEEE, 2008.
- [2] Halevy Alon Y et al "Piazza: Data management infrastructure for semantic web applications." *Proceedings of the 12th international conference on World Wide Web*. ACM, 2003.
- [3] Lin Guoxin, Haivong Shen, and Lee Ward. "An efficient and trustworthy P2P and social network integrated file sharing system." *IEEE transactions on computers* 64.1 (2015): 54-70.
- [4] Lin Fangming, et al. "Onota: Rationing server resources in peer-assisted online hosting systems." *Network Protocols, 2009. ICNP 2009. 17th IEEE International Conference on*. IEEE, 2009.
- [5] Ratnasamy, Sylvia, et al. "Topology-aware overlay construction and server selection." *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE. Vol. 3*. IEEE, 2002.
- [6] T. Asano, D. Ranjan, T. Roos, E. Welzl, and P. Widmaier, "Space filling curves and their use in geometric data structure," *Theoretical Comput. Sci.*, vol. 181, no. 1, pp. 3–15, 1997.
- [7] H. Song, S. Dharmapurikar, J. Turner, and J. Lockwood, "Fast hash table lookup using extended bloom filter: An aid to network processing," in *Proc. Conf. Appl., Technol., Archit., Protocols Comput. Commun.*, 2005, pp.181–192.
- [8] PlanetLab. (2012) [Online]. Available: <http://www.planet-lab.org/>
- [9] K. Psounisa, P. M. Fernandezb, B. Prabhakarc, and F. Papadopoulosd, "Systems with multiple servers under heavytailed workloads," *Perform. Eval.*, vol. 62, pp. 456–474, 2005.
- [10] F. A. Haight, *Handbook of the Poisson Distribution*. New York, NY, USA: Wiley, 1967.
- [11] H. Shen, "PAIS: A proximity-aware interest-clustered P2P file sharing system," in *Proc. IEEE/ACM Int. Symp. Cluster Comput. Grid*, 2009, pp. 164–171.