pFusion: A P2P Architecture for Internet-Scale Content-Based Search and Retrieval
Demetrios Zeinalipour-Yazti, Vana Kalogeraki and Dimitrios Gunopulos

Charalampos S. Nikolaou
charnik@di.uoa.gr

Department of Informatics and Telecommunications

3 June 2008
Introduction

Unstructured P2P Networks
Gnutella v4.0
Paper Objectives

Content-Based Search in P2P
Problem Definition
Search in Unstructured P2P Networks
Experimental Evaluation

Topologically-Aware Overlay Networks
Network Mismatch
Network Topologies
Experimental Evaluation

pFusion
pFusion Architecture

Conclusions
Unstructured P2P Networks

- A set of nodes (peers) with same responsibilities (no client-server differentiation).
- Every node is connected to a set of other nodes and all form an overlay network (a logical/application level network over a physical one).
- No global knowledge of any kind is maintained at any peer.
- Resource discovery is done by message exchange between neighbouring peers.
Bootstrapping is the process during which a peer connects to a P2P network.

- A Gnutella complied P2P network provides a number of central servers.
- Each server maintains a set of online peers in its so called hostcache (or GWebCache).
- The peers in a hostcache satisfy certain properties (long uptime / light-loaded / permanent servents) that make them ideal to respond to pings.

Note that the neighbours of a peer may span over geographically long distances (e.g. from Alaska to Madagascar).
Bootstrapping (cont.)

- To join such a network (e.g. LimeWire), a peer must connect to one of its servers and get a list of online peers.
- Then it sends to a subset of them a *ping* message and connects to a specific number of peers which will answer back with a *pong* message.
- These peers form the *neighbours* (or *routing table*) of the newly connected peer and may span over geographically long distances.
Query Routing

A query is submitted to the network in a BFS-like manner. The notion of TTL (Time-To-Live) is inserted in order to bound the depth of the search space. A peer issuing a query:

- sends to all of its neighbour peers the query,
- each such peer decrease the TTL, forwards the query respectively and
- matches the query against its local storage returning the results (QueryHit) to the peer from which it received the query.

Figure: Query/QueryHit
The paper proposes the **pFusion** architecture which endeavours to improve:

1. the accuracy of the query results (deals with query routing) and
2. the network latency between geographically distant peers (deals with the neighbour set).
Problem Definition

Setting:
- A network of peers where each node maintains a collection of documents.

Goal:
- Effectively query the distributed documents by keywords.
- Consume the less possible network resources.
Agnostic Techniques

a) TTL-based Breadth-First-Search (BFS)
   - Each peer forwards the query to all its neighbors.
   - Excessive network and resource consumption.

b) Random BFS (RBFS)
   - Each peer forwards the query to a random subset of neighbors.
   - Some important segments may become unreachable.
Techniques using Past Statistics

a) Most Results in Past Heuristic (>RES)
   - Query peers with the most results in the last K queries.
   - Usually explores the larger network segments, but fails to explore peers with the most relevant content.

b) Intelligent Search Mechanism (ISM)
   - Each peer maintains a query/queryhit profile for its neighbours.
   - Uses the cosine similarity to drive the queries to the results.

Figure: Querying P2P network using ISM
Intelligent Search Mechanism (ISM)

1. Profile mechanism: (LRU replacement policy)

<table>
<thead>
<tr>
<th>Query</th>
<th>Connection &amp; Hits</th>
<th>Timestamp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stevie Ray Vaughan</td>
<td>(peer1,20), (peer4,50), ...</td>
<td>100002222</td>
</tr>
<tr>
<td>Bireli Lagrene</td>
<td>(peer2,10)</td>
<td>100065652</td>
</tr>
<tr>
<td>Eva Cassidy</td>
<td>NULL</td>
<td>100022453</td>
</tr>
</tbody>
</table>

2. Relevance Rank (RR): Ranking neighbours by similarity and queryhit. For a query $q$ and each neighbour $P_i$ the $RR$ is defined as:

$$RR(P_i, q) = \sum_{j=\text{QueryHit by } P_i} Qsim(q_j, q)^a \times S(P_i, q_j)$$

, where $Qsim$ is the cosine similarity and $S(P_i, q_j)$ is the number of results returned by $P_i$ for query $q_j$.

Note: if $v_1$ and $v_2$ are the featured vectors of resources $r_1$ and $r_2$ then the similarity between them is defined as the inverse of the angle of $v_1$, $v_2$. 
Search: Experimental Evaluation

- The ISM achieves in some cases 100% Recall Rate while using 40 – 50% less Messages and 30 – 40% less Time than BFS.
- Scales well to large environments (since only local information is utilized).
- Performs best with high locality of queries.

Figure: Recall Rate and Message exchange for different querying methods
Network Mismatch

Nature of P2P networks:

- they are usually *network-agnostic* (recall that a peer in Alaska may have a neighbour in Madagascar).
- *Physical* with *Overlay* network mismatch (messages are routed physically through the Internet, but logically, peers constitute the application routers).
- The network mismatch between the Physical and the Overlay layer results in high latencies and excessive network resource consumption.
- Smaller latency $\Rightarrow$ Faster interaction and higher data transfer rates.
Network-Efficient Topologies


- Random topology (Network-agnostic).
- Short-Long (SL) topology (Network-aware).
- Binning SL (BinSL) topology (Network-aware).
Random Topology

- Each peer randomly connects to $k$ other peers.
- This is the technique used in most systems (implementing Gnutella v0.4 protocol — FrostWire, LimeWire, etc.).

**Advantages:**
- Simplicity.
- Needs only local knowledge.
- Leads to connected topologies if $\text{degree} > \log n$.

**Disadvantages:**
- Doesn’t take into account the underlying network.
- Excessive network resource consumption.
Short Long Topology

- Build a global latency adjacency matrix.
- Each peer connects to $k/2$ closest peers (Short Links).
- It then connects to $k/2$ random peers (Long Links).

*Note:* Choosing only Short Links yields disconnected topologies.

**Consequences:**
- The construction of the adjacency matrix requires global knowledge (e.g. each peer pings its neighbors and sends this info to a centralized index).
- Impractical technique due to index size.
BinSL Topology

1. Each node calculates the $RTT$ to $k$ well-known landmarks.
   - The numeric ordering of the landmarks defines the bin of a node.
   - Furthermore latencies are divided into level ranges, e.g. $Level_0 = [0, 100) \text{ms}$, $Level_1 = [100, 200) \text{ms}$, $Level_3 = \text{rest}$.
   - $\text{BinCode} = \text{Landmarks} : \text{Levels} = l_2 l_1 l_3 : 011$

2. Each peer then connects to $k/2$ peers that have the same bin code.

3. It then connects to $k/2$ random peers.

Consequences:
- Depends on the number and quality of landmarks.
- Bin codes have to be stored in a central database.
- Both the central database and the landmarks may become a point of bottleneck.
DDNO — Distributed Domain Name Order

**Motivation:** 58% of the Gnutella network (300,000 IPs) belongs to only 20 ISPs.

- Connect to $d/2$ nodes (siblings) in the same domain (locate them without any global knowledge).
- Connect to $d/2$ random nodes.

**Solution:** Deploy a ZoneCache which tells a node towards which direction to move (done by the DDNO Module).

*Figure:* Domain Name Lookup in a DDNO topology
Domain Name Lookup

1. A peer connects to $d/2$ (geographically random) peers according to Gnutella’s bootstrapping method and sends to one of them a *lookupDN* msg.

2. Each peer that receives *msg* forwards it to one of its neighbours consulting its *ZoneCache*.

3. When a peer in the same domain name receives *msg* broadcasts it to its siblings and they all respond with a *LookupOK* msg to the initial peer.

<table>
<thead>
<tr>
<th>Split-Hash</th>
<th>Neighbor</th>
<th># Hops</th>
<th>Timestamp</th>
</tr>
</thead>
<tbody>
<tr>
<td>9A78DF</td>
<td>Socket3</td>
<td>3</td>
<td>100000000</td>
</tr>
<tr>
<td>421CDE</td>
<td>Socket1</td>
<td>2</td>
<td>10012000</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>2AB356</td>
<td>Socket1</td>
<td>2</td>
<td>10160000</td>
</tr>
</tbody>
</table>
Random/DDNO Performance

- We perform a query and measure the delay until the expected answer arrive.
- We observe that a DDNO network minimizes this delay for all search methods (BFS, RBFS, >RES and ISM) by 30% over Random.

**Figure:** Query answering delay in Random and DDNO network topologies
pFusion Architecture

By Merging the ISM method and the DDNO Module over a DDNO network topology we take the pFusion architecture.

Query Routing:
- Only pose the query to the sibling peers.
- If the results are not satisfactory reissue the query to all neighbours.
Time-efficient when there is a locality of interests (e.g. news / events).
Conclusions

1. Organizing the overlay network using only local information is **feasible** and it leads to **significant improvement in query latency**.

2. ISM succeeds high recall rates using a bare minimum of messages.

3. But... what about the time needed for bootstrapping in a DDNO network? The paper does not provide any experiment on this issue.
References

Demetrios Zeinalipour-Yazti, Vana Kalogeraki and Dimitrios Gunopulos,
*pFusion: A P2P Architecture for Internet-Scale Content-Based Search and Retrieval*, IEEE Transactions on Parallel and Distributed Systems, vol. 18, no. 6, June 2007.

Demetrios Zeinalipour-Yazti,

Karbhari, P. Ammar, M. Dhamdhere, A. Raj, H. Riley, G. F. Zegura, E.

*The Gnutella Protocol Specification v0.4*. 1.
The End

Thank you!