PERFORMANCE MEASUREMENTS AND HUMAN BEHAVIOUR ON E-COMMERCE SYSTEMS

M. Arce-Urriza\textsuperscript{1}, J.E. Armendáriz-Iñigo\textsuperscript{2}, M. Liroz-Gistau\textsuperscript{2}, J.R. Juárez-Rodríguez\textsuperscript{2}, J.R. González de Mendivil\textsuperscript{2}

\{marta.arce, enrique.armendariz, jr.juarez, mendivil\}@unavarra.es, liroz.47682@e.unavarra.es

\textsuperscript{1}\textsuperscript{1}Departamento de Gestión de Empresas
Universidad Pública de Navarra
Campus de Arrosadía, s/n – 31006
Pamplona (Spain)

\textsuperscript{2}\textsuperscript{2}Departamento de Matemática e Informática
Universidad Pública de Navarra
Campus de Arrosadía, s/n – 31006
Pamplona (Spain)

Abstract. E-commerce on the Internet is becoming an efficient economic model. This technology provides business organizations with useful systems to trade their products. Therefore, there is a great interest to adapt these applications to user behaviour. Benchmark tools can provide information about the performance of these systems. However, these benchmark tools simulate a specified environment, usually very tied with a specific kind of application, what means that obtained results can only be discussed in such environments. As it is reflected in a study we have performed, it is quite difficult to describe consumer behaviour completely, but there are certain circumstances in which some consumer patterns can be identified. Thus, this information could be used to develop better benchmark tools, which provide more accurate results to real user behaviours.

1 INTRODUCTION. E-commerce has become widely used nowadays, from hotel booking to buying a romantic novel or even groceries. Thus, enterprises are very interested in understanding how customers behave online to create environments that satisfy their needs and that facilitate consumer choice, whilst developing and using systems providing higher availability and performance. Data replication through a cluster of databases is an interesting approach for enterprises facing dynamic web content generation and their associated e-commerce transactions, Wu and Kemme (1), Elnikety et al. (2) and Lin et al. (3). This permits to attend a greater number of clients at different sites geographically distributed, which gain access to the system through their closest node and, hence, increasing performance. Besides, in case of failures the load supported by the faulty node may be redirected to the rest of available nodes and therefore system availability is also increased. However, as there are clients concurrently accessing the system, special attention to data consistency has to be paid.

Data consistency is crucial for this kind of applications. The strongest correctness criterion for database replication is One Copy Serializable, Bernstein et al. (4), which means that the interleaved execution of the transactions must be equivalent to a serial execution of those transactions in a single database. This is not an attractive approach for e-commerce systems, since read operations may become blocked, which are most of the operations performed in a transaction. Besides, this criterion requires each replica database providing Serializability, Berenson et al. (5). However, most database vendors and database programmers take advantage of consistency levels below serializability to achieve better performance.

This is the case of the Snapshot Isolation (SI), also defined in (5), that states that transactions read committed data as of the transaction started. Read operations are never blocked as long as the snapshot can be maintained. On the other hand, transaction's writes are reflected in its snapshot. Concurrent updates are invisible to the transaction. In order to prevent lost updates (5), it applies the First-Committer-Wins rule.

\textsuperscript{*} This work has been partially supported by the Spanish Government under research grant TIC2003-09420-C02-02.
Hence, an alternative correctness criterion for database replication has been introduced, Generalized Snapshot Isolation (GSI) (2), considering that database replicas only guarantee SI. GSI transactions may use older snapshots of the system instead of the latest snapshot required in SI, being able to provide better performance without significant increase in the abortion rate when write conflicts are low. Another correctness criterion is One Copy SI, introduced in (3), where a transaction gets the latest snapshot of the system, but may cause read transactions to block occasionally.

It is important to note that there are different alternatives to consider about the implementation of the database cluster. It may be implemented by modifying the core of the Database Management System (DBMS) of each site composing the cluster, including the addition of a communication module such as in (1), (4), Carey and Livny (6), Kemme and Alonso (7), Kemme et al. (8) and Holliday et al. (9). This solution minimizes the system overhead but it is a restricted solution, since it greatly depends on the DBMS used. The alternative approach is to put a middleware layer between clients and DBMSs, such as in (3), Patiño-Martínez et al. (10) and Irún et al. (11), what increases the overhead but permits to get rid of DBMSs dependencies. These systems may employ different replication protocols to maintain data consistency. Basically, all recent replication protocols perform reads on one replica and writes on all (available) replicas since this performs best for read-intensive applications. Since several possibilities are available, it is interesting to compare the performance of them with a standard benchmark tool.

TPC-W (12) is a standard tool proposed by the research and industrial community to measure the performance of DBMSs. It simulates the behaviour of users accessing an online bookstore where they search and buy books. TPC-W has been used by (1), (2) and (3) so as to measure the performance of their different replicated data solutions which are implemented over SI DBMS replicas. TPC-W has different purchase behaviour options, but the selection criterions of these options are exclusively based on randomly generated data.

However, although characterizing consumer behaviour is a difficult task\(^1\), the way consumers behave cannot be said to be totally defined by a random pattern\(^2\). That is why marketing research has tried to describe consumer behaviour patterns so as to help firms and practitioners to develop better marketing strategies. Nevertheless, consumer behaviour online does not necessarily mimic that observed at physical stores, which has been the one traditionally analyzed. There are important differences between physical and virtual stores that make consumers behave differently. For instance, the amount and quality of the information that is available at each channel, the perceived risk or the possibility of using or not personal purchase lists are not the same at these two shopping environments, Burke et al. (13), Alba et al. (14), Otto et al. (15). The amount of information available online is said to be extremely high, but this information is always visual, neither tactile nor from the sense of smell (14), Degeratu et al. (16). So, for products such as towels, in which softness is a valued characteristic, online shopping seems not to be the best shopping option. Thus, a consumer purchasing a towel online will probably show a different behaviour than that the one he would have had at a physical store—it would be logic to think that he could be more loyal to some brand he had previously purchased.

However, his behaviour would not be so different while purchasing a cinema ticket online or offline. Besides, online shopping is also perceived to be more risky than offline shopping, due to the belief that the probability of a fraudulent credit-card transaction (15), Van den Poel et al. (17), and/or getting a product with a lower quality than expected is higher at a virtual store than at a traditional store. On the other hand, the existence of some online tools (e.g. search engines,

\(^1\) We are unable to know and record every factor each consumer considers while purchasing.

\(^2\) In fact, marketing probabilistic models consider that the consumer utility function has a deterministic and a non-deterministic (random) component.
personal purchase lists) that help consumers decide their choice, Häubl et al. (18), and the 24-7 service possibility that is easily offered through the Internet have also made some consumers to choose the online store for convenience. These diverse advantages and disadvantages of the virtual store obviously affect not only the type of consumers who buy online or the type of products purchased, but also how products are bought.

Therefore, despite the fact that TPC-W offers a rich environment that emulates many web service applications, it does not reflect the entire range of web service or application server requirements (12). In addition, the extent to which a customer can achieve the results reported by a vendor is highly dependent on how closely TPC-W approximates the customer application. The relative performance of systems derived from this benchmark does not necessarily hold for other workloads or environments and extrapolations to any other environment are not recommended.

In this paper, we outline the main parameters and their characterization in the TPC-W benchmark tool. Besides, we study the online and offline consumer behaviour through a small empirical application to the purchases made at the online and offline stores of a Spanish grocery chain. This permits us to further discover the consumer behaviour in online environments, which is not total random nature, as opposite to TPC-W. Despite the fact that this study is not based on an online bookstore, it may reflect the great influence of the kind of application considered. This kind of application characterization allows us to consider that measurement of database replication systems are highly dependent of the kind of application considered. Thus, TPC-W will favour some replication protocols or solutions whose operation fits with the proposed bookstore model; however, these replication protocols will not fit into other environments, such as a grocery chain website.

The rest of the paper is organized as follows: Section 2 is devoted to explain the fundamentals of the TPC-W benchmark. In Section 3, we study the user behaviour in online and offline environments analyzed from purchases performed during a year in a Spanish grocery chain for a product category. Section 4 discusses comparisons of both approaches.

2 TPC-W. A TRANSACTIONAL WEB E-COMMERCE BENCHMARK. TPC-W is a transactional web benchmark designed to evaluate e-commerce systems. It specifies a workload that simulates the activities of an online bookstore. Three separate components take part in the interaction: The System Under Test, SUT, comprises all components which are part of the application being simulated; the Remote Browser Emulator, RBE drives the TCP-W workloads creating an Emulated Browser, EB, for each user interacting with the system; and the Payment Gateway Emulator, PGE, represents an external system which authorizes payment of funds.

TPC-W specifies 14 different pages which must be implemented in the SUT. Moreover it defines the schema of the database where data will be stored. There are described 8 tables: CUSTOMER, ORDER, ADDRESS, COUNTRY, ORDER_LINE, CC_XACTS, ITEM and AUTHOR.

An EB emulates the communication between the customers and the system. The interaction is done through sessions, which are a set of consecutive requests to execute some function. The session duration is controlled by a User Session Minimum Duration (USMD) time, defined as the minimum duration for which a session must last. It is generated using the following equation:

\[ USMD = -\ln(r) \mu \]  

(1)

where \( r \) is a random number uniformly distributed between 0 and 1.0, and \( \mu \) is the mean duration of 15 minutes.
Between two requests, the EB waits for a period of time, called Think Time, that is defined as:

\[ TT = -\ln(r) \cdot \mu \]  

(2)

where \( r \) is defined as below and \( \mu \) is the mean think time included between 7 and 8 seconds.

After each interaction the EB must decide which of the navigation options will be chosen. In the TPC-W specification the probabilities of these navigation options are well defined for each of the pages. TPC-W provides three diverse patterns of behaviour for the EBs, called web mixes, varying the ratio of read-only transactions vs. update transactions. The Browsing Mix presents the 95% of read-only transaction as opposed to the 5% of update transactions. The Shopping Mix specifies 80% vs. 20% and the Ordering Mix 50% vs. 50% respectively.

The workload can be adjusted by modifying the values of the mean think time and the number of EBs which take part in the simulation. The size of the database tables is calculated in function of the number of items that the system offers and the number of EBs that participates in the interaction, in order to maintain the scalability of the system.

The TPC-W primary metrics are the Web Interactions Per Second (WIPS) and system cost per WIPS, $/WIPS, calculated using the Shopping Mix. There are also defined another two secondary metrics, corresponding with the browsing mix, WIPSB, and with the ordering mix, WIPSO. TPC-W establishes a response time requirement for each type of web interaction. At least 90% of each type of web interaction must be returned in the time specified.

During the session, the item selection (that guides the TPC-W requests) can be done from three different webs or by clicking at promotional items:

- **The Best Seller Web Interaction**: it shows the 50 most popular items for a concrete subject among the 3,333 most recent orders sorted by descending number of ordered items. The item is selected using a uniform random distribution.
- **The New Products Web Interaction**: it shows the list of the 50 newest products for a concrete subject sorted by descending release date. The item is selected from the list as before.
- **The Search Result Web Interaction**: it shows the list of items which match the criteria given on the previous page, Search Request Web Interaction. There the user selects a search type and defines a search string. TPC-W chooses the search type from a uniform distribution over the values author, title and subject. The search string is filled in function of the selected search type so that a specific match rate is guaranteed. This is achieved by using similar generation functions in the search string and in the field when the database is populated.
- **The Promotional Items**: the promotional items are 5 items whose images are shown on the top of some pages.

TPC-W was developed in response to the increasing need for an industry-standard benchmark that could measure the performance of both hardware and software in an e-commerce environment. However, it does not simulate realistic customer behaviour, as can be inferred from the following Section.

### 3. CONSUMER BEHAVIOUR IN ONLINE AND OFFLINE ENVIRONMENTS

Current advances in technology development and usage, especially those related to the Internet, have brought up an increasing research interest towards the marketing principles guiding consumer choice behaviour at virtual environments, Parasuraman and Zinkhan (19). However, neither are physical stores equal to virtual stores, nor the drivers behind the decision of purchasing online...
or offline, nor the way consumers purchase at this new channel. Thus, it is not so easy to develop good electronic tools and systems that help consumers while purchasing online.

Physical stores allow a real inspection of the products, including their touch and smell (14) and a good notion of their size and volume (13), whilst online stores are much less informative in these issues, but perfect communicating those facts that can be easily transformed into written information or reproduced graphically, such as price, sound reproduction’s speed of a hi-fi system or the place we select to sit on a cinema. However, e-commerce has increased significantly during the last decade. A survey from Morganosky and Cude (22) shows that the main reasons argued by e-grocery shoppers to use the Internet to buy groceries are convenience and time saving. This, joint to the fact that some products face characteristics easily communicated online, could explain why the online store is frequently used to buy products such as concert tickets, airline tickets or hotel booking (17).

Several papers have tried to understand online consumer behaviour during the last ten years (13), (14), (18), Lynch and Ariely (23). Some recent empirical work on grocery shopping (16), Danaher et al. (24), Andrews and Currim (25) have found significant differences between how consumers attracted to shopping online behave relative to consumers shopping in traditional stores. Andrews and Currim show in (25) that the majority of online shoppers behave differently than those purchasing offline. They verify that the parameters that describe the shopping process (brand, size package, loyalty, price, promotion…) are different for online and offline customers. Compared to offline consumers, online consumers are less price sensitive, have a stronger preference for big-size packages, present a stronger loyalty to the package size, and search products by brands more frequently. Danaher et al (24) find strong evidence of brand loyalty being higher online than offline while analyzing more than 100 brands from 19 categories of products. Degeratu et al (16) observe that for products whose principal characteristics are difficult to communicate online, the role of brands is even stronger online than for those whose main characteristics are easy to visualize online. They also observe that the joint effect of price and promotion online is lower online than offline.

But it is also important to note that we can even observe a different consumer choice behaviour for the same product being purchased only at a unique channel. It would depend on the drivers behind that purchase. De Figuerido (21) suggest the category of books as an example, what also helps us to connect this part of the paper with that explaining the origin of TPC-W. Imagine three different books that might be purchased: i) a textbook, ii) a tour guide, and iii) a rare book or one-of-a-kind used book. In the first situation, a student would need to buy a book with a specified title, author and even edition. In such a case, the student would probably try to search over books at the lowest price. In the second situation, however, the tourist may want to search first over different titles, formats and then over price. In the third and last situation, a buyer may want to check the appearance and state of the book at the same time as the price. Consumers are used to face these different types of purchases offline, but not at the new virtual stores.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Online</th>
<th>Offline (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brand loyalty</td>
<td>4,060</td>
<td>3,055</td>
</tr>
<tr>
<td>Size loyalty</td>
<td>2,841</td>
<td>2,453</td>
</tr>
<tr>
<td>Price (2)</td>
<td>-2.479</td>
<td>-3.998</td>
</tr>
</tbody>
</table>

(1) Parameters corrected by Swait and Louviere’s procedure (1993) so as to make results comparable. \( \mu_{\text{eff}} = 0.9 \).
(2) Price is measured in ln.
(3) p-value < 0.05 for all parameters.
(4) We do not show brand and size specific parameters although they are available from the authors.

Table 1: Canned tuna estimation.
3.1 A small empirical application. Considering a one-year record of purchases from one of the five most important Spanish supermarket chains, we observe that some variables –as loyalty and price– have a different impact on consumer choice, while considering the purchases of the same set of consumers online and offline. In concrete, our results for the frequently purchased category of canned tuna\(^3\) show that the effect of loyalty (both brand and size loyalty) is significantly stronger online, whilst lower for price.

These results are in concordance with previous insights in the literature (16, 24, 25), as we have seen. Table 1 depicts the main results obtained. This example shows how a consumer may buy the same category differently while shopping at different channels. Further, it also shows that the previous purchase affects considerably next purchase, since loyalty variables appear to be significantly important. This is even more notable for the case of online purchases.

4. DISCUSSION. We have described TPC-W, a standard tool that has been used as a benchmark to compare the performance of different DBMS solutions. TPC-W has different purchase behaviour options but its criterion selection is totally random. However, although we are unable to describe consumer behaviour completely, there are certain circumstances such as the type of product purchased (e.g., books, groceries, clothes…), the environment in which the purchase is done (online, offline) and the reasons behind the purchase (e.g., a textbook, a novel as a present…) in which some consumer patterns can be identified.

Research in marketing tries to do this through the analysis of consumers purchase histories, surveys of purchase intention or even experiments. If we used this information, we could develop better tools to measure systems performance. Otherwise, it would be possible that we understated or overstated the value of a system that could be very good for the booking of hotels, but not for books.

The paper raises the question of the importance of considering these circumstances while determining the performance of a system. That is, we do not want to say that the use of TPC-W is wrong. We just say that it would be interesting that the tools that measure performance, such as TPC-W, took into account the purpose a system was developed to. Therefore, when evaluating a concrete system, it is important to define precisely where it is going to be used. Then, even if there is not a benchmark which fits exactly to the system, divergence points have to be found so that the results could be more carefully studied.

Nevertheless TPC is working to embrace a widely range of services. In short, it has published a new benchmark, TPC-App, which simulates the activities of a business-to-business transactional application server operating in a 24x7 environment. Moreover, it is working on the release of two new benchmarks: TPC-E, for on-line transaction processing application environments, which defines a mixture of read-only and updates intensive transactions; and TPC-DS, a decision support benchmark that models some aspects of decision support system. Besides they are working on the draft of a new version of TPC-W.

---

\(^3\) The category of canned tuna was selected because it was a category of product frequently purchased online and offline. The choice set includes 11 alternatives (from different combinations of 6 brands, 4 package sizes and olive/other vegetal oil). The data considered for the estimation are 7171 purchase observations (3933 offline and 3238 online).
5 REFERENCES.


