Evaluation of Top-k queries in Peer-to-Peer Networks using Threshold Algorithms

Ioannis Chrysakis¹, Constantinos Chalkidis², Dimitris Plexousakis¹,²

hrysakis@ics.forth.gr  conholk@csd.uoc.gr  dp@ics.forth.gr

FORTH, Institute of Computer Science & ² Computer Science Department, University of Crete, Greece

Motivation:
• Threshold top-k query processing algorithms promise limited bandwidth consumption and fast query answering.
• No thorough evaluation in distributed environments has been conducted to date.

Contributions:
• A Detailed Evaluation of exact top-k threshold algorithms in peer-to-peer networks.
• Two extended versions of the Hybrid Threshold algorithm.
• Measuring applicability of threshold techniques in peer to peer data management systems.

The HT-p2p Algorithm

Key Concepts: Super-Peer Topology, Storing Ability, (Fully) Distributed Top-k Query Processing

Evaluation:
Measured: Bandwidth usage and execution time.
Data distributions: normal, zipfian, uniform, synthetic (IMDB), horizontal data partitioning.
Parameters: Number of peers, number of hosted objects per peer, data distribution.
Method: Simulation through PlanetSim Simulator.

Main Results:
• HT-p2p plus is in general the most economic in terms of bandwidth usage.
• Low calculation times as compared to the total ones in HT-p2p plus.
• Linear increase for HT-p2p and HT-p2p plus up to 200 peers and become sharper for 400 or 500 peers in total execution time.
• Bad performance of TPUT in all cases except uniform data distribution.
• Naïve Algorithm had the worst bandwidth performance, but optimal time performance. More experiments with fully distributed simulation are required to evaluate HT-p2p plus time performance.