COMPARATIVE ANALYSIS OF THE IMPACT OF SERVER OPERATING SYSTEMS ON WEB SITE PERFORMANCE

Petar Jakić*,
Ali Elsadai,
Milan Tair

Singidunum University,
Belgrade, Serbia

Abstract:

This work analyses the impact of the server operating system on the overall performance of hosted web sites. In the paper, we are presenting the results of comparing the influence of the two most common server operating system families, Windows and the GNU/Linux distributions. Through experiments, we have collected data based on which we have concluded the impact of these operating systems on the performance of a web site developed using the most commonly used web platform - WordPress. The experiment was set up in such a way as to eliminate the impact of hardware on our measurements, which is done by using the same hardware, network and matching conditions. The experiment was conducted with a large number of repetitions to reduce relative errors in measurements. The performance analysis covers files of different sizes and types, including text and multimedia content typical for modern web sites.

Keywords:

GNU/Linux, Windows, web site, performance.

INTRODUCTION

Users tend to leave web sites that require extensive time to load their content [1]. In such situations, users search for alternative sources of information and services [2]. It is confirmed by research that at least one of the important reasons why users leave web sites is the time it takes to load them, compared to other web sites [3] [4]. This can lead to a reduction in sales, for commercially oriented web sites [5]. Whenever a web page loads only one second longer than other similar web sites, it reduces user satisfaction by as much as 16%, page browsing by 11% [6] and conversion rate by 7% [7] [8]. Therefore, the time to load the web site is one of the more important features that affect the satisfaction and success of the web site.

Improving the performance of the web site can lead to greater satisfaction of visitors of web site services, which can have a positive impact on business. Other than optimizing the web site's application code, the

Correspondence:

Petar Jakić

e-mail:
petar.jakic.993@gmail.com

underplaying platform can also impact performance. Therefore, research is needed to determine what actions in preparing the web server platform can be performed which can affect the reduction in the load time of the hosted web site. Since several server features may affect performance, as well as other features of the web application itself [9], this research analyzes the impact of a single feature independently. The research aims to determine the influence of the server operating system on performance.

2. METHODOLOGY

To reduce the error of measuring the speed of loading the content of the web site on selected operating systems the experiment was designed to use the content of different types and sizes. Repeating the measurement many times reduces the relative measurement error. The content used in the experiment was formed to simulate different ranges of content sizes typical of modern web sites [10]. The file sizes used in the experiment were divided into ranges of typical textual content, smaller and larger images, as well as large video files. The research aims to show to what extent and in what way the choice of an operating system affects the performance of the site. Other characteristics of the architecture of the web server on which the web site is hosted and executed are unchanged so that they cannot affect the measurement results. To ensure virtually unchanged conditions regarding other features of this experimental platform, such as local network connection speed, the experiment was performed on a local network, thus excluding network capacity as a possible bottleneck of the model. This reduces the impact of changing the speed of the Internet connection due to changes in the connection load at different times of the day [11]. The web site was loaded from a local server exclusively purposed for the experiment. For both operating system families, the same computer was used as the server, each in term. The computer from which the content download request was made was disconnected from the Internet connection during the experiment. Also, all installed programs were turned off during the measurement. In this way, the influence of other software on the measurement results for the duration of the experiment was reduced. In the study, we tested the effects of systems from the two most commonly used families of server operating systems - Windows and GNU/Linux distributions [12] [13]. In the case of Windows, the latest version at the time of the research was 2019, build version 1903. In the case of the GNU/Linux

system distribution, the Ubuntu Server operating system distribution, version 19.04, was selected, which was up-to-date at the time [14]. The servers was set-up following the required server configuration, using an Apache2 web server [15] with a PHP interpreter [16] and a MySQL relational database management system [17] was used as a web server, and as a platform for web site hosting. The WordPress CMS [18] was installed with default settings, without additional configuration of layout, functionality and optimization. The operating systems have not been in any way modified and additionally configured from their default initial states, automatically configured during their respective setup processes.

The content used for the experiment for the download time measurement was divided into two categories: (a) files and (b) page content. Files are divided into four groups: (1) a small file of 12.9KB, (2) a medium file of 1.08MB, (3) a large file of 90.7MB, and (4) a very large file of 1GB, and pages into two groups: I) a small web page with textual content, with 3 paragraphs of text and (II) a large web page, with 50 paragraphs of text, 4 tables of average size, 30 embedded images, 5 sheets of 10 items and with a total of 100 links. The experiment is set up in such a way that a large number of consecutive requests are made to the server. The server is asked to deliver specified types of content from the web site hosted on the server with the tested operating system. For each request, we record the response time, as well as the response time for all request in the sequence. Requests were not sent competitively, but one at a time, each upon the completion of the previous one. The data collection process was performed for both operating systems.

We have used a Bash script to automatically query each category of content from the web site, which is delivered from pre-defined web site paths. For each content group, two content delivery times were recorded, the total time for the entire cycle of all iterations, and the average individual download time. The number of repetitions in a cycle is automatically increased from 1 to 100, and over 100, it is increased in steps of 10 repetitions, up to 500 repetitions of downloads for each content group. In this way, it is possible to determine whether a web site behaves differently when it needs to serve the same amount of content one or more times in a very short time. Averages and total times are recorded.

A listing of the part of the script that downloads content from the site, measures the total time and writes the results to a file for processing and averaging is shown in the following computer code listing:

```
#!/bin/bash
for repetitions in `cat sizes.txt`
do
    echo "(1) small file, x $repetitions..."
    start=`date +%s%N`
    for i in {1..$repetitions}
    do
        wget --cache=off \
             --no-check-certificate \
             --no-cache --no-cookies \
             -q $URL -O- > /dev/null
    done
    end=`date  +%s%N`
    time=$(( $end - $start ))
    echo "$SIZE $repetitions $time" >>
results.txt
done
```

Listing 1 - A segment of the data collection script

The script uses a supplementary file whose content supports the process of automated data collection for processing. This file is called sizes.txt and contains pre-defined repetition numbers from 1 to 100, with steps of 1 and over 100 to 500, with steps of 10.

Two essential variables used in the script are:

- $URL is the link from which the content from the web page will be downloaded in the given iteration. The value of the variable is set outside the script and changes before each script run for each subsequent content group by size.

- $SIZE is the symbolic name of the content group by size that is retrieved in that iteration.

The script ignores the content downloaded from the web site. It does not store it anywhere, but erases it immediately after the end of an iteration of the current cycle. After completing one download and after it records the measurement, it continues with the next planned iteration. This script is repeated for each group of content sizes and complements the results of the measurement with the results.txt file. Executing the script generates a total of 140 results for each content group, grouped by size. By repeating the process for all content groups, a total of 840 rows of results are created, which contain the symbolic name of the content group, the number of repetitions and the total duration, expressed in ms in the form of a real number value. Based on the duration and number of repetitions, the average content download time was subsequently calculated for each of the 840 result rows in the file and compared to the average calculated on the fly during the download process. This comparison is done for verification.

3. EXPERIMENTAL RESULTS

The completion of the experiment, which included data collection for the two tested operating systems and all content groups, grouped by size on the same hardware and using the same network infrastructure and configuration took 154 hours. Based on the collected measurement results, we were able to perform an analysis and create a report. This report explains the web site's performance in terms of response speed and delivery of requested content.

This section presents the measurement results. Speeds are calculated based on the total and average download times for all content groups. After the calculations were done, the maximum calculated speed for both operating systems was taken into account and the results were recalculated and set on a scale from 0 to 100 MB/s. The speed values are shown in the measurement results section graphically and in tabular form and explained in the following analyses.

3.1. ANALYSIS OF THE RESULTS FOR THE UBUNTU SERVER

Figure 1 shows a graph of average content download speeds for all six content categories, grouped by size, for each number of iterations, from 1 to 500, for the web site hosted on a server running Ubuntu Server. For clarity, the graph is drawn so that the numbers of consecutive repetitions of content retrieval ranges from 1 to 500 and are placed along the abscissa, and the values of average content retrieval speeds are shown in logarithmic scale, shown from 0 to 100, values from 0 to 100MB/s are along the ordinate.
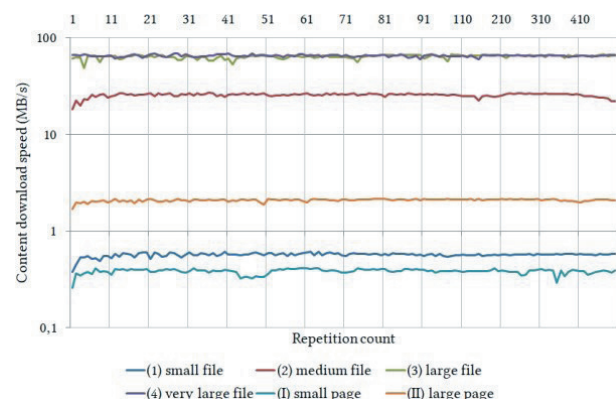

Figure 1: Results when hosting on the Ubuntu system

From these measurement results we can notice that when downloading (I) a small page, the average download speed is lower when the page is downloaded consecutively a smaller number of times (up to 5), and increases for multiple consecutive downloads. The same is the case with (1) a small file. It is noticeable that the download speed is unstable when downloading

(1) small and (2) medium files, as well as (I) small pages, while in the case of downloading (3) large and (4) very large files and (II) large pages speed uniform.

We have measured oscillations in average download speeds during download (I) of a small page in cycles of up to 100 consecutive downloads. For some consecutive downloads, the average speed is the same for this content group.

The least number of oscillations was recorded when downloading the (II) large page. There are fewer oscillations observed when loading this page than when loading (2) medium, (3) large and (4) very large files.

This finding is contrary to our expectations, given that when downloading a file from a web site, the Apache web server accesses the served file directly from storage, without passing the request through the PHP interpreter, which would in term run the logic of the WordPress site. Direct access thus bypasses the connection to the MySQL database server data, and was expected to reduce the overall amount of processing time needed to deliver the file.

When measuring the average download speed (2) of the average file for the whole frame of the number of repetitions, it is noticeable that there is acceleration in situations when the same content is served consecutively up to 20 times. However, over 460 times, a lower average download speed is recorded, but the oscillations are smaller. When downloading (3) large and (4) very large files, the least number of oscillations were noticed, regardless of the number of consecutive downloads of the same file from the server.

Finally, it was observed that the average download speed of (2) medium, (3) large and (4) very large files is in the upper 50% of the speeds compared to other content groups.

3.2. ANALYSIS OF THE RESULTS FOR THE WINDOWS SERVER

Figure 2 shows a graph with the average relative download speeds for all six content categories, for each number of iterations from 1 to 500, from a web site hosted on a server running the Windows Server 2019 operating system.
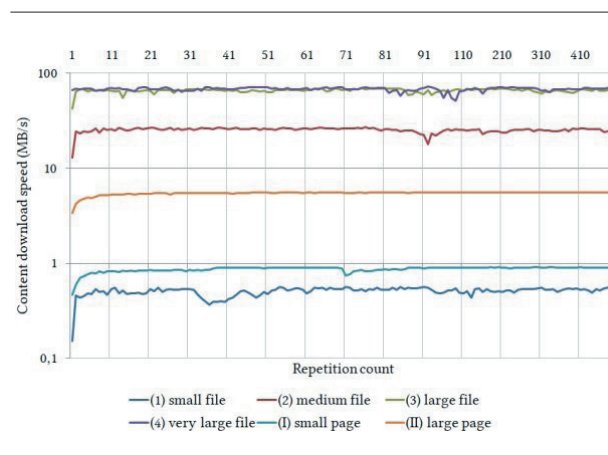

Figure 2: Results when hosting on the Windows Server

When hosting the web site on the Windows Server, we notice similar performance in the results, with some difference in performance evident with certain content groups. In terms of speeds expressed and calculated relative to the maximum speed in this configuration, the least number of oscillations in terms of average speeds were recorded in (2) medium and (4) very large files, while (3) large files recorded more oscillations for various ranges of consecutive download repetitions. When reviewing the results of downloading the contents (1) of a small file, we see that the average download speeds are unstable in the whole range from 1 to 500 repetitions, and it is particularly noticeable that up to 5 consecutive downloads, the speeds are significantly lower. In web page download scenarios, most oscillations were observed when downloading (I) a small file, while downloading (II) a large file was stable, except for the range of 90 to 95 iterations.

In the shown graph, we can see that the download speed for large and giant files is comparable, regardless of the large difference in file sizes.

Finally, we have observed that the average download speed of (3) large and (4) very large files is within the top 50% of speeds compared to other content groups.

3.3. COMPARISON OF THE RESULTS

In this section, we explain how we compared the obtained measurements. We express measurements as values from 0 to 100 MB/s, corresponding to absolute data transfer rates. We present them graphically.

To make it easier to compare the results of the average download speeds, the values were aggregated into four groups, based on the number of performed repetitions of content downloads, for each size group:

- A - The number of repetitions of content download is from 0 to 99;
- B - The number of repetitions of content download is from 100 to 190;
- C - The number of repetitions of content download is from 200 to 290;
- D - The number of repetitions of content download is from 300 to 500.

For each content group, we have made four subgroups, filtering the results by averaged speeds labelled as A, B, C, and D, as was explained above. These groups and their subgroups are hereinafter collectively referred to as Content analysis grouping labels. For further review of the performed analysis, we have labelled these aggregates as:

- A-1, A-2, A-3, A-4, A-I and A-II;
- B-1, B-2, B-3, B-4, B-I and B-II;
- C-1, C-2, C-3, C-4, C-I and C-II and
- D-1, D-2, D-3, D-4, D-I and D-II.

The results are shown in the graphs in Figures 3 and 4 with logarithmic scale in base 2.
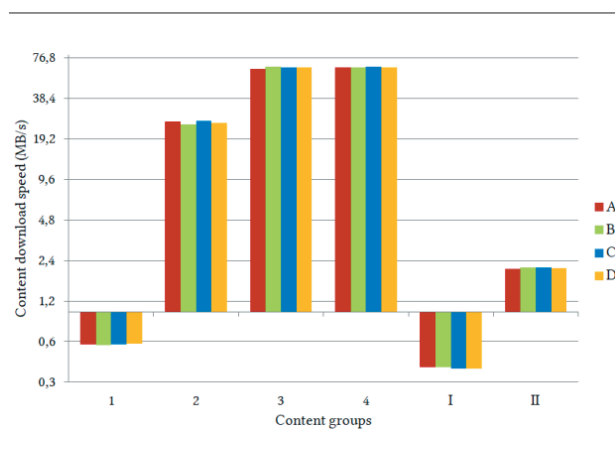
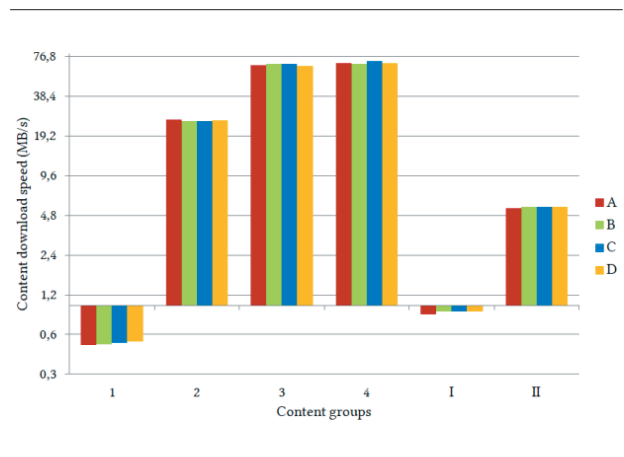Figure 3: Measurement results on GNU/Linux

Figure 4: Measurement results on Windows Server

After reviewing the presented charts, we can conclude that the average download variance is low within the same repetition number group (A through D). Charts show that download speeds of content groups for files (1), (2), (3), and (4) are nearly the same. The variances are better presented in Table 1, shown below. Greater differences are observed in the area showing average download speeds of page contents (I) and (II) between the two experiments. Transfer rates of contents of (I) small and (II) large pages are lower, relative to their respective sizes when downloading from the web site hosted on the Ubuntu Server GNU/Linux distribution. When reading the chart, it is evident that the download speed of the (II) large page from the web site hosted on the Windows Server is double than the achieved download speed of the content of the same page from the web site hosted on the Ubuntu Server.

Table 1 shows the differences in average download speeds for contents of different sizes (1), (2), (3), (4), (I), and (II), grouped by repetition number ranges A, B, C, and D, from web sites hosted on Ubuntu Server and Windows Server operating systems. Table cells include additional indicators which show when the download speed difference is greater than 0 or less than 0, represented by collared arrow symbols, facing up or down, respectively. In our experimental results, we did not come across a situation when the difference was 0. Therefore, we did not include an indicator to show this scenario. From the given table, we can confirm that download speeds across different repetition number groups for (1) small files are nearly the same for both operating systems and that the Ubuntu Server has had better performance in this scenario. When it comes to the (2) medium file size, we can see increasing instability, while on average, the difference is still negligible in this scenario.

Starting from the (3) large file size, it is evident that the download speed difference is steadily increasing and is growing in favour of Windows Server. Regarding web page contents, the download speed difference for (I) small pages is also negligible, while it does show that Window Server outperforms Ubuntu Server.

However, the difference for (II) large pages is comparably larger. On average, across all repetition number groups, it shows that the average download speed is 3,51MB/s greater when the pages are loaded from web sites hosted on the Windows Server, compared to that hosted on Ubuntu Server.

4. DISCUSSION

A comparative analysis of the results of the web site performance measurement performed on Ubuntu Server 19.04 and Windows Server 2019, build 1903 operating systems showed that the web site's performance on the Ubuntu Server system was slower for most content groups compared to the speed of content download from the same web site hosted on the Windows Server operating system. When it comes to content downloads which completely bypass the WordPress application code, by performing file downloads exclusively through the Apache2 web server file download handler, the performance of the web site hosted on the Ubuntu Server better only for the smallest file sizes. As file sizes grow, the download performance gradually shifts in favour of the Windows Server hosted web site. For content that does rely on the WordPress application code for generation and delivery, the download speed was always faster when downloaded from the web site hosted on the Windows Server, compared to the Ubuntu Server. In the case of downloading content (1) small files and (I) small pages, performance differences are negligible. When it comes to (II) large pages, it is noticeable that the average download speed was higher on the Windows Server system. Measurement results for (2) medium, (3) large, and (4) very large files clearly show that performance on the Windows Server system was better. There was a difference in average download speeds of (4) very large files for recurrence number groups A, B, C, and D, as well as for the download speed of (II) large pages, for all the same recurrence number groups. The most significant difference was observed during the transfer of (4) very large files. In this scenario, the performance was unequivocally better when the file was being downloaded from the Windows Server.

5. CONCLUSION

In this paper, we have described an experiment conducted to determine the impact of choosing an operating system on the performance of a web site hosted on the server. Both operating systems were installed with their default setup configurations without performing any post setup optimizations. Results of the experiment are measurements of average download speeds of the content of different types and sizes.

The types of content we used in the experiment were web pages and linked file resources. The four file size ranges are chosen to simulate typical text, multimedia, and audio-video content on modern web sites. The two pages are designed to represent pages of typical sizes, with common content, including text, referenced pictures, parsed links, rendered tables, lists, and so on. We have provided equal conditions for executing the web site on servers with two selected operating systems using the same hardware architecture, using the same local network, without actively connecting to the Internet for the duration of the experiment and deactivating all other software except those necessary for the web site. We used operating systems from the two largest families, GNU/Linux and Windows families. We used the server version of the popular distribution of GNU/Linux operating system - Ubuntu Server and the server version of Windows system - the Windows Server. For our experiment, we used the latest versions that were available at the time of conducting this research. Server software, database management system and web site development platform have been selected in the same way by analyzing market share data and overall representation of these technologies and we have not made any optimizations to these systems after their installation. The collection of measurements was automatic, performed using a script executed on a computer prepared exclusively for this experiment, and the data collected was subsequently imported, normalized, and processed. By analyzing the measurements, we have concluded that overall performance in terms of average content download speed from the web site is better when the web site is hosted on the Windows Server operating system compared to the one hosted on the Ubuntu Server distribution of the GNU/Linux operating system family.

6. FUTURE WORK

This experiment was intended to measure only the impact of a single factor of the platform configuration - the operating system. The research results have shown that choosing an operating system does have an impact on the web site's performance. For smaller contents, the performance is negligible. For larger contents it should not be ignored. Given that other features of the server platform configuration were not taken into account, and that their performance impacts were not considered in this research, it is conceded that additional research is needed to measure and show the impacts of other features of the server platform configuration, as well as to determine whether they influence each other in any way.

REFERENCES

[1] M. Poulos, N. Korfiatis and S. Papavlassopoulos, "Assessing Stationarity in Web Analytics: A study of Bounce Rates," in Assessing Stationarity in Web Analytics: A study of Bounce Rates. Expert Systems, 2019.

[2] L. Mathews, "AlternativeTo offers social software recommendations," 14 4 2009. [Online]. [Accessed 13 03 2020].

[3] A. Willson, "11 Reasons Your Website Can Have a High Bounce Rate," Search Engine Journal, 1 10 2019. [Online]. Available: searchenginejournal.com/website-bounce-rate/332439. [Accessed 15 03 2020].

[4] J. Little, "High Bounce Rate? Here are the Reasons & What You Should Do," 11 5 2020. [Online]. Available: crazyegg.com/blog/high-bounce-rate. [Accessed 10 7 2020].

[5] S. Woessner, "Increasing Revenue by Decreasing Bounce Rate," 18 6 2013. [Online]. Available: online-behavior.com/analytics/decreasing-bounce-rate. [Accessed 2 7 2020].

[6] J. Nielsen, "Conversion Rates," Nielsen Norman Group, 24 10 2013. [Online]. Available: nngroup.com/articles/conversion- rates. [Accessed 19 1 2020].

[7] A. Price, "Infographic: Web Performance Impacts Conversion Rates," LoadStorm, 9 4 2014. [Online]. Available: loadstorm.com/2014/04/infographic-web-performance- impacts-conversion-rates. [Accessed 20 5 2020].

[8] M. Beasley, Practical Web Analytics for User Experience: How Analytics Can Help You Understand Your Users, Newnes, 2013.

[9] S. Hill, "Operating system and hardware requirements, server sizing guide," Entuity, 4 2020. [Online]. Available: support.entuity.com/hc/en-us/articles/360006172733- Operating-system-and-hardware-requirements-server-sizing- guide. [Accessed 5 7 2020].

[10] GreenNet, "Understanding file sizes," 21 10 2015. [Online]. Available: greennet.org.uk/support/understanding-file-sizes. [Accessed 15 3 2019].

[11] S. Bauer, D. Clark and W. Lehr, "Understanding Broadband Speed Measurements," in TPRC 2010, 2010.

[12] W3Techs, "Usage statistics of operating systems for websites," 2019. [Online]. Available: w3techs.com/technologies/overview/operating_system. [Accessed 3 8 2019].

[13] W3Techs, "Usage statistics of Unix for websites," 2019. [Online]. Available: w3techs.com/technologies/details/os-unix. [Accessed 3 8 2019].

[14] W3Techs, "Usage statistics of Linux for websites," 2019. [Online]. Available: w3techs.com/technologies/details/os-linux. [Accessed 3 8 2019].

[15] W3Techs, "Usage statistics of web servers," 2019. [Online]. Available: w3techs.com/technologies/overview/web_server. [Accessed 3 8 2019].

[16] W3Techs, "Usage statistics of server-side programming languages for websites," 2019. [Online]. Available: w3techs.com/technologies/overview/programming_language. [Accessed 8 3 2019].

[17] Explore Group, "The Most Popular Databases 2019," 21 4 2019. [Online]. Available: explore-group.com/blog/the-most-popular- databases-2019/bp46. [Accessed 3 8 2019].

[18] W3Techs, "Usage statistics of content management systems," 2019. [Online]. Available: w3techs.com/technologies/overview/content_management. [Accessed 3 8 2019].