Toward an Efficient Support of Complex Queries on Structured Peer-to-Peer Networks

Mots clés :
- Directeur de thèse : Pierre Sens
- Co-encadrant(s) :
- Unité de recherche : Laboratoire d'informatique de Paris 6
- Ecole doctorale : École Doctorale Informatique, Télécommunications, Électronique de Paris
- Domaine scientifique principal: Divers

Résumé du projet de recherche (Langue 1)

Supporting complex queries in large-scale, dynamic networks and acquiring an exhaustive list of results in a timely manner is now a common issue for a wide variety of distributed applications such as music/movie storage, peer-to-peer persistent games, scientific computation, data mining, and many types of data warehouses. Peer-to-peer (P2P) systems offer an abundance of resources, with a higher data storage potential than most organisations would ever be able to afford individually. Moreover, P2P systems tend to be resilient to faults due to the diversity and decentralised nature of their structure. Although, not all P2P topologies combine scalability and efficient data localisation, but structured P2P systems do. Firstly, they have shown to be highly scalable, with the potential to scale up to millions of nodes. They build a Distributed Hash Table (DHT) which achieves load balancing and scalability by using a uniform hash that maps data to a node in a common name-space. Secondly, a DHT is particularly efficient in exact match queries. DHT-based P2P overlays, such as Chord, Pastry and Kademlia, can perform a lookup operation in $O(\log N)$ hops, where $N$ is the size of the network. Regrettably, they do not support complex queries efficiently, since the use of uniform hashing destroys data locality. Several solutions in the literature attempt to provide efficient support of complex queries over DHT-based P2P networks. However, most of them provides poor performance in terms of load balancing, message traffic, query latency and system scalability. This thesis proposes 3 methods to support in an efficient manner complex queries over DHT-based P2P networks: We propose a new cache for Prefix-Trees called Tabu Prefix Table Cache (TPT-C) inspired on the Tabu search heuristic. TPT-C allows to reduce message traffic, query latency and also to improve load balancing among nodes. We propose a novel architecture and search method called PORQUE. Our solution build an additional ring to support complex queries in an efficient manner. The additional ring called leaf ring is composed only by nodes that stores data. To access the leaf ring, each node keeps an history of the leaf nodes contacted, thus complex queries are resolved in $\log(L)$, where $L$ is the number of leaf nodes in the ring. We propose a new indexing model called ECHO. ECHO is self-aware indexing structure which allows to improve the complex queries performance over prefix trees.