Electronic Commerce Research and Applications 13 (2014) 320–340

Contents lists available at ScienceDirect

Electronic Commerce Research and Applications

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

Behavioral-based performability modeling and evaluation of e-commerce systems

Pece J. Mitrevski [a,*], Ilija S. Hristoski [b]

[a] Department of Computer Science and Engineering, Faculty of Technical Sciences, "St. Clement of Ohrid" University, Ivo Lola Ribar St, 7000 Bitola, Macedonia
[b] Department of Business Informatics, Faculty of Economics, "St. Clement of Ohrid" University, Gjorche Petrov St, 7500 Prilep, Macedonia

ARTICLE INFO

Article history:
Received 22 November 2013
Received in revised form 18 June 2014
Accepted 18 June 2014
Available online 28 June 2014

Keywords:
Consumer behavior
Discrete-event simulation
Performability
Stochastic Petri nets
System availability
Service reliability
Workload characteristics

ABSTRACT

Assuring high quality of web services, especially regarding service reliability, performance and availability of e-commerce systems (unified under the term performability), has turned into an imperative of the contemporary way of doing business on the Internet. Recognizing the fact that customers' online shopping behavior is largely affecting the conduct of e-commerce systems, the paper promotes a customer-centric, holistic approach: customers are identified as the most essential "subsystem" with a number of important, but less well-understood behavioral factors. The proposed taxonomy of customers and the specification of operational profiles is a basis to building predictive models, usable for evaluating a range of performability measures. The hierarchical composition of sub-models utilizes the semantic power of deterministic and stochastic Petri nets, in conjunction with discrete-event simulation. A handful of variables are identified in order to turn performability measures into business-oriented performance metrics, as a cornerstone for conducting relevant server sizing activities.

1. Introduction

As e-commerce becomes mainstream, achieving and retaining customers' satisfaction regarding the quality of web services being delivered becomes an ultimate goal: the principal element of the contemporary way of doing business on the Internet, a crucial factor that affects both the existence and the progress of Internet companies. The complexity to face and cope with this challenge is even bigger, knowing the fact that e-commerce web services rely heavily on large-scale systems, consisting of thousands of computers, networks, software components, and users. Large systems are inherently complex, whilst the randomness in the way customers demand those web services initiate the problem of managing and planning the capacity of hardware resources.

In fact, the true challenge is to achieve an optimal balance among: (1) implementation investments, (2) the costs needed to continually upgrade and maintain a particular e-commerce web site, (3) the performability of the underlying system, (4) the achieved level of customers' satisfaction, all related to (5) the delivered quality of web services. It can be accomplished solely through a proactive, continuous and scientifically-based capacity planning procedure, not by the application of ad hoc, non-regular procedures and rules of thumb, based on one's subjective opinion and experience. In that context, Menascé and Almeida (2002) define the capacity planning as being "the process of predicting when future load levels will saturate the system and determining the most cost-effective way of delaying system saturation, as much as possible," taking into account "the evolution of the workload due to existing and new applications, and the desired service levels".

The absence of such methodology can easily push a company into a "no-win situation:" the inappropriate capacity of hardware resources implies poor performance (long response times), reduced service reliability and system availability (unexpected and frequent crashes, long-lasting downtimes). These, in turn, imply poor QoS levels being delivered to customers, who, due to their increased dissatisfaction, abandon e-commerce web sites massively and prematurely. This can be devastating for the Internet companies, imposing financial and sales losses, decreased productivity, customers' disloyalty and endangered reputation. The average downtime cost per hour may range from thousands to millions of dollars, depending on the type of industry (Patterson 2002).

Based on a Ponemon Institute (2011) study, Emerson Network Power (2011) released a report that reveals and analyzes the financial impact of infrastructure vulnerability. For instance, downtime

* Corresponding author. Tel.: +389 47 207 702; fax: +389 47 203 370.
E-mail addresses: pece.mitrevski@uklo.edu.mk (P.J. Mitrevski), ilija.hristoski@uklo.edu.mk (I.S. Hristoski).

http://dx.doi.org/10.1016/j.elerap.2014.06.005
1567-4223/

can be particularly costly for e-commerce companies with revenue models that depend on the data center's ability to deliver IT and networking services to customers, with the highest cost of a single event topping US$1 million (more than US$11,000 per minute). According to the Information Technology and Intelligence Corp. (2011), their high availability survey and trends for 2010 revealed that while companies cannot achieve zero downtime, one out of ten companies said it needs greater than 99.999% availability.

Moreover, according to Dun & Bradstreet, Inc., which licenses information on businesses and corporations for use in business-to-business marketing and supply chain management, 59% of Fortune 500 companies experience a minimum of 1.6 h of downtime per week. If one takes the average Fortune 500 company (at least 10,000 employees, paid an average of US$56 per hour (US$40 per hour salary plus US$16 per hour in benefits)), the labor part of downtime costs for an organization this size would be US$896,000 weekly, translating into more than US$46 million per year (Arnold 2010). Faced with ever-increasing financial problems, Internet companies cannot keep pace with ever-increasing needs for upgrading the existing hardware infrastructure due to steadily growing service demands, which closes the vicious circle.

In order to avoid such an unpleasant scenario, predictive models have to be built and evaluated on a regular basis. To investigate the impact of the inherent performance characteristics of the hardware infrastructure on the quality of service levels being delivered by the e-commerce system, it is inevitable to take into account the unexpected and stochastic changes in the workload, causing service reliability degradation and traffic bursts (or peak rates) that can easily exceed the average traffic levels by the order of ten (Banga and Druschel 1999, Jiang and Dovrolis 2005). This is crucial, since even small increases in the traffic intensity can significantly degrade system performance, whereas large ones can even cause functional breakdowns, making the e-commerce systems completely unavailable to customers' service requests.

Analyzing the genesis of the workload itself, a single fact emerges: customers' online shopping behavior is the main underlying factor affecting the behavior of e-commerce systems' hardware infrastructure. Actually, customers' behavior during the course of an online shopping session is the true generator of all performability-related problems that arise at the server-side. It is, to a certain extent, irrational and stochastic, as being a reflection of people's intrinsically different mindsets, resulting in rather different and unexpected decision-making processes when shopping online. As a result, customers exhibit heterogeneous, yet predictable navigational patterns, when surfing throughout the e-commerce web site for the duration of online shopping sessions. Capturing those navigational patterns and defining them quantitatively, in line with a relevant taxonomy of customers' behaviors, is a basic premise to performing both modeling and evaluation of a wide gamut of performability measures.

Buying decisions have costs and benefits and qualitative rationale for doing them. But, this also involves people making these decisions – people with their own backgrounds, incentives and experiences. The last couple of years have seen an upswing of researchers who are trying to explain these kinds of unexpected behaviors and why consumers don't make purchasing decisions purely based on micro-economic principles. This blend of the economics and emotions has been subject to investigation of a separate branch in economics, known as behavioral economics (Ariely 2009, Curran 2010). What can we learn from behavioral economics to improve the way we make business decisions and improve business successes? According to Latino (2000): "reliability is a combination of two trains of thought. The first is what one calls the hard side of reliability. The hard side consists of all the tools that you physically go and measure reliability with ... The other train of thought is the behavioral side." What causes people to perform tasks in such a way that it causes performability setbacks?

Scarcity is the basic economic problem that ascends because people have unlimited wants, but limited resources. Because of scarcity, various economic decisions must be made to allocate resources in an efficient manner. Standard economics assumes that we are rational – that we know all the relevant information about our decisions, that we can estimate the value of the different options, and that we are unobstructed in assessing the implications of each possible choice (Ariely 2009). The outcome is that we are supposed to be making rational and pragmatic decisions. On the basis of these assumptions, economists draw in-depth conclusions about shopping trends.

We are all far less rational in our decision-making than standard economic theory assumes. Moreover, our irrational behaviors are neither random, nor meaningless – they are systematic and predictable: some people tend to spend more when shopping online than when they count out the cash; automated bill pay, another favorite of retailers and service providers, makes use of a similar lack of "tangible" payment; many decisions involve weighing the value of money now versus its more unclear future value; and, in many situations consumers are driven by emotion, not thought (Dratch 2014).

A survey by University of Southern California's Center for the Digital Future (2013) says that 76% of adult Internet users buy online, and for 59% of the Internet users "online purchasing has reduced their buying in traditional retail stores" (the most popular online purchases are for books, clothes, and travel, each reported by 66% of online buyers). Nevertheless, 48% of the Internet users are "very concerned or extremely concerned about the privacy of personal information when or if buying online," whereas 44% are "very concerned or extremely concerned about the security of credit card information when buying online."

To put it differently, some customers know exactly what they want (they are focused and passionate) and frequently place an order, whereas others browse or search repetitively and almost never place an order (they are just curious or reluctant), being affected by those security related issues, anxious about the privacy of their personal information when purchasing from their favorite online stores. In conventional economics, the assumption that we are all rational implies that, in everyday life, we compute the value of all the options we face and then follow the best possible pathway. But, we do not always base these judgments on our essential values – our likes and dislikes.

In our holistic approach, we identify customers as the most essential "subsystem" of an e-commerce system, with a number of important, but less well-understood behavioral factors. The modeling of customers' classes and the specification of the operational profiles is a basis to building predictive models, suitable for evaluation of a number of performability measures. Performance and dependability are often modeled separately, based on the assumption that the faults of an individual subsystem do not necessarily affect performance. Conjointly, performability assessment is defined as "quantifying how well the object system performs in the presence of faults over a specified period of time" (Meyer and Sanders 2001).

In this study, the performability evaluation strategy is based on a hierarchical composition approach that combines results from: (1) dependability (continuity of and readiness for correct service, based on system availability and service reliability); and (2) performance models. The proposition is that "the faults" in the system are the behavioral factors at play here, that can undermine one's business goals, and "a failure" is an instance in time when a customer displays behavior that is contrary to some micro-economic principles. Dependable behavior, in our context, refers to human