

Comparing Hybrid P2P Systems

Authors:

Beverly Yang

Hector Garcia-Molina

Discussant: Wenli Liu
University of Waterloo

Outline

- Initial Thoughts / Contributions
 - Why Hybrid P2P, Not C/S?
 - Problems With This Paper
 - Directions of Further Research
 - Conclusion
 - Discussion
-

Initial Thoughts / Contributions

- A brand new research area Opened
 - Sufficient background information provided
 - Four architectures & two login flavors generalized
 - Preliminary recommendations provided
-

Why Hybrid P2P, Not C/S?

With Hybrid P2P:

- Query flooding
- Reliability
- Data synchronization
- Selfish peers
- Higher transmission cost
- Longer client response time

With Client/Server:

- Could establish a number of identical servers, and distribute them as close to clients as possible
-

Why Hybrid P2P, Not C/S? (cont.)

- Could event other servers immediately about any changes at a local server
- Storage is very cheap now
- Caching technologies could be adopted

Benefits with C/S

- Central administration
 - Single query
 - Shorter client response time
 - Better use of the underlying network resource
-

Problems With This Paper

- Improper title. "comparing hybrid, music-sharing p2p systems" instead.
 - Models are supposed to be general while most parameters are derived from and verified in OpenNap or Napster without consideration of systems from other domains.
 - Most of the formulas remain unexplained even in the longer version
-

Problems With This Paper (cont.)

- Too many assumptions to be very convincing.
 - No explanations on the fitting procedure in verifying query model.
 - The Maximum supportable users is the only performance metric. Other necessary metrics such as response time, etc.
-

Directions of Further Research

- Verification of the models has to be done in domains other than music-file sharing.
 - Query behavior varies with time, so time should be incorporated into to the query model.
 - Response time is simply neglected when comparing performance while it really matters.
 - The interaction between performance metrics.
-

Directions of Further Research

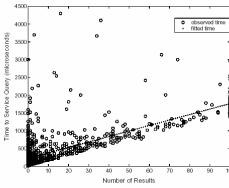


Figure 6: Time to Service a Query vs. Number of Results Returned

- Most of the formulas are linear in essence. So it's hard to capture the characteristics of complicated problems such as P2P performance. Quadratic, cubic, or even higher order formulas are necessary.
-

Conclusion

- Does P2P justify the overhead it introduces?
 - Generally, the models are not very convincing.
 - Further researches are needed in the area.
-

Discussion

Thanks for listen to me patiently!
