The Hidden Locality in Swarms

John S. Otto
Fabián E. Bustamante
Northwestern U.
Locality in P2P systems

- Avoid cross-ISP traffic
- Better performance
- Lower ISP costs, reduced peering link congestion

- Much work toward finding and connecting to local peers
  - Ono, P4P, ALTO

- Why are we still talking about this?
  - Does much locality exist?
  - How can a peer find it?
Widespread locality

- Focus on popular torrents → where challenge of finding locality is hardest
- Use Dasu platform of widely-distributed nodes
- Repeated requests to all trackers → approximate full knowledge of the swarm
Widespread locality

Average number of peers seen

![Graph showing widespread locality]

Top 1%: hundreds of peers

Peers per network vs. Percent of networks
Widespread locality

Average number of peers seen

20% of networks:
5+ local peers
...but it matters when you look

For instance, in a single network (Virgin UK)

![Graph showing normalized number of peers online over local time, with a significant increase of 150% between 11/10 and 06.

150% increase over the day]
Widespread locality, but it matters when you look

**distance = diurnal pattern strength**

**Small networks:**
larger change

**Time-of-day plays a larger role where locality is hardest to find!**
Widespread locality, but it matters when you look

At the peak hour, another 30% of networks have 5+ local peers

Locality exists, but can peers find it?
Peer discovery uses random sampling

- Trackers
- Distributed Hash Table (DHT)
- Peer EXchange (PEX)
- All use random sampling, making it difficult to find locality

Why focus on trackers?
- DHT, PEX disabled in darknets, private torrents
- We turn a key disadvantage of trackers to our benefit!

If you can’t beat ‘em…
- Can we work around the limitations of random sampling?
More is better

- Ask for larger sample
  - 2x increase in peers

- Query more frequently
  - Trackers did not rate-limit
  - Order-of-magnitude increase

*Client-side ways to speed peer discovery*
Torrent site diversity works against the user!

Torrent site choice limits to a subset of trackers and the peer population...

Solution: use all the trackers!
Some torrent sites are better than others

Benefits of using all trackers:
1) More sources of peers
2) Access to more of the swarm, more local peers
More trackers = access to more of the swarm

20% of torrents: miss part of the swarm no matter what site used

CDF of torrents

% increase in available swarm population

Best choice

median

Worst choice
More trackers = access to more of the swarm

![Graph showing CDF of torrents vs. % increase in available swarm population. The graph indicates that up to 50% increase is possible.]
Summary

• Widespread locality exists
  – 50% of networks have 5+ local peers
  – Time-of-day is a key factor: 200% locality increase during peak hour
• Client-side modifications improve discovery rate by >100x
• Using all trackers is essential to discover available locality
• These techniques boost the efficacy of end-host-based biased neighbor selection
Why we are still talking about locality

- Sustainable growth requires efficient use of resources
  - local copies of content offload the network
- When systems ossify or are beyond our control, we...
  - cope as best we can
  - build on top – end-host solutions