A Tool for Frequent Items Discovery in Large-Scale P2P Networks

Emrah Cem, Sena Efsun Cebeci and Oznur Ozkasap
The University of Texas at Dallas & Koç University

Introduction

For large scale distributed systems, designing energy efficient protocols and services has become significant while considering conventional performance criteria like **scalability**, **reliability**, f**ault-tolerance** and s**ecurity**.

Due to its extensive applicability in diverse areas, we consider frequent item set discovery problem in this context.

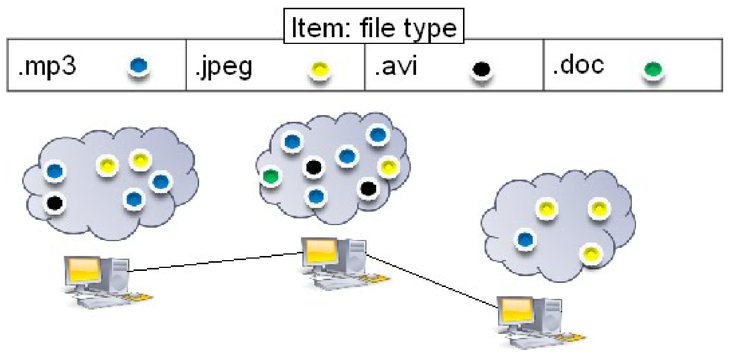
A **simulation model** is developed for ProFID protocol, which is a distributed protocol is developed to find frequent item set discovery in unstructured networks on Peersim.

Objectives

• Providing users to analyze the effects of the protocol and network parameters on different network topologies such as Barabasi-Albert, Erdos-Renyi etc.

• Providing users to do experiments and analysis on different algorithms:
ProFID, adaptive ProFID, hierarchical ProFID and Push-Sum.

ProFID: Protocol for Frequent Item Set Discovery



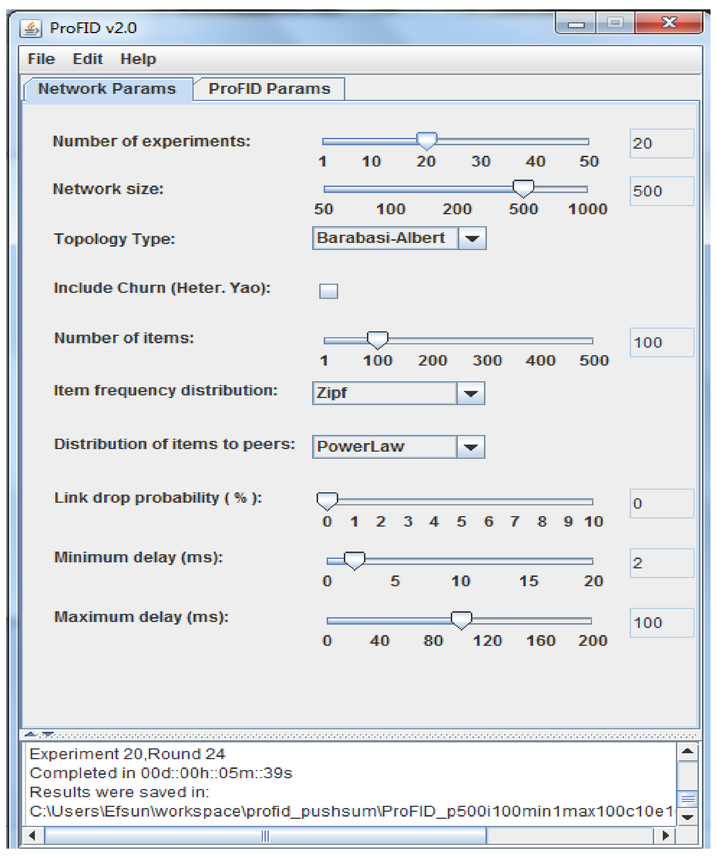
Which file types are frequent?

• Items with global frequencies above a threshold is detected.

• Supports various distributed applications such as cache management, military attack detection, worm detection, DDoS attack detection and topology optimization.

ProFID Toolkit

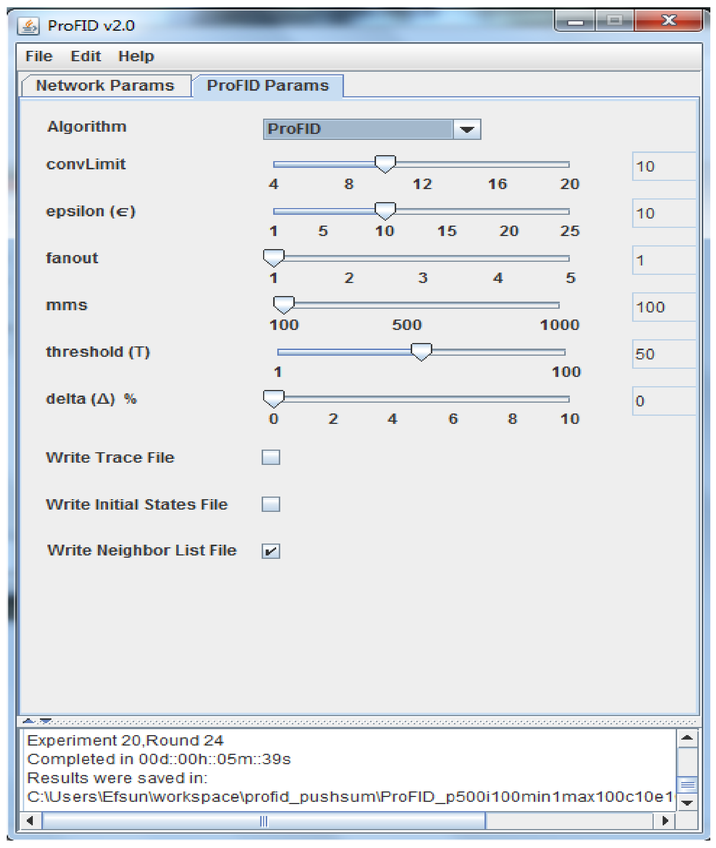
Network Parameters



• Network parameters and protocol parameters can be easily set from the interface.

• Experiments , completion time and destination folder is shown in the console.

Protocol Parameters



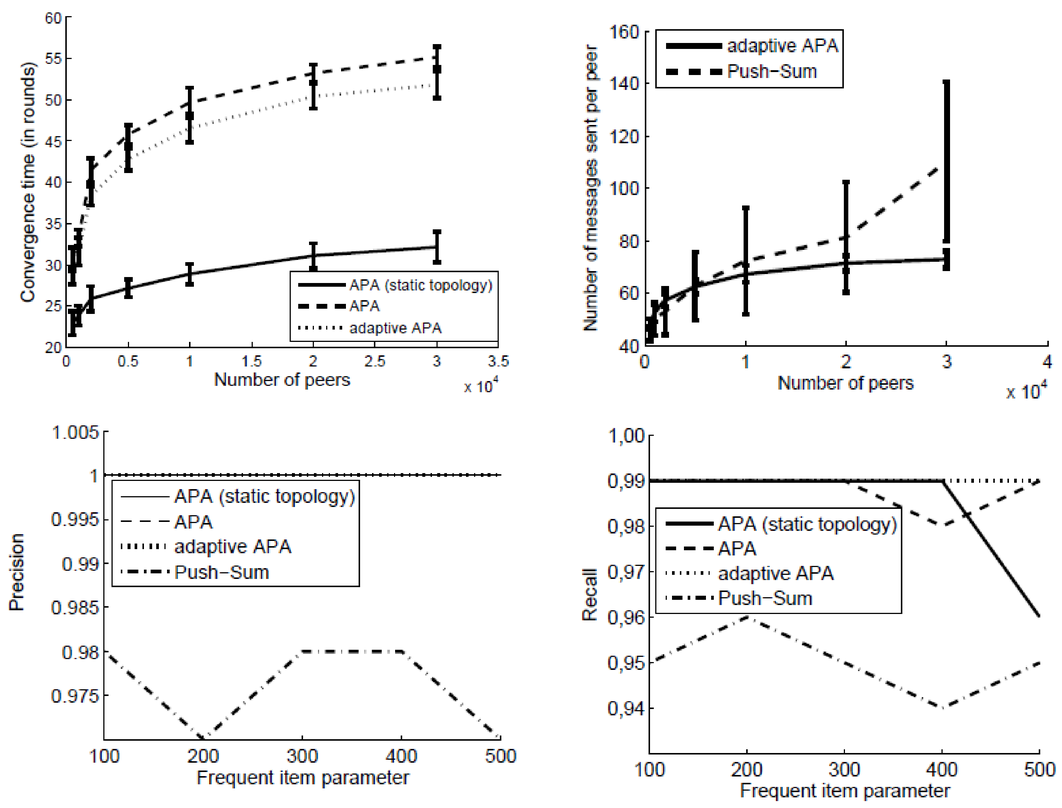
We focus on **performance metrics** :

Convergence time: measures how fast the algorithm converges

Number of messages sent per peer: measures the energy efficency of the algorithm

Precision / Recall: measures how accurate the actual and estimated frequent items based on true positives, false positives and false negatives.

Experimental Results



Conclusions

• Provide a toolkit for extensive analysis on large scale P2P networks for different type of frequent item discovery (FID) algorithms.

•Propose a distributed hierarchical gossip-based approach using dominating set algorithm

References

1. Indranil Gupta, Anne-Marie Kermarrec, Ayalvadi J. Ganesh, Efficient and Adaptive Epidemic-style Protocols for Reliable and Scalable Multicast, IEEE Transactions Parallel and Distributed Systems, 2006
2. Davide Frey, Rachid Guerraoui, Anne-Marie Kermarrec, Boris Koldehofe, Martin Mogensen, Maxime Monod, Vivien Quéma, Heterogeneous Gossip, Proc. of Middleware, Berlin, 2009
3. Emrah Cem, Oznur Ozkasap, ProFID: Practical Frequent Item Set Discovery in Peer-to-Peer Networks, ISCIS, London/UK, September 2010.
4. Emrah Cem, Ender Demirkaya, Ertem Esiner, Burak Ozaydın, Oznur Ozkasap, Energy Cost Model for Frequent Item Set Discovery in Unstructured P2P Networks, ISCIS, London/UK, September 2011
5. http://peersim.sf.net

Supported by the COST (European Cooperation in Science and Technology) framework under Action IC0804, and by TUBITAK (The Scientific and Technical Research Council of Turkey) under Grant 109M761.