Should Internet Service Providers Fear Peer-Assisted Content Distribution?

Thomas Karagiannis, UC Riverside

Pablo Rodriguez, Microsoft Research Cambridge
Konstantina Papagiannaki, Intel Research Cambridge

http://www.cs.ucr.edu/~tkarag
P2P networks emerge as content distribution solutions

• No major infrastructure investments.
  – Capitalizing on the bandwidth of end-nodes

• Self-scalable
  – Capacity grows at the same rate as the demand

• Resilient to “flash crowd” events
  – The network spontaneously adapts to the demand
The distribution cost is shifted to the Internet Service Providers!

- ISPs indirectly act as distribution servers
  - Peers become servers
  - Increase of ISP egress traffic

- No revenue from serving the content
  - Increased bandwidth requirements but no extra revenue
Client/server vs. P2P content distribution

Transit ISP

Broadband ISP

Transit ISP

Broadband ISP

Transit ISP

Internet

server

Broadband ISP

Broadband ISP

http://www.cs.ucr.edu/~tkarag
Client/server vs. P2P content distribution

Increase in egress traffic!!

$$ \implies \$$$$$
Locality or caching can reduce egress link usage
Locality or caching can reduce egress link usage

http://www.cs.ucr.edu/~tkarag
Our contributions: An empirical cost-benefit analysis using real Internet traces

• We quantify the impact of peer-assisted content distribution solutions on:
  – the ISPs
  – the Content Providers
  – the end users

• We establish the potential for locality-aware “peer-assisted” solutions.

• We evaluate easily deployable architectures for efficient peer-assisted content distribution.
BitTorrent

- Tit-for-tat
  - Choke/unchoke
  - No free-riding

- Three entities:
  - Tracker
    - Coordinates the distribution
  - Torrent
    - Meta-info file
  - Peers
    - Seeds, Leechers

http://www.cs.ucr.edu/~tkarag
Outline

• P2P content distribution: The view from an edge network
  – Examine the potential for locality:
    • File hit ratios
    • Peer overlap in time
  – Potential bandwidth savings
  – Performance implications for the end user

• Impact on ISPs: A global perspective
  – Impact on downloaded/uploaded traffic volumes per ISP
  – Impact on the content provider

• Locality Algorithms and their Performance
• Implications of locality

http://www.cs.ucr.edu/~tkarag
The view from an edge network: Traces

- Packet-traces with machine readable headers
  - Residential (3 traces)
    - 25/34/29 hours, 110 - 130 Mbps
    - 1M-5M IPs
    - web (35%), p2p (32%)

- BitTorrent:
  - 13%-15% of the traffic
The view from an edge network: Methodology

1. Reconstruct all BT flows
   - Tracker requests/responses
   - Peer messages (e.g., handshake, HAVE, etc)

2. Identify individual peers per file
   - Pitfalls: NATs, Proxies, Random peer IDs

3. Quantify savings if locality were present
   - Identify “unnecessary” downloads

http://www.cs.ucr.edu/~tkarag
The view from an edge network: Hit ratios & user overlap

• Hit ratio: How many users request the same content?

<table>
<thead>
<tr>
<th></th>
<th>January</th>
<th>April</th>
<th>May</th>
</tr>
</thead>
<tbody>
<tr>
<td>File Hit Ratio</td>
<td>14%</td>
<td>10.4%</td>
<td>18.2%</td>
</tr>
<tr>
<td>Byte Hit Ratio</td>
<td>12%</td>
<td>9.6%</td>
<td>13%</td>
</tr>
<tr>
<td>Piece Hit Ratio</td>
<td>6%</td>
<td>6%</td>
<td>11.8%</td>
</tr>
</tbody>
</table>

• User overlap: Number of simultaneous active users for the same file?
  – 30%-70% of the time peers coexist
The view from an edge network:
Potential savings

• Two scenarios:
  – Caching (all downloaded bytes are available)
  – Peer-assisted (bytes in active users are available)

70%-90% of existing pieces are downloaded externally while 50% of these pieces exist in active users
The view from an edge network: Implications for end-user

• Locality will improve end-user performance:
  – Wider bottlenecks locally
  – Higher throughput paths

24% of the clients experience >50% faster downloads

Peer id

http://www.cs.ucr.edu/~tkarag
Outline

• P2P content distribution: The view from an edge network
  – Examine the potential for locality:
    • File hit ratios ---- (6% -18%)
    • Peer overlap in time ---- (~60%)
  – Potential bandwidth savings ---- (50% p2p, 70%-90% cache)
  – Performance implications for the end user ---- (50% faster for 24% of the population)

• Impact on ISPs: A global perspective
  – Impact on downloaded/uploaded traffic volumes per ISP
  – Impact on the content provider

• Locality Algorithms and their Performance
• Implications of locality
Impact of Peer-Assisted Content Distribution on ISPs: A global perspective

• Traces:
  – BT Tracker log of Redhat v9.0 distribution.
  – April-August 2003

• Network partition in ASes using BGP tables
  • May and August 2003 BGP tables
Content distribution scenarios

1. Server /server farm/CDN
2. P2P random-matching
3. BitTorrent-like P2P
4. Peer-assisted content distribution + locality
5. Distributed caching

http://www.cs.ucr.edu/~tkarag
A global perspective: Metrics of interest

• ISPs:
  – Ingress traffic per ISP (total & 95th percentile)
  – Egress traffic per ISP (total & 95th percentile)
  – Performance vs. ISP size
  – P2P vs. caching

• Content provider
  – Bytes served
A global perspective: Ingress traffic

Ingress traffic is reduced by a factor of 2 with locality

Requires only roughly 1.5 times the peak capacity compared to caching

Downloaded data (in MB) by each ISP. Percentages show savings compared to client/server.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Average</th>
<th>95th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client/server</td>
<td>14137</td>
<td>804</td>
</tr>
<tr>
<td>P2P</td>
<td>13954 (1.3%)</td>
<td>794 (1.3%)</td>
</tr>
<tr>
<td>BT</td>
<td>13784 (2.5%)</td>
<td>786 (2.2%)</td>
</tr>
<tr>
<td>P2P+locality</td>
<td>6710 (52.5%)</td>
<td>625 (22.3%)</td>
</tr>
<tr>
<td>Caching</td>
<td>1191 (91.6%)</td>
<td>459 (42.9%)</td>
</tr>
</tbody>
</table>

http://www.cs.ucr.edu/~tkarag
A global perspective: egress traffic

Average uploaded data (in MB) by each ISP. Percentages show savings.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Average</th>
<th>95th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client/server</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>P2P</td>
<td>17239</td>
<td>750</td>
</tr>
<tr>
<td>BT</td>
<td>17551</td>
<td>759</td>
</tr>
<tr>
<td>P2P+locality</td>
<td>2827 (84%)</td>
<td>238 (68%)</td>
</tr>
<tr>
<td>Caching</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Each ISP is required to upload just over a copy of the file (1.9 GB)

http://www.cs.ucr.edu/~tkarag
A global perspective: Savings vs. ISP size

ISPs with more than 30 active users experience >60% savings

http://www.cs.ucr.edu/~tkarag
Impact of Peer-Assisted Content Distribution on ISPs: Content Provider

Locality results in less than half the resource requirements compared to the client-server scenario

Total egress server capacity

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Average</th>
<th>95&lt;sup&gt;th&lt;/sup&gt; percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client/server</td>
<td>59.8 TB</td>
<td>17 TB</td>
</tr>
<tr>
<td>P2P+locality</td>
<td>28.4 TB (52.5%)</td>
<td>8.1 TB (52.3%)</td>
</tr>
<tr>
<td>Caching</td>
<td>5 TB (91.6%)</td>
<td>1.6 TB (91%)</td>
</tr>
</tbody>
</table>

http://www.cs.ucr.edu/~tkarag
Locality algorithms and their performance

• Locality algorithms:
  – implemented by ISPs
    • proxy-trackers
    • consistent with peer-assisted locality analysis
  – imposed by content providers
    • IPs grouped by prefix/domain rules
• Imposed solutions are not as efficient
  – Fail to match AS boundaries (contrary to proxy-trackers)
  – 50% of the optimal solution

Downloaded data (in MB) by each ISP for different locality algorithms.

<table>
<thead>
<tr>
<th></th>
<th>/24</th>
<th>/16</th>
<th>DOMAIN</th>
<th>Hierarchical</th>
<th>Proxy Tracker</th>
</tr>
</thead>
<tbody>
<tr>
<td>P2P Locality (Avg)</td>
<td>13964 (1.2%)</td>
<td>11643 (17.7%)</td>
<td>10864 (23.1%)</td>
<td>10227 (27.5%)</td>
<td>6710 (52.5%)</td>
</tr>
<tr>
<td>P2P Locality (95th)</td>
<td>779 (3.1%)</td>
<td>698 (13.2%)</td>
<td>709 (11.8%)</td>
<td>689 (14.3%)</td>
<td>625 (22.3%)</td>
</tr>
</tbody>
</table>

http://www.cs.ucr.edu/~tkarag
Issues and implications

• Peer-assisted vs. existing content distribution solutions
  – Peer-assisted solutions need to address:
    • Availability when population is limited
    • e2e connectivity (NATs)
    • Security
    • Reliability

• Impact of peer-assisted content distribution on internal ISP traffic
  – Re-engineering of internal traffic may prove costly for certain ISPs
Summary

• Current P2P solutions are not “ISP-friendly”
  – Unnecessary traffic downstream & upstream.

• Locality-aware peer-assisted solutions:
  – Decrease egress traffic by a factor of two.
  – Provide >60% savings for ingress traffic.
  – Approximate the performance of a caching architecture in terms of peak load.
Everybody wins!

- Peer-assisted + locality content distribution:
  - CDNs:
    - Push more content with less infrastructure
  - ISPs:
    - Serve more content at the same cost
  - End-users:
    - More content faster