Chapter 8

Load Balancing in Peer-to-Peer Systems

Haiying Shen
University of Arkansas, USA

ABSTRACT

Structured peer-to-peer (P2P) overlay networks like Distributed Hash Tables (DHTs) map data items to the network based on a consistent hashing function. Such mapping for data distribution has an inherent load balance problem. Thus, a load balancing mechanism is an indispensable part of a structured P2P overlay network for high performance. The rapid development of P2P systems has posed challenges in load balancing due to their features characterized by large scale, heterogeneity, dynamism, and proximity. An efficient load balancing method should be flexible and resilient enough to deal with these characteristics. This chapter will first introduce the P2P systems and the load balancing in P2P systems. It then introduces the current technologies for load balancing in P2P systems, and provides a case study of a dynamism-resilient and proximity-aware load balancing mechanism. Finally, it indicates the future and emerging trends of load balancing, and concludes the chapter.

1. INTRODUCTION

Peer-to-peer (P2P) overlay network is a logical network on top of a physical network in which peers are organized without any centralized coordination. Each peer has equivalent responsibilities, and offers both client and server functionalities to the network for resource sharing. Over the past years, the immense popularity of P2P resource sharing services has produced a significant stimulus to content-delivery overlay network research (Xu, 2005). An important class of the overlay networks is structured P2P overlays, i.e. distributed hash tables (DHTs), that map keys to the nodes of a network based on a consistent hashing function (Karger, 1997). Representatives of the DHTs include CAN (Ratnasamy,
Load Balancing in Peer-to-Peer Systems

In a DHT overlay, each node and key has a unique ID, and each key is mapped to a node according to the DHT definition. The ID space of each DHT is partitioned among the nodes and each node is responsible for those keys the IDs of which are located in its space range. For example, in Chord, a key is stored in a node whose ID is equal to or succeeding the key’s ID. However, a downside of consistent hashing is uneven load distribution. In theory, consistent hashing produces a bound of $O(\log n)$ imbalance of keys between nodes, where $n$ is the number of nodes in the system (Karger, 1997). Load balancing is an indispensable part of DHTs. The objective of load balancing is to prevent nodes from being overloaded by distributing application load among the nodes in proportion to their capacities.

Although the load balancing problem has been studied extensively in a general context of parallel and distributed systems, the rapid development of P2P systems has posed challenges in load balancing due to their features characterized by large scale, heterogeneity, dynamism/churn, and proximity. An efficient load balancing method should flexible and resilient enough to deal with these characteristics. Network churn represents a situation where a large percentage of nodes and items join, leave and fail continuously and rapidly, leading to unpredicted P2P network size. Effective load balancing algorithms should work for DHTs with and without churn and meanwhile be capable of exploiting the physical proximity of the network nodes to minimize operation cost. By proximity, we mean that the logical proximity abstraction derived from DHTs don’t necessarily match the physical proximity information in reality. In the past, numerous load balancing algorithms were proposed with different characteristics (Stoica, 2003; Rao, 2003; Godfrey, 2006; Zhu, 2005; Karger, 2006). This chapter is dedicated to providing the reader with a complete understanding of load balancing in P2P overlays.

The rest of this chapter is organized as follows. In Section 2, we will give an in depth background of load balancing algorithms in P2P overlays. We move on to present the load balancing algorithms discussing their goals, properties, initialization, and classification in Section 3. Also, we will present a case study of a dynamism-resilient and locality-aware load balancing algorithm. In Section 4, we will discuss the future and emerging trends in the domain of load balancing, and present the current open problems in load balancing from the P2P overlay network perspective. Finally, in Section 5 we conclude this chapter.

2. BACKGROUND

Over the past years, the immerse popularity of the Internet has produced a significant stimulus to P2P file sharing systems. A recent study of large scale characterization of traffic (Saroiu, 2002) shows that more than 75% of Internet traffic is generated by P2P applications. Load balancing is an inherent problem in DHTs based on consistent hashing functions. Karger et al. proved that the consistent hashing function in Chord (Karger, 1997) leads to a bound of $O(\log n)$ imbalance of keys between the nodes. Load imbalance adversely affects system performance by overloading some nodes, while prevents a P2P overlay from taking full advantage of all resources. One main goal of P2P overlays is to harness all available resources such as CPU, storage, and bandwidth in the P2P network so that users can efficiently and effectively access files. Therefore, load balancing is crucial to achieving high performance of a P2P overlay. It helps to avoid overloading nodes and make full use of all available resources in the P2P overlay.
Related Content

Optimizing Techniques for OpenCL Programs on Heterogeneous Platforms
[www.irma-international.org/article/optimizing-techniques-opencl-programs-heterogeneous/69805/](www.irma-international.org/article/optimizing-techniques-opencl-programs-heterogeneous/69805/)

Resource Allocation Mechanism with New Models for Grid Environment
[www.irma-international.org/article/resource-allocation-mechanism-with-new-models-for-grid-environment/78893/](www.irma-international.org/article/resource-allocation-mechanism-with-new-models-for-grid-environment/78893/)

Adaptive Mobile Architecture with Utility Computing
[www.irma-international.org/chapter/adaptive-mobile-architecture-with-utility-computing/139854/](www.irma-international.org/chapter/adaptive-mobile-architecture-with-utility-computing/139854/)

Pragmatic Correlations of Quality-of-Experience and Quality-of-Service in IMS-Based IPTV Network
[www.irma-international.org/article/pragmatic-correlations-quality-experience-quality/76922/](www.irma-international.org/article/pragmatic-correlations-quality-experience-quality/76922/)

Software Performance Estimate using Fuzzy Based Backpropagation Learning