Finding Information in a Complex Yet Sharing World: A Decentralized System Architecture for Multidimensional Search

NSF Highlight - Award # CNS-0615055/0753066 (2006-2009)
Principal Investigator: Dr. Duc A. Tran, UMass Boston

Lessons from the way financial services recovered after the 09/11 attack suggested for greater decentralization in the nation's communication networks so that any future attack on a single trading site would not stop the system. Decentralization would also avoid the capacity limitation of central servers such as that incurred by traditional search engines which can only index a tiny portion of the entire accessible Internet content.

Investigation on decentralized networks is the main focus of Duc Tran's research group at UMass Boston. These networks, whose examples include grid networks and sensor networks, link together geographically distributed resources and allow them to be shared by any node and used collectively to support execution of large-scale applications, whether single or collaborative.

Of particular interest is the support for information retrieval services. How do we find information in a complex yet sharing world, without using a central agency such as Google? Nowadays, anybody, anywhere, virtually, can generate data thanks to advanced multimedia technologies. The search for useful information, however, remains challenging. A known reason is the curse of dimensionality. As the information is rich with many attributes associated, and the network can be large, it is very difficult for a system's performance to sustain its growth.

Also, there lacks a decentralized system design that can serve as the unified foundation to deploy a wide range of future search applications. This is where Tran's project makes its contribution. The developed solution allows for the network to be efficiently organized in a decentralized manner. This networking design is independent from the data, yet offers convenient mechanisms to deploy retrieval services that may differ from each other. In retrospect, earlier solutions design the network based on the characteristics of the data in deployment, thus much less flexible in support of new data services. For example, if a network is built to support a search application in 10 dimensions, the network would need be rebuilt for a new search application dealing with 20 dimensions. This is not the case in Tran's technique which also offers fast search time and fairly balanced search load.

The results have been published this year in the Journal of Computer Communications and will be presented in the upcoming NSF-PI workshop at IPDPS 2008. Not only exciting scientifically, but the project has its broader impacts. In a complex and decentralized world it could now be easier to search for information located anywhere from anywhere. Educationally, the project has provided research opportunities to several students including two female graduate students and one African-American undergraduate student.