

# Practical Aspects of Performance Analysis (PAPA 2002)

## June 15, 2002

**A joint initiative of the ACM Sigmetrics and the Computer Measurement Group (CMG).  
Sponsorship provided, in part, by HP Labs, HP, Palo Alto, CA.**

### Abstracts/Program

**8:00-8:15 Welcome** (PC co-chairs Danny Menascé and Chris Loosley)

**8:15-10:00 Keynote Speaker:** *“Towards Planetary Scale Computing: next generation internet computing,”* Rich Friedrich, Director, Internet Systems and Storage Laboratory, Internet and Computing Platforms Research Center, HP Labs, Palo Alto, CA.

In the not-too-distant future, billions of people, places and things could all be connected to each other and to useful services through the Internet. In this world scalable, cost-effective information technology capabilities will need to be provisioned as a service, delivered as a service, metered and managed as a service, and purchased as a service. Consequently, processing and storage will be accessible via utilities where customers pay for what they need, when they need it, where they need it. Processing and storage utilities will become as ubiquitous as electrical and water utilities are today. Cooperation among utilities leads towards planetary-scale computing. This talk will describe the rise of the Internet Data Center and why planetary scale computing is an important research topic. We introduce a programmable data center paradigm as a flexible architecture to achieve dynamic resource allocation from a pool of shared resources. The results of workload characterization motivate this work. We will examine a few important results in automated server and storage management. Due to increasing system density and CPU power consumption, we will also explore the critical issue of dynamic thermal management in the data center of the future. Finally, key open research questions pertaining to next generation Internet computing are summarized.

**10:00-10:30 Coffee Break**

**10:30-12:00 Session 1 - Web and Internet**

*“Analyzing a Web based systems performance at multiple time scales,”* Virgilio Almeida, Federal University of Minas Gerais (UFMG), Brazil; Martin Arlitt, Jerry Rolia, HP Labs, USA.

Web and e-commerce workloads are known to vary significantly from hour to hour, day to day, and week to week. The causes of these fluctuations are changes in the number of users visiting a site and the mix of services they require. Since the workloads are known to vary over time, one should not simply choose an arbitrary time interval and consider it as a reference for performance evaluation. We conclude that time scales are of great importance for operational analysis, particularly for systems with bursty loads. Service level agreements must certainly take into account measurement time scales. Similarly, input parameters for predictive models are sensitive to time scale. Ultimately, a time scale should be chosen for service level requirements that best express the needs of end-users and the price the owner of a site is willing to pay for QoS.

*“Performance study of dispatching algorithms in multi-tier Web architectures”* Mauro Andreolini, University of Roma Tor Vergata, Italy; Michele Colajanni, University of Modena, Italy; Ruggero Morselli, University of Modena and University of Maryland.

The number and heterogeneity of requests to Web sites are increasing also because the Web technology is becoming the preferred interface for information systems. Many systems hosting current Web sites are complex architectures composed by multiple server layers with strong scalability and reliability issues. In this paper we compare the performance of several combinations of centralized and distributed dispatching algorithms working at the first and second layer, and using different levels of state information. We confirm some known results about load sharing in distributed systems and give new insights to the problem of dispatching requests in multi-tier cluster-based Web systems.

*“On the Stability of Network Distance Estimation,” Yan Chen, Khian Hao Lim, Randy Katz, UC Berkeley, USA; Chris Overton, Keynote Systems, USA*

Overlay network distance monitoring and estimation can benefit many new applications and services, such as peer-to-peer overlay routing and location. However, there is a lack of such scalable system with small overhead, good usability, and good distance estimation accuracy and stability. Thus we propose a scalable overlay distance monitoring system, Internet Iso-bar, which clusters hosts based on the similarity of their perceived network distance, with no assumption about the underlying network topology. The centers of each cluster are then chosen as monitors to represent their clusters for probing and distance estimation. We compare it with other network distance estimation systems, such as Global Network Positioning (GNP) [1]. Internet Iso-bar is easy to implement and use, and has good scalability and small communication and computation cost for online monitoring. Preliminary evaluation on real Internet measurement data also shows that Internet Iso-bar has high prediction accuracy and stability. Finally, by adjusting the number of clusters, we can smoothly trade off the measurement and management cost for better distance accuracy and stability.

**12:00-1:30 Lunch (on your own)**

**1:30-3:00 Session 2 - I/O and P2P**

*“Disk scheduling policies with lookahead,” Alexander Thomasian and Chang Liu, NJIT, USA*

Advances in magnetic recording technology have resulted in a rapid increase in disk capacities, but improvements in the mechanical characteristics of disks have been quite modest. For example the access time to random disk blocks has decreased by a mere factor of two, while disk capacities have increased by several orders of magnitude. High performance OLTP applications subject disks to a very demanding workload, since they require high access rates to randomly distributed disk blocks and gain limited benefit from caching and prefetching. We address this problem by re-evaluating the performance of some well known disk scheduling methods, before proposing and evaluating extensions to them. A variation to CSCAN takes into account rotational latency so that the service time of further requests is reduced. A variation to SATF considers the sum of service times of several successive requests in scheduling the next request, so that the arm is moved to a (temporal) neighborhood with many requests. The service time of further requests is discounted, since their immediate processing is not guaranteed. An SATF policy prioritizes reads with respect to writes and processes “SATF winner” write requests conditionally, i.e., when the ratio of their service time to that of the winner read request is smaller than a certain threshold. We review previous work to put our work into the proper perspective and discuss some of our plans for future work.

*“A Note on SCSI Bus Waits,” Alexandre Brandwajn, UC Santa Cruz, USA*

In the SCSI-2 standard, the unique IDs of devices on the bus define a fixed priority whenever several devices compete for the use of the bus. Although the more recent SCSI-3 standard specifies an additional fair arbitration mode, it leaves such fair mode an optional feature. Despite a number of allusions to potential unfairness of the traditional SCSI bus arbitration scattered in the trade literature, there seem to be few formal studies to quantify this unfairness. In this paper, we propose a simple model of SCSI bus acquisition in which devices on the bus are viewed as sources of requests with fixed non-preemptive priorities. We use the model to assess the expected extent of unfairness, as measured by the mean bus wait, under varying load conditions. Effects of tagged command queueing are not considered in this note. Numerical results obtained with our model show that there is little unfairness as long as the workload is balanced across devices and the bus utilization is relatively low. Interestingly, even for medium bus utilization a significant fraction of bus requests find the bus free which might correlate with the service rounds noted in a recent experimental study. For unbalanced loads and higher bus utilization, the expected wait for the bus experienced by lowest priority devices can become significantly larger than the one experienced by highest priority device. This appears to be especially true if the higher priority devices have higher I/O rates and occupy the bus for longer periods. As might be expected, even for balanced workloads, unfairness tends to increase with the number of devices on the bus.

*“Probabilistic Scalable P2P Resource Discovery Location Services,” Daniel A. Menascé and Lavanya Kanchanapalli, George Mason University, USA*

Scalable resource discovery services form the core of directory and other middleware services. Scalability requirements preclude centralized solutions. The need to have directory services that are highly robust and that can scale with the number of resources and the performance of individual nodes, points to Peer-to-Peer (P2P) architectures as a promising approach. The resource location problem can be simply stated as “given a resource name, find the location of a node or nodes that manage the resource.” We call this the *deterministic* location problem. In a very large network, it is clearly not feasible to contact all nodes to locate a resource. Therefore, we modify the problem statement to “given a resource name, find with a given probability, the location of a node or nodes that manage the resource.” We call this a *probabilistic* location approach. We present a protocol that solves this problem and develop an analytical model to compute the probability that a directory entry is found, the fraction of peers involved in a search, and the average number of hops required to find a directory entry. Numerical results clearly show that the proposed approach achieves high probability of finding the entry while involving a relatively small fraction of the total number of peers. The analytical results are further validated by results obtained from an implementation of the proposed protocol in a cluster of workstations.

### **3:00-3:30 Coffee Break**

### **3:30-5:00 Panel: “Challenges and Future Directions in Data Analysis for Performance Management.”**

Panelists: Nevil Brownlee, CAIDA; Rich Friedrich, HP Labs; Chris Loosley (chair), Keynote Systems; Chris Overton, Keynote Systems; and Jerry Rolia, HP Labs.

There are many challenging problems to solve as systems become more distributed. Applications that exploit the Web, and associated technologies such as Web services, can be subject to extremely large and volatile workloads, generating massive volumes of measurement data. How can we extract meaning from such a measurement fire-hose and use it for timely performance management decisions? If the traditional techniques that worked for the single system environment are not scalable, what replaces them?