Data Replication and Synchronization in P2P Collaborative Systems

Fatos Xhafa
Technical University of Catalonia
Barcelona, Spain
fatos@lsi.upc.edu

Abstract—Data replication and synchronization techniques have recently attracted a lot of attention of researchers from P2P computing community. Such techniques are fundamental to increase data availability, reliability and robustness of P2P applications. However, several issues arise, such as data consistency and designing cost-efficient solutions, due to the highly dynamic nature of P2P systems. Although there has been a lot of research work from distributed systems and database research community on replication and synchronization, P2P systems have many features that call for investigating on data replication and synchronization techniques in such systems. Recently, techniques for P2P systems are also being investigated in the context of Cloud computing system, to develop Cloud based applications that benefit from features of P2P systems.

While data replication and synchronization have shown useful in the broad context of P2P systems, in this talk, we focus on data replication and synchronization for P2P super-peer collaborative systems. In such systems peers are organized in peer-groups and collaborate together synchronously and/or asynchronously to accomplish a common project by sharing documents and data (contacts, calendar information, etc.), also known as P2P groupware systems. Such systems are attractive for several application contexts such collaborative work in online teams in virtual campuses and small to medium corporates. These applications are especially interesting due to their low cost of deployment and maintenance as compared to the rather high cost centralized groupware applications; also, they provide facilities to support opportunistic collaboration by giving full control to the users.

We discuss the requirements on replication techniques in view of their suitability to supporting efficient peer-group collaboration in P2P systems. Some important requirements include the dynamic nature of the P2P collaborative systems, the need for peers to work both online and offline through available consistent document replicas. We consider dynamics at both peers level, that is to assume peers leaving and joining the system, and also regarding the dynamics of the documents generated during the peer-group activity. The former requires supporting late peer joining so that a peer can update the local copies of its shared documents. In the later case, in many peer-group activities (e.g. in a software project), documents are usually partially changed over time (e.g. some parts of the document are added, revised, deleted, etc.). We discuss XML representation of documents so that changes at an original document at a peer can be efficiently propagated and updated to its replicas at other peers of the peer-group. Some computational results for optimistic replication techniques for groupware systems in multicast and unicast environments are discussed. Additionally, assuming that some peers are mobile, we require the system to be replication-aware so that mobile peers can be aware of changes in the system through real-time notifications of document changes.

On the other hand, we discuss the synchronization of common data to all peers, such as contacts and calendar information, in super-peer model when several mobile peers are connected to the super-peer, which behaves as a small server for the mobile peers. The motivation for data synchronization comes from the needs of geographically distributed teams in business processes to work with intermittent connectivity and thus data synchronization becomes a basic feature. The objective is again to support collaboration of teams of mobile users by enabling anytime and anywhere access to shared contact data. The problem can be formulated as a distributed system problem. We discuss some important requirements such as no modification of server algorithms (the super-peer needs not to perform the contact synchronization task, conflict resolution can be performed at the client side.) and no complex setups at super-peer in order to achieve simple contacts and calendar information sharing. Some computational results will be discussed when using mobile peers under Android system.

Keywords-Data replication, data synchronization, P2P systems, groupware.

Biography: FATOS XHAFA holds a PhD in Computer Science from the Department of Languages and Informatics Systems (LSI) of the Technical University of Catalonia (UPC), Barcelona, Spain. He was a Visiting Professor at the Department of Computer Science and Information Systems, Birkbeck, University of London, UK (2009/2010) and a Research Associate at College of Information Science and Technology, Drexel University, Philadelphia, USA (2004/2005). Dr. Xhafa holds a permanent position of *Professor Titular* at the Department of LSI, UPC (Spain). His research interests include parallel and distributed algorithms, combinatorial optimization, approximation and meta-heuristics, networking and distributed computing, Grid and P2P computing. Dr. Xhafa has widely published in peer reviewed international journals, conferences/workshops, book chapters and edited books and proceedings in the field. Dr. Xhafa has an extensive editorial and reviewing service. He is Editor in Chief of the *International Journal of Space-based and Situated Computing*, and of *International Journal of Grid and Utility Computing*, Inderscience Publishers. He is an associate/member of Editorial Board of several international peer-reviewed scientific journals. He has also guest co-edited several special issues of international journals. Dr. Xhafa is actively participating in the organization of several international conferences. His email is fatos@lsi.upc.edu and his personal webpage at http://www.lsi.upc.edu/~fatos/

IEEE
computer
society