Chapter 13
Publish/Subscribe Techniques For P2P Networks

Cuong Pham
University of Massachusetts, USA

Duc A. Tran
University of Massachusetts, USA

ABSTRACT

P2P is a popular networking paradigm in today's Internet. As such, many research and development efforts are geared toward services that can be useful to the users of P2P networks. An important class of such services is that based on the publish/subscribe paradigm to allow the nodes of a network to publish data and subscribe data interests efficiently. This chapter is focused on the techniques that enable these services in P2P networks.

INTRODUCTION

A publish/subscribe networking system is one in which the nodes can serve the role of a publisher or a subscriber to publish data or subscribe for data of interest, respectively. The publish/subscribe model differs from other request/response models in that a query of the former model is submitted and stored in advance, for which the result may not yet exist but the query subscriber expects to be notified if and when the result later becomes available. The publish/subscribe model is thus suitable for search applications where queries await future information, as opposed to the traditional applications where the information to be searched must pre-exist.

Enabling publish/subscribe services in peer-to-peer (P2P) networks is a topic that has received a lot of attention in recent years. As P2P can be adopted for distributed networking as an effective way to share resources, minimize server costs, and promote boundary-crossing collaborations, a publish/subscribe functionality should be useful to these networks. For example, a monitoring operator in a P2P-based geographical observation network (Teranishi et al., 2008) will be able to

DOI: 10.4018/978-1-61350-110-8.ch013

subscribe a query to receive alerts of fire occurrences so that necessary rescue efforts can be dispatched quickly; or, in a P2P-based scientific information sharing network (Shalaby & Zinky, 2007), a subscriber will be notified when new scientific information is published.

Usually, a publisher node does not know who is interested in its data, and, vice versa, a subscriber node does not know where in the network its data of interest is available. Thus, a challenging problem is to design mechanisms for the subscribers and publishers to find each other quickly and efficiently. A simple way is to broadcast each query to all the nodes in the network or to employ a centralized index of all the queries subscribed and information published. This mechanism is neither efficient nor scalable if applied to a large-scale network.

Consequently, a variety of distributed publish/subscribe mechanisms have been proposed. They follow two main approaches: gossip-based and structure-based. The first approach is designed for any unstructured networks, in which the subscriber nodes and publisher nodes find each other via exchanges of information using the existing peer links, typically based on some form of randomization. The other approach organizes the nodes into some overlay structure and develops publish/subscribe methods on top of it. Examples of such an overlay are those based on Distributed Hash Tables (e.g., CAN (Ratnasamy et al., 2001), Chord (Stoica et al., 2001). The gossip-based approach's advantage is its applicability to any unstructured network, while the structure-based approach is favored for better efficiency.

This chapter provides a survey on the publish/subscribe techniques for P2P networks. First, we will provide some necessary preliminaries. We then discuss several representative techniques in each of the following categories: structure-based, gossip-based, and a hybrid of these two. We conclude the chapter with some remarks.

PRELIMINARIES

Peer-to-Peer Networks

A P2P network is a decentralized network of equivalent-role nodes. A node can serve in either a "server" role or a "client" role, or both, depending on circumstances. Unlike traditional client/server networks, P2P networks have no limit for growth and no single point of failure. The capability to share resources and the freedom to join and leave the network at any time are among the properties that make P2P networks very popular on the Internet today.

There are two main types of P2P networks: structured and unstructured. In unstructured P2P networks, e.g., [Gnutella] and [Freenet], the links between nodes are formed in an ad hoc manner without any predefined structure. Unstructured P2P networks are easy to maintain under network dynamics. They are fully decentralized with a high degree of fairness. However, they are not efficient in search operations. Search in a unstructured network usually requires broadcasting of the query, thus incurring a high communication cost.

Structured P2P networks are designed for better search operations. In such a network, the nodes are arranged in an overlay structure which provides efficient routing and lookup mechanisms. Distributed Hash Tables (DHT) is the most popular structure for structured P2P networks (e.g., CAN (Ratnasamy et al., 2001), Chord (Stoica et al., 2001), Pastry (Rowstron & Druschel, 2001), and Tapestry (Zhao et al., 2004). As CAN is used later in this chapter, let us describe briefly how it works. In CAN (abbr. of "Content Addressable Networks"), the network is viewed virtually as a multi-dimensional geometric space, called the CAN space, in which each node is assigned a location. The CAN space is partitioned into rectangular zones and the node location assignments are determined such that there is only one node in each zone. An overlay neighbor link is created between two nodes if their zones are adjacent.