Journal of Network and Computer Applications 34 (2011) 20–29

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Coordinated session-based admission control with statistical learning for multi-tier internet applications

Sireesha Muppala, Xiaobo Zhou *

Department of Computer Science, University of Colorado at Colorado Springs, CO 80918, USA

ARTICLE INFO

Article history:
Received 12 March 2010
Received in revised form
9 September 2010
Accepted 11 October 2010

Keywords:
Scalable internet architecture
Session admission control
Statistical learning
TPC-W benchmark
Effective session throughput

ABSTRACT

Popular Internet applications deploy a multi-tier architecture, with each tier provisioning a certain functionality to its preceding tier. In this paper, we address a challenging issue, session-based admission control for peak load management for multi-tier Internet applications. The session-based admission control approach (SBAC) designed for a single Web server is not effective for a multi-tier architecture. This is due to the fact that the bottleneck in a multi-tier website dynamically shifts among tiers as client access patterns change. Admission control based on only the bottleneck tier is not efficient as different sessions impose different resource consumptions at the different tiers. First, we propose a multi-tier measurement based admission control (MBAC), which pro-actively accepts different session mixes based on the utilization state of all tiers. More importantly, we design a coordinated session-based admission control approach (CoSAC) based on a machine learning technique. It uses a Bayesian network to correlate the states of all tiers. The probability with which a session is admitted is determined by the probabilistic inference of the network after applying the evidence in terms of utilization and processing time at each tier to the network. We compare CoSAC with MBAC and a Blackbox approach tailored from SBAC, using the industry standard TPC-W benchmark in a typical three-tier e-commerce website. Experimental results demonstrate the superior performance of CoSAC with respect to the effective session throughput.

1. Introduction

Due to the dynamic nature and scale of the Internet, Internet applications pose great challenges including scalability and availability (Zu, 2005; Zhou et al., 2007). Today, popular Internet applications deploy a multi-tier architecture, with each tier provisioning a certain functionality to its preceding tier and making use of the functionality provided by its successor to carry out its part of the overall request processing (Bouchenak et al., 2006; Chen et al., 2006; Diao et al., 2006a, 2006b; Horvath et al., 2007; Kamra et al., 2004; Lama and Zhou, 2009; Liu et al., 2008; Muppala and Zhou, 2009; Tang et al., 2008; Urgaonkar et al., 2005, 2008; Villela et al., 2007; Weng et al., 2009; Zhang et al., 2007). A typical e-commerce application usually consists of three tiers; a front-end web tier that is responsible for HTTP request processing, a middle application tier that implements core application functionality say based on Java Enterprise platform, and a backend database that stores product catalogs and user orders. The multi-tier computer systems keep growing in scale and complexity. They become so complicated that it is even a big challenge to get a good understanding of the entire system dynamic behaviors (Rao and Xu, 2008a, 2008b).

Admission control and dynamic resource allocation are two critical mechanisms for peak load management and quality-of-service provisioning for Internet services. Admission control is critical when an Internet service is operating at or close to its maximum available capacity before additional resources are provisioned. An admission control mechanism should accept new clients only when the service can guarantee successful transaction completion. Deferring clients at the very beginning of their transactions rather than in the middle is essential to minimize resource wastage (Cherkasova and Phaal, 2002).

Most popular Internet applications are session-based. A session is a sequence of individual requests of different types made by a customer during a single visit to a website (Menascé et al., 2000). SBAC (session-based admission control) is an important approach on e-commerce websites (Cherkasova and Phaal, 2002). It originally proposes to use the effective session throughput defined as the number of completed sessions, instead of request throughput, to evaluate the web server performance. However, SBAC is not effective for peak load management in a multi-tier architecture. This is mainly due to the fact that the bottleneck tier in a multi-tier website dynamically shifts among tiers as client access pattern changes.

The industry standard TPC-W benchmark has browsing, shopping, and ordering mixes. One may argue that under heavy load conditions, only ordering sessions be accepted since they are more likely than other sessions to result in economic benefits for an

* Corresponding author. Tel.: +1 719 255 3493.
E-mail addresses: sdasaraj@uccs.edu (S. Muppala), xzhou@uccs.edu (X. Zhou).

1084-8045/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jnca.2010.10.007