CS 856: Project Proposal
W Anthony Young - 20161423
October 28th, 2004

Database systems have been in use for many years as a means to store information. Large organizations store records related to employees, customers and products, while healthcare professionals and researchers store information related to patients and research projects. Databases have the power to organize our information, and provide a quick and easy way to access it.

For the past several years, distributed databases have entered the information storage arena. They allow for local autonomy, improved query performance, improved reliability of data and availability of access, and a high level of expandability and easy data sharing. Distributed databases provide an organization the flexibility to tune storage and access protocols to suit their infrastructure. For example, data that is used frequently can be replicated many times to speed up access. As well, data can be spread amongst many servers to increase redundancy.

Recently, peer-to-peer architecture has been employed in database systems. Peer-to-Peer architecture allows a system to act as both the client (for performing queries and interacting with a user) as well as a server (to provide results to queries posed by other clients on the network).

Peer-to-Peer databases pose many interesting implementation challenges. They are also fundamentally different from distributed databases in several key ways:

- Nodes may join and leave a peer-to-peer network at any time: In distributed databases, nodes are added out of necessity (i.e. for redundancy or growth) and are known to the cluster ahead of time.
- The schema for a peer-to-peer database is not global: In distributed databases, the schema is standardized across each node. In a peer-to-peer database, there may be several schemas used to represent the same data on different nodes.
- The data in a peer-to-peer database might not be complete: Distributed databases contain a complete set of information in each cluster. However, a group of peer-to-peer systems might not have the complete set of information required to accurately and completely answer a query.
- Queries in peer-to-peer databases must be routed to many nodes: In distributed databases, the query can be routed to a relatively small set of nodes. In peer-to-peer databases, the query must be passed to many nodes in order to return an accurate result set.

Peer database systems can be applied to many fields. For example:

- Contact information management: Suppose user A has the contact information for user B, and user C requires it. Instead of phoning user B or looking up the information in an address book, user C could query the contact database of user A to get the information.
- Development environment configuration management: In order to ensure that developers working on a software project are making use of the same development conventions, standard formatting and naming, etc. are required to ensure that each

piece of the project looks and works the same. A peer database system can distribute this information and important project documents to developers.

- Genomics: gene researchers often wish to share data with others and search to see if a discovery they have made has already been found. Each researcher makes use of their own database. A peer system would be capable of searching each database in parallel, possibly obtaining more results.
- Healthcare: suppose John is taken to hospital in Toronto after fainting. How will doctors know that the heart surgery John recently had in Ottawa could be the cause of his ailment? A peer database system would allow doctors to query hospitals around the country for John's medical history, possibly saving his life.

There are several key issues of importance to peer-to-peer database systems. They help govern how well peer-to-peer database systems function. Among them is security and user authentication. Until now, security services have not been implemented in peer-to-peer database systems, as the idea of organizing databases in a peer manner is relatively new. I propose to investigate current trends in peer-to-peer security and determine how they might be applied to peer-to-peer database systems. Specifically, I would like to investigate:

- Authentication of users: User authentication is extremely important in order to facilitate sharing of sensitive information across an organization without the fear of that information being used by those who are not privileged to see it.
- Data Encryption: To prevent data from being stolen while in storage or transit, it should be encrypted.

The paper will begin with an overview of peer-to-peer database systems. Following the overview, critical analysis of the security services offered by APPOINT, DBGlobe PeerDB, and Edutella (some peer-to-peer database systems) will take place. Next, some current security proposals will be presented and evaluated. Finally, a summary will be presented to discuss how the security protocols might be applied in a peer-to-peer database system. An implementation will be suggested and, should time permit, be completed and analyzed.

Milestones

a. Things I will achieve
    i. Complete a thorough literature survey of peer database systems
    ii. Complete a thorough literature survey of authentication and encryption protocols
    iii. Complete a critical analysis of some selected peer database systems
    iv. Complete a critical analysis of some selected authentication and encryption protocols
b. Things I plan to achieve
    i. Complete a discussion of how selected authentication and encryption protocols might be applied, and how well they might perform with, selected peer database systems

ii. Propose an experiment to implement some authentication and encryption protocols on top of some peer database systems as well as metrics to measure how well the systems perform under attack

c. Things I might achieve
   i. Complete a test implementation of some authentication and encryption services in a peer database system and determine if the security they provide is adequate to safeguard critical documents