Improving the Speed of Peer to Peer Backup Systems with BitTorrent

Dan Bouffard
Marc Green
Backup Solutions

- Local Storage
- Cloud Services
  - Dropbox
  - Google Drive
  - Microsoft OneDrive
- Dedicated Backup Services
- Peer to Peer (P2P) Backup Systems

P2P Backup Systems
P2P Backup Systems
P2P Backup Systems

Data

Data

Data

Data

Data
P2P Backup Systems
P2P Backup Systems
P2P Backup Systems
Research in P2P Backup Systems

- Research in this area
  - Handling node churn
  - Completely decentralizing
  - Efficient routing algorithms
  - Minimizing storage and network overhead

- We focus on increasing the speed of data backup and recovery
BTBackup

- P2P Backup System that uses BitTorrent
  - Peers that hold replicas become a BitTorrent swarm
BTBackup

- P2P Backup System that uses BitTorrent
  - Peers that hold replicas become a BitTorrent swarm

Peer E parallelizes the recovery of its data in the backup system
BitTorrent

- Protocol for quickly transferring large amounts of data between a group of peers
  - Peers download chunks of data from the initial seeder
  - While downloading, they share chunks with each other
  - Upon receipt, they, themselves, become seeders

- Peers download from multiple seeders simultaneously, increasing speed
  - Due to upload/download bandwidth differences
A BitTorrent peer parallelizes the download of data
Implementation

- Prototype
- Fast, modular, platform independent
- C++11
  - Boost
  - cURLpp
  - Jsoncpp
- BitTorrent Sync
BitTorrent Sync

- P2P synchronization tool allowing users to share files between devices
- Uses a modified version of the BitTorrent protocol
- Web API
- Closed source
Methodology

- Environments
  - Minimal network for speed tests (1 peer, 5 nodes)
  - 100 node P2P network using SZTAKI Cloud for the churn test, with realistic churn and bandwidth limitations

- What we tested
  - Speed of backup and recovery
  - Performance differences between large and small files
  - Overhead caused by churn
Results (1 of 3)

Stacked Speed of Recovering Data from Nodes' Perspective (Moving Average)
Results (2 of 3)

Speed of Recovering Data from Peer's Perspective
(5 per. Moving Average)

- **One 1 GB File**
- **Ten 100 MB Files**
- **Hundred 10 MB Files**
Results (3 of 3)

- 100 peers each backed up a 100 MB file
- Test scenario simulated churn for 1 hour
  - Table 1 shows churn created an additional 146.12 MB of data per peer
- This test scenario simulated a 16 day period, meaning, on average, a peer can expect to replace a replica every 11 days

<table>
<thead>
<tr>
<th>Data without Churn</th>
<th>Data with Churn</th>
</tr>
</thead>
<tbody>
<tr>
<td>58 GB</td>
<td>72.27 GB</td>
</tr>
</tbody>
</table>

Table 1. Amount of data transferred in a 100 node network over 1 hour
Conclusion

- We take advantage of the existence of replicas to turn them into a BitTorrent swarm
  - BitTorrent can download files much faster by combining limited upload speeds of peers
- Data transfer speed increases by up to 300%
- Average amount of churn creates 150% overhead per file
Future Work

- Incentivizing uptime
- NAT Traversal
- Finding the optimal number of replicas to create per backup
Acknowledgements

Mihály Héder, our SZTAKI Project Advisor

Gábor Sárközy, our WPI Project Advisor

Worcester Polytechnic Institute

MTA SZTAKI

The employees of MTA SZTAKI
Thank you
Questions?