

CS 398 Research:

**Improving Content Management
System Performance and the Web**

GEOFFREY PYRZYNSKI

**Geoffresh Inc,
Tinley Park, IL 60487, USA
<http://www.geoffresh.com/contact>**

MITCHELL D. THEYS

**University of Illinois at Chicago,
Department of Computer Science (MC 152)
851 S. Morgan St., RM 941,
Chicago, IL 60607-7053, USA
mtheys@uic.edu**

Table Of Contents

1 Introduction	3
2 Motivation	3
<i>2.1 Attributes of a Well Designed Website</i>	<i>3</i>
<i>2.2 Page Rank</i>	<i>5</i>
<i>2.3 Search Engines and Page Rank</i>	<i>5</i>
<i>2.4 Meta Tags and Meta Descriptions</i>	<i>6</i>
<i>2.5 Internal and External Linking</i>	<i>7</i>
<i>2.6 Portability</i>	<i>8</i>
<i>2.7 CMS Versus Static Sites</i>	<i>9</i>
3 Improving the website	10
<i>3.1 XHTML and CSS Validation</i>	<i>10</i>
<i>3.2 CMS Configuration</i>	<i>12</i>
<i>3.3 Initial Benchmarks</i>	<i>13</i>
4 Experiments	14
<i>4.1 Benchmarking Trials</i>	<i>14</i>
<i>4.2 Average Benchmark Data</i>	<i>16</i>
<i>4.3 Private Server Tests</i>	<i>21</i>
<i>4.4 Final Server Tests</i>	<i>22</i>
5 Conclusion	25
6 Bibliography	26

1 Introduction

What makes a good website? While creating my own website, I have experimented with many designs to improve www.geoffresh.com in any way I could. Geoffresh.com originally started off as a collection of HTML pages that were very light, but in order to add new content, I would have to add a new page every time new content needed to be added to the site. This would slow down any new content from getting to the site. Since then, Geoffresh.com has made the transition to a Content Management System (CMS) called Joomla!. [Ope10] With a CMS, a single webmaster can more easily manage the site and add content. Also, the site can dynamically generate pages. However, with a CMS, the site has the chance to become very slow since the webmaster is no longer in control of the core code. CMS sites tend to get bogged down with many HTTP requests, validation errors, and unneeded components. For a website, speed is critical! Users who visit a slow website will become frustrated with it and not return and slow sites may not even show up in search results. The goal of this project is to speed up [geoffresh.com](http://www.geoffresh.com) and I believe that the three areas that most affect CMS performance are: valid XHTML and CSS, the number of HTTP requests, and the configuration of the CMS itself.

2 Motivation

2.1 Attributes of a Well Designed Website

If a webmaster was to go and design their own website, what characteristics must it show to be considered by many to be a well designed website? This goes well beyond just the template of the site but the template is what the visitors will see first. A website template is the look and feel of the website. It is common between all pages of the website. In a template, the designer can specify the color of

the site, the font size and type, and even where different material on the site will show up. The template is tasked with organizing the content on the site. To have a well designed template, the designer must make sure that the visitors can easily access the information that they are looking for. Also, for people who are using screen readers, it is important to put a well written description for any of the links and pictures on the website so when the screen reader “sees” the description, it can properly relay the information to the user. For example on an image, the description of it should tell what the image is instead of having it listed as “image0001”.

For websites, content is king! That is why users want to come to a website. They come for the information, data, or the entertainment the site is offering. Without content, users will only visit for a short period of time, leave, and possibly never return. Even worse, they may not even see your site. In order to have a site that visitors want to see, the website must offer them something they want. The content on the site can usually be described in a few key words called meta tags. Meta tags are words that appear in the page code of a website that tell the search engines what this page is about.

To make it easier for people to find specific content, the webmaster can put tags in the code of the page with commas separating each key word. Example meta tags for a video documentary on wolves would include but are not limited to: wolves, wolves habitat, types of wolves, wolves mating habits. These tags describe what the content is about. It would not be a good idea to just simply put “video about wolves” as a meta tag because it does not describe the details inside the video. Also you must anticipate about what the person would be searching for. If the visitor is

searching for information about the mating habits of wolves, they would be directed to your site because of the “wolves mating habits” tag.

2.2 Page Rank

Just because there are proper tags for the content of the site, still does not guarantee that potential visitors will be able to see it. When searching for information, users will use a search engine such as Google to type in their search terms. Google will then retrieve links to websites that contain the most valid information for their search terms and put them in a list with the most relevant one first. As it turns out, many users will not navigate farther than the first search page and sometimes not farther down that the first 3 sites that pop up! [Fai09] This is called page rank.

In order to increase page rank, a web designer must make sure that all of the pages have relevant tags on it. The website's pages must have something in common with each other or else they belong on separate website. For example, in geoffresh.com, all of the pages are related to the video productions of the crew of geoffresh.com. These pages do differ in content such as some have pictures of the crew while others have the video productions. Some pages have media on them while others have just text talking about the productions. Still all of the pages have a central theme.

2.3 Search Engines and Page Rank

It is important at this point to understand how search engines actually find the meta tags on a website. Search engines such as google use small programs called spiders that move from webpage to webpage and “the spider will begin with a popular site, indexing the words on its pages and following every link found within the site. In this way, the spidering system quickly begins to travel, spreading out across the most

widely used portions of the Web.” [Fra00] These spiders collect information on the webpage including all of the meta tags and documents them in its database.

A web developer should put meta tags that is part of the common theme on all pages, but have more specific tags as well on each page. In Geofffresh.com’s meta tag list, every page has the following tags: Imperial Hawk, Geofffresh, Geoffffresh, 'The Legend Of Imperial Hawk', jefffresh, jeff–fresh,jeffffresh. Notice how they have alternate spellings of each word as well as phrases. For each page, they have more specific tags that state more clearly what is on this exact page. The visitor and the search engine know that this website is dedicated to internet broadcasting, but with specific page meta tags, the visitor and search engine could see exactly what is on this specific page. For example, if someone searches for Imperial Hawk Act II, they will be directed to a specific page that contains the second act of The Legend Of Imperial Hawk because of the tags “ Imperial Hawk” which is on every page and the tag “ACT II” was specifically on the page containing the second act of The Legend Of Imperial Hawk.

2.4 Meta Tags and Meta Descriptions

Another tag that all websites should employ is the Meta Description tag.

This tag is similar to a slogan for the website. Usually it is one short sentence that describes exactly what the website is about. For Geofffresh.com, the meta description is “Geofffresh Imperial Hawk Productions Internet Broadcasting bringing you premium quality shows straight to your browser!” This tag tells the name of the website and what the website does. By including a meta description in a website’s head, search engines will use this to display directly underneath the link. [Fai09] This will give searchers a quick overview of a website and see if this is what they were looking for. If the title of the

website does not tell exactly what the website is about, then the meta description will pick up the slack as long as it is well written.

Since meta tags play an important role in helping boost page rank, web designers place as many tags into the webpage as possible in order for the web spider to associate the website with those tags. However, some web developers have abused this by including words that are not relevant to the site. Sometimes a web developer would simply copy a digital dictionary and place all of the words into the head of the site. This would successfully capture a searcher by making the search engine think that there was information relevant to the search on the page, but in reality there was not. In order to combat against that, search engines are now detecting if a website abuses the meta tags. If it does detect an abuse, the search engine will give it a page rank starting at 1,000 and climbing. Since most “users do not even look farther than the first 3 pages that come up in a search”, let alone past the first four line items, the site will not even be seen by searchers. [Fai09] So it is most effective if only the proper search terms are used to describe what is on the page.

2.5 Internal and External Linking

It is also important to know how visitors can come about to a website. If a website has many links to it from other websites, it will draw in more traffic. For example, if an article was written about a certain website, the reporter may include a link to that website in his or her report. When someone is reading that article, they may click on that link if they are interested in what the site has to offer. This has successfully caught another view who may have not seen the website. Mrs. Claiborne from searchengineguide.com commented on this saying “Links are like roads to your site. With every new link, more people can find your site. Just as more drive-by traffic helps a

physical business, the more links there are to your site, the more traffic you are likely to get.” [Cla05] and this is an important factor in visibility on the net.

There are three different ways a site can be linked. There may be an incoming link from another website, an outgoing link to another website, or it may go both ways. The last way is most likely a “link swap” where the webmasters of both sites decide to help each other out by linking to each other. Every link helps a website become more visible.

Of course, it also helps to have many internal links within a site. This helps users quickly move about the website. If a website does not have many internal links, visitors may find it difficult to move around and find the content they are looking for. For example, if an article was written on the home page of a website talking about a new video that was released on the site, it would be best to have a link to this video or to have it implanted directly into the article. This helps the visitor find the content they want quickly and easily.

2.6 Portability

Another attribute of a well designed website would be its portability. While a web designer may spend hours upon hours programming their website, it can all come crashing down if the server that it resides on fails. When building a website, it is important to create backups of it to save the current state of the site and be able to revert back to it if the site gets hacked. Also, by keeping the most current backups, the web designer will be able to move his or her site to another server in the event that the server becomes unstable or is no longer suitable for the website.

All of the above information is essential to making a good website, but now the standards have risen again. Now search engines such as google are having

their search algorithm check for the performance of the website. So now even if a website has the most relevant information to the search, it can be bumped down due to poor performance. Of course the opposite can be true as well. If a website has a high performance rating, it will be higher up on the search results. So now it is imperative for slow running sites to improve their performance or else they will be left behind in the search results as other sites, with less pertinent information, will be higher up on the search results.

2.7 CMS Versus Static Sites

Now, with geoffresh.com completely transformed into a CMS, it has experienced a severe slowdown in performance! With the new standards for a well designed website requires a website to have high performance, why would someone even transfer their site over if it caused a slow down? The big advantage of having a CMS site is the ease of managing the material on it. In fact, one webmaster could control an entire site and keep it updated. This is because the main core code is now controlled and edited by a third party; in this case Joomla! [Ope10] A webmaster no longer has to worry about designing the core system that his or her site must run on.

Also, with a content management system such as Joomla!, content on the site can be dynamically generated. This means that users can enter data into the site and can expect to see a different result from time to time. For example, if someone went to geoffresh.com and put in the search terms “bachelor’s grove” they would find the videos pertaining to the bachelor’s grove mini documentary. If articles were added to the site later and the user came back to search for “bachelor’s grove” they would find that there would be the video plus the articles that show up in their search results.

Dynamically generating content would also help the website grow much quicker when the webmasters are not the only ones adding content. When the community of visitors and users are able to add to the site, the site can grow and become more vibrant than a static site that only displays information. By allowing users to engage with the material on the site and interact with it, the site takes on a community feel to it. By converting to a CMS, a webmaster can easily allow users to create their own page on the site and interact with others. If the webmaster allows, users can even add their own content or comment on other's work.

3 Improving the website

3.1 XHTML and CSS Validation

An issue with Content Management Systems that directly correlates with its speed issue is the XHTML and CSS validation. Validation is important because there are standards that should be followed when adding code to a website. This has to deal with how browsers "see" the web page. When browsers access the web page, they have to do some "guessing" to place everything in the browser window correctly. In order to have the most accurate "guessing" a standard was developed by the World Wide Web Consortium (W3C) [W3c10] to make sure that web browsers and web developers were moving together in the correct direction instead of moving apart. For example in figure 3.1, with these lines of code there is a specific way to set page min and max width.

```
body.width_2 div#wrap {  
  min-width:760px;  
  max-width:1800px;  
  width:auto !important; /*ie6 hack*/  
  width:960px; /* ie6 hack*/  
}
```

Figure 3.1 Code Snippet for Setting Page Width

Notice the syntax of the code and how that the pixels are abbreviated as “px” and not written out as “pixels.” This is one example of the standard that was established by the W3C. In this example, there is a line with “/*ie6 hack*/.” If the web code is becoming standardized, then why is there a “hack” in there? As it turns out, the Microsoft Corporation has decided to not value to W3C’s standardization. In fact, Microsoft’s Internet Explorer browser has come under fire from the European market because they are “not complying with W3C web standards when it comes to how IE renders pages.” [Ban07] For their Internet Explorer browser to display web pages properly, web designers must place special codes or “hacks” into their CSS files so visitors who are connecting with Internet Explorer have the same experience as someone who is connecting with a browser that supports the W3C standards. Examples of browsers that support the W3C are Firefox, Safari, Google Chrome, and Opera.

The main validation issues I was having involved small spelling errors and incorrect notation. For the notation, it would be as small as placing in a colon where a semi-colon should have been to signify the end of a line. But while I was debugging my XHTML code, I noticed that my URLs were throwing errors after sections of the URL that had an ampersand. This is yet another example of a W3C standard. In order to properly code in an ampersand, the line of code must read “&” instead of just “&.” This is so the browser knows that the ampersand is a special character and not just part of a string.

How does the validation effect the performance of the site? Since a browser has to interpret how to display a page and then deliver the results to the user, it

will take longer if the browser has to “think” more about the page. If a site was written with a validation error, the error would cause the browser to try harder to interpret the code and figure out how to display the site. Having just one or two small validation errors will not have a massive impact on the performance of the site. However, a large amount of errors will severely slow down the site. The site may transfer the requested data to the visitor’s computer quickly, but it will take a while for the browser to make sense of it all and finally display the page to the user.

3.2 CMS Configuration

Geoffresh.com had many performance issues going into this research.

The site was first performance benched at 216 HTTP requests, a ping of 74.955 milliseconds with 12 hops, and CSS and XHTML errors that numbered into the hundreds. The first issue that had to be dealt with was the legacy system that was running in the background of the site. The legacy code was needed for the video component, Seyret, to operate because Seyret was not compatible with the new version of Joomla! that was installed. The Seyret video component was really only compatible with the original version of Joomla! 1.0. With Geoffresh.com operating on Joomla! 1.5, it was limited by the extra bulky and slow running code in the background. In January of 2010, Geoffresh.com made a major upgrade to its video system and removed the old Seyret component and replaced it with HWDVideoshare. This new component was designed to work in the native Joomla! 1.5 environment however it did contain many XHTML and CSS validation errors.

The next step was the replacement of the commenting system.

JComments was also running on the old legacy code and it required too many HTTP requests. Specifically, it was loading too many images for no reason. This component

was replaced with JComments and the new program was also Joomla! 1.5 native as well. JComments still does load many images but it was still less than that of JComments.

From there, I upgraded the photo system to the Joomla! 1.5 native version. With the upgrade came many XHTML and CSS validation fixes. The photo system is not a heavily used component; however, it does take more time to load all of the images on the page VS a page with just text. Upgrading this would still help the site because all of the components of the site would have to be on the native 1.5 code before removing the legacy code plugin to operate.

3.3 Initial Benchmarks

Figure 3.1 First HTTP Requests

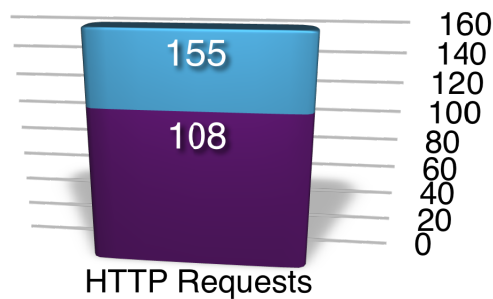
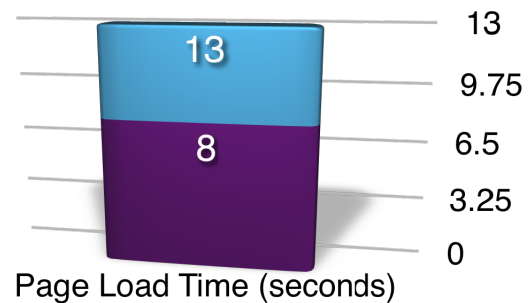


Figure 3.2 First Page Load Time



■ Benchmark 2

■ Benchmark 1

With all of these upgrades, I decided to do another benchmark on the site.

Just by upgrading the components to the Joomla! 1.5 Native code, I was able to reduce my http requests by 31%! With my page load times going down from an average of 13 seconds to 8 seconds, I was able to reduce my page load time by 39%! Although this was a major decrease in my http requests and load times, it was still not enough.

My next step was to get rid of the XHTML and CSS errors since it is difficult to determine exactly how they affecting the load time of the site. I decided to create my own custom template for Geoffresh.com because there were too many unnecessary images being loaded with the template I currently had. With my own template, I was having difficulties with the XHTML and CSS validation because I was unfamiliar with the XHTML and CSS standards. Many issues included spelling errors, incorrect or old syntax, and splitting the CSS files. I originally had the CSS of the site distributed between two files but in order to save an HTTP request, I combined both files into one.

4 Experiments

4.1 Benchmarking Trials

With the XHTML and CSS errors fixed, I did another benchmark on the site. This time I looked at several different categories. First and foremost, I was looking for how many HTTP requests the site was making. The number of HTTP requests was looking like a major performance factor from my last data trial, but this time I wanted to look at different categories as well. Specifically, I was looking at the time it took to download, the average bytes per second of the download, the download speed and the generation time. There were nine trials taken at one hour intervals.

Figure 4.1- Site Download Time in Seconds

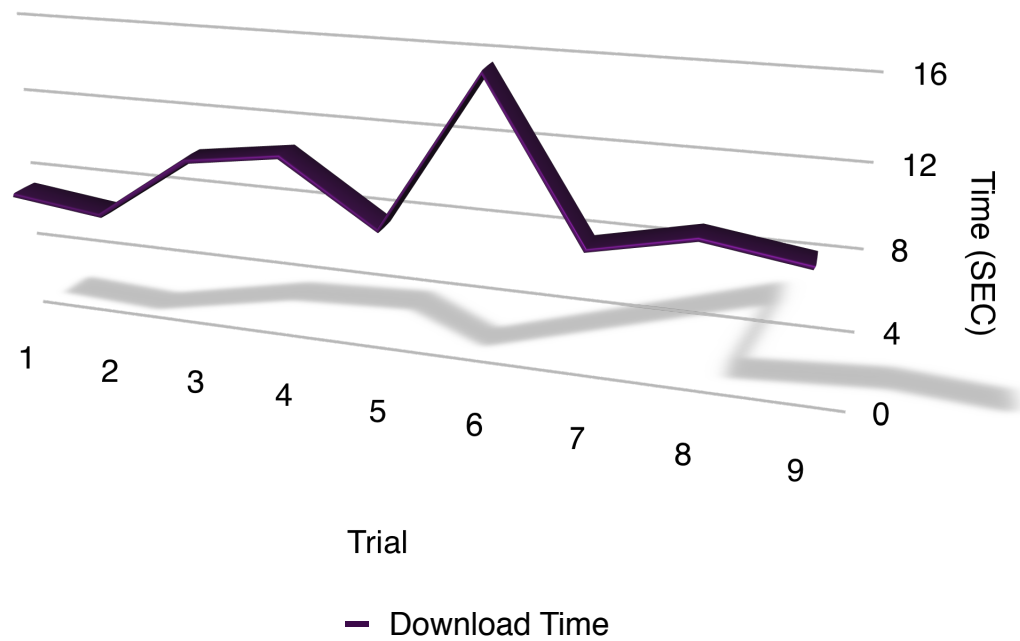


Figure 4.2- Site Download Time in Bytes per Second

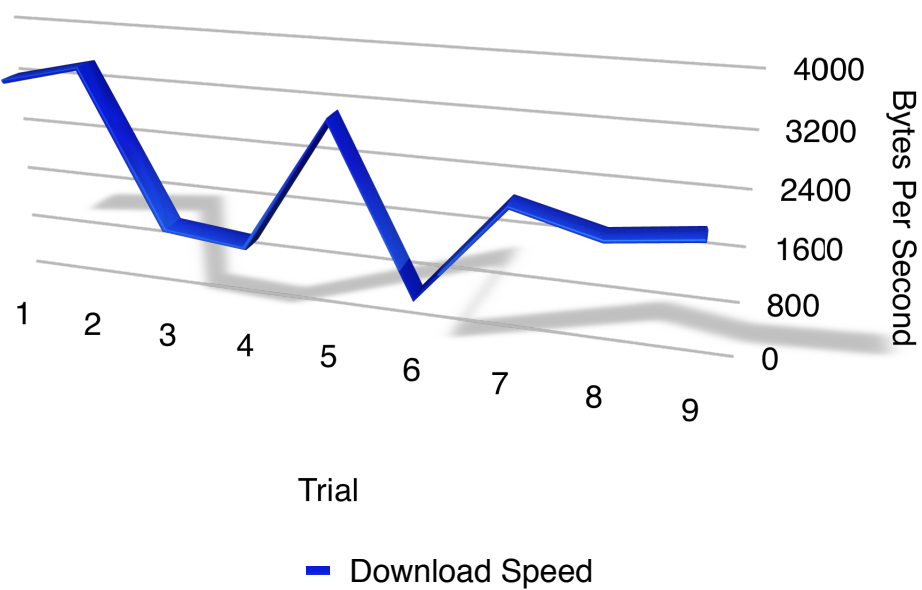
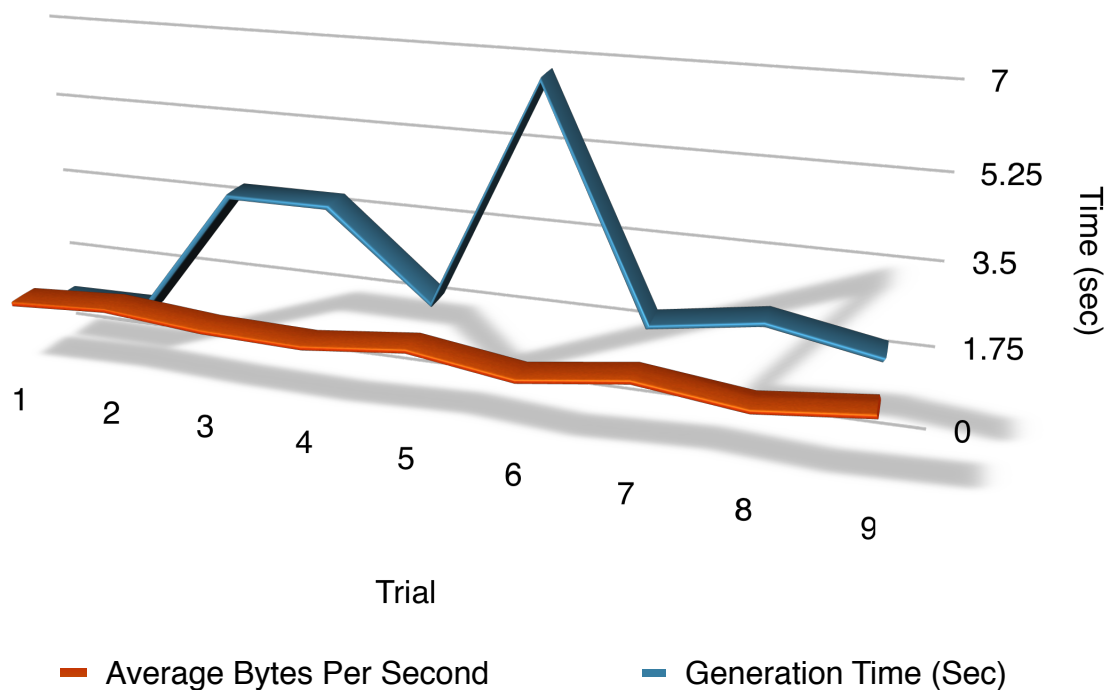


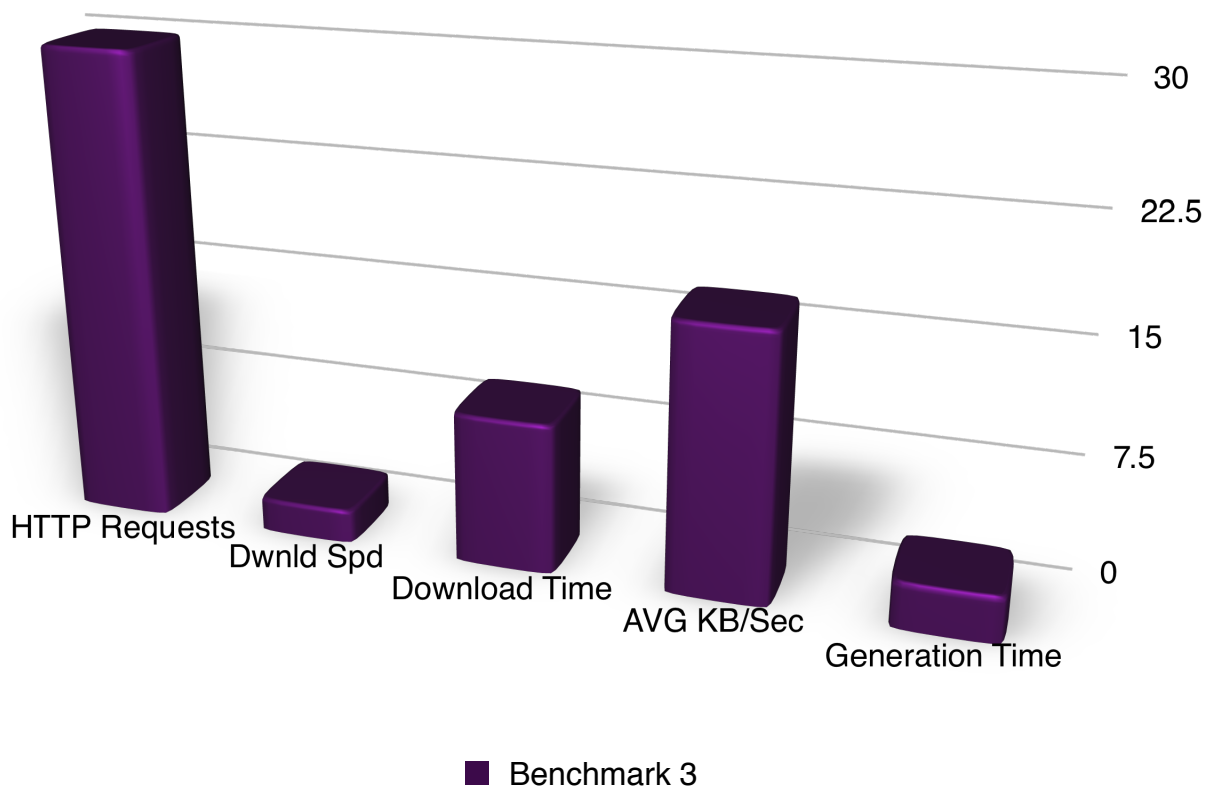
Figure 4.3- Site Generation Time and Average Bytes Per Second



4.2 Average Benchmark Data

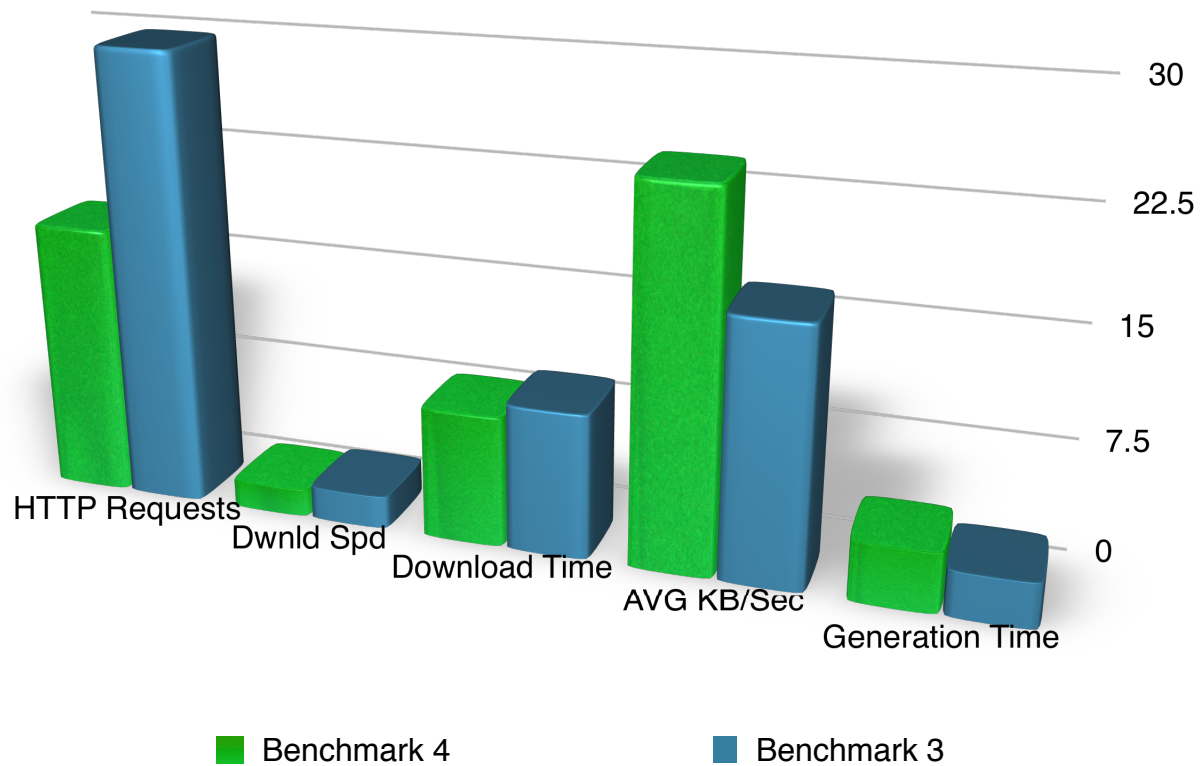
After reviewing this data, I found it difficult to tell what was going on with the site. So I decided to average all of the data trails in each category. This would allow me to see how the site was performing over all over the course of one day. See table 4.7 for exact values.

Figure 4.4- Benchmark 3 With 1.5 Native Code Averages



With this benchmark, I was able to see how geoffresh.com was performing over all. This was a good snapshot to see how things were going over an entire day. From here I was able to see that I was able to reduce my HTTP requests even more by removing some redundant CSS files and images from my custom template along with adding in drop down menus. I thought by having less HTTP requests, the site should generate faster because there were less items to download and place on the page. With that done, I took another benchmark of the site over a day. See table 4.7 for exact values.

Figure 4.5- Benchmark Four With Custom Template



After reviewing this data I found that my average page generation time had actually gone up! My HTTP requests had gone down by 12 requests and as expected my average download time had gone down. But the download speed from the server had gone down as well. I was expecting this value to stay very close to the same, but while reviewing the individual data trials, I found that the download speed varies by 77.37% at its greatest point! This was not something that I could control because the download speed has to deal with the connection from the server to the computer and the network in between. With those three different areas all working together, it was difficult to tell where the issue was residing. Was it the website's code that was slowing

it down or perhaps the server was bogged down or maybe the network was very busy at the time the test was taken? It is difficult to say at this point. So I decided to turn on the cache for my site to see if there was an increase in performance. See table 4.7 for exact values.

Figure 4.6 -Benchmark Five With Cache On

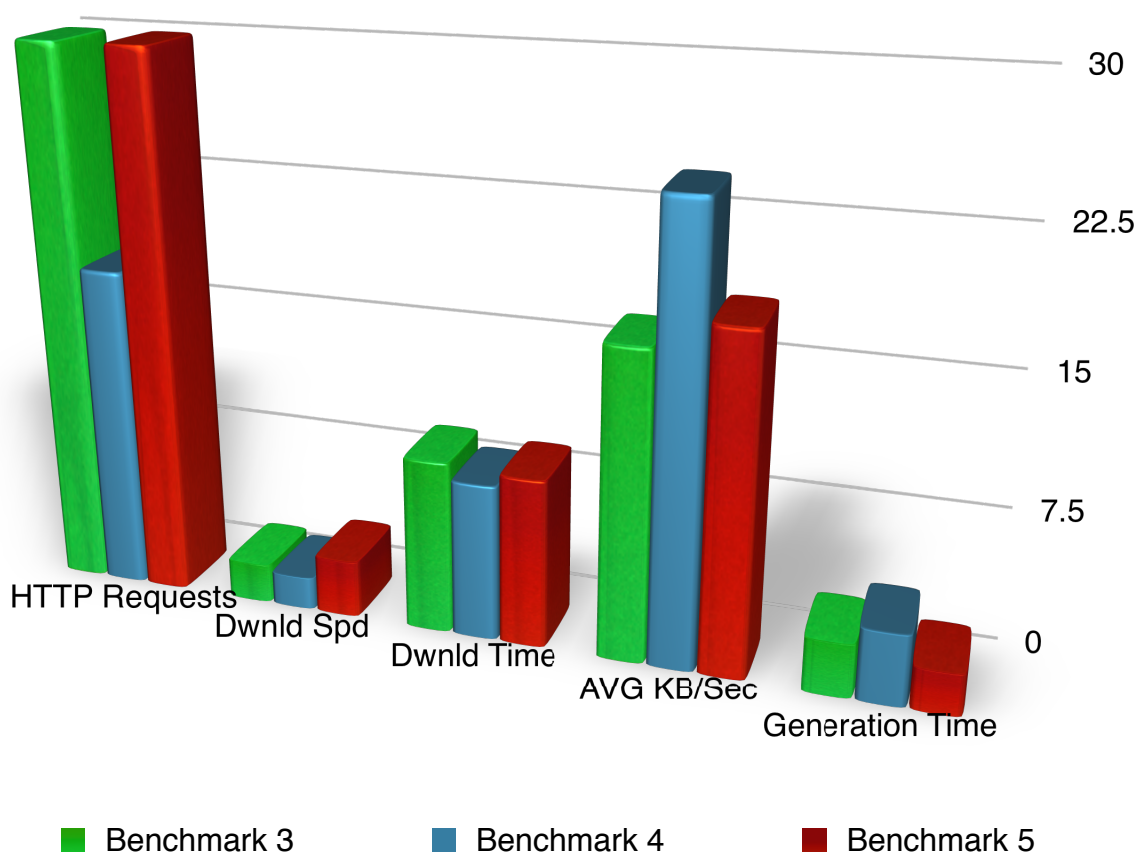


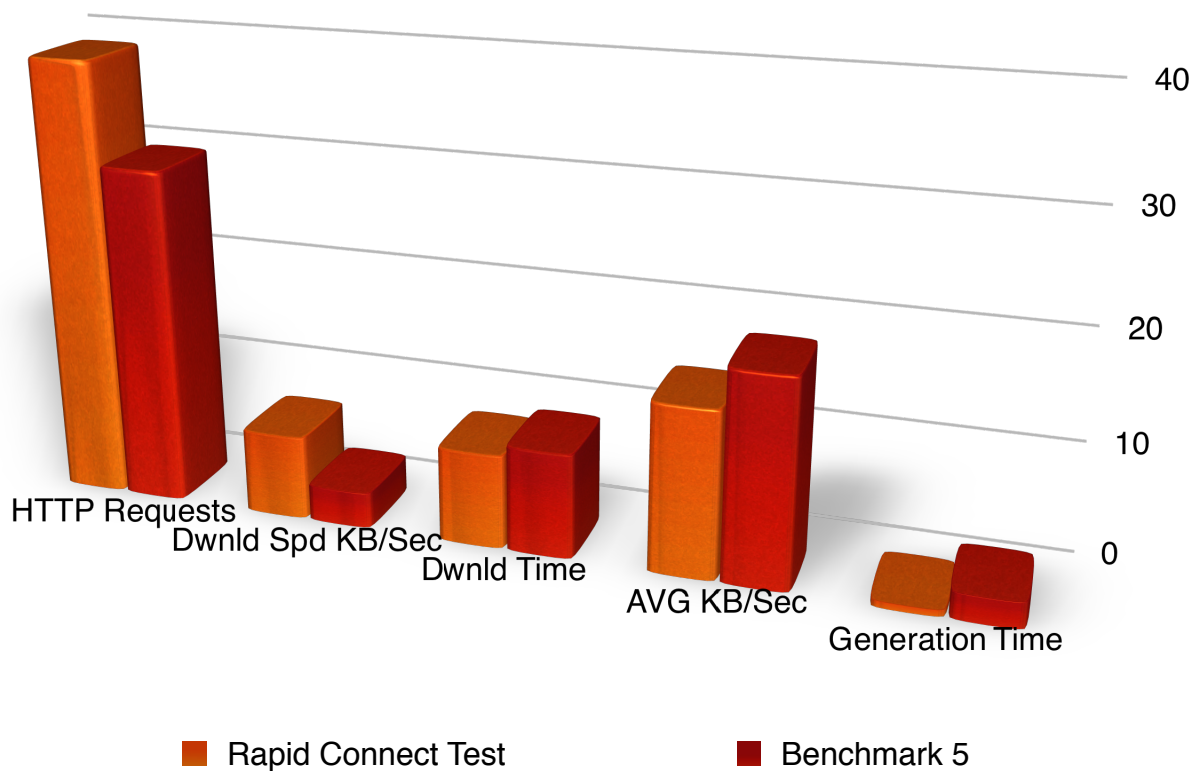
Table 4.7 - Benchmark Values

	HTTP Requests	Download Speed	Download Time	AVG KB/ Sec	Generation Time
Benchmark 3	30	2.163	9.61	17	2.97
Benchmark 4	18	1.803	8.67	24.6	3.94
Benchmark 5	30	3.168	9.34	18.5	2.3
Rapid Connect Test	39	7.496	8.39	14.9	0.56
Private Server Test	17	45.496	0.536	50.4	0.27
Scale	Number of Requests	Kilobytes per second	Seconds	Kilobytes per second	Seconds

The generation time had gone down by 22% with the cache on. Also, the download speed increased as well with an average of 3168 bytes per second along with the average bytes per second of each downloaded item to 1.85. The site was loading much more quickly now and users were able to add to the site much more quickly now.

Since all of the previous tests were taken once every hour, I decided to do a rapid connection test. I would test the server for nine trials with only 30 seconds in between each trial. This would show how the site was performing in a short period of time instead of the daily overall performance.

Figure 4.8- Rapid Connect Test VS Benchmark 5



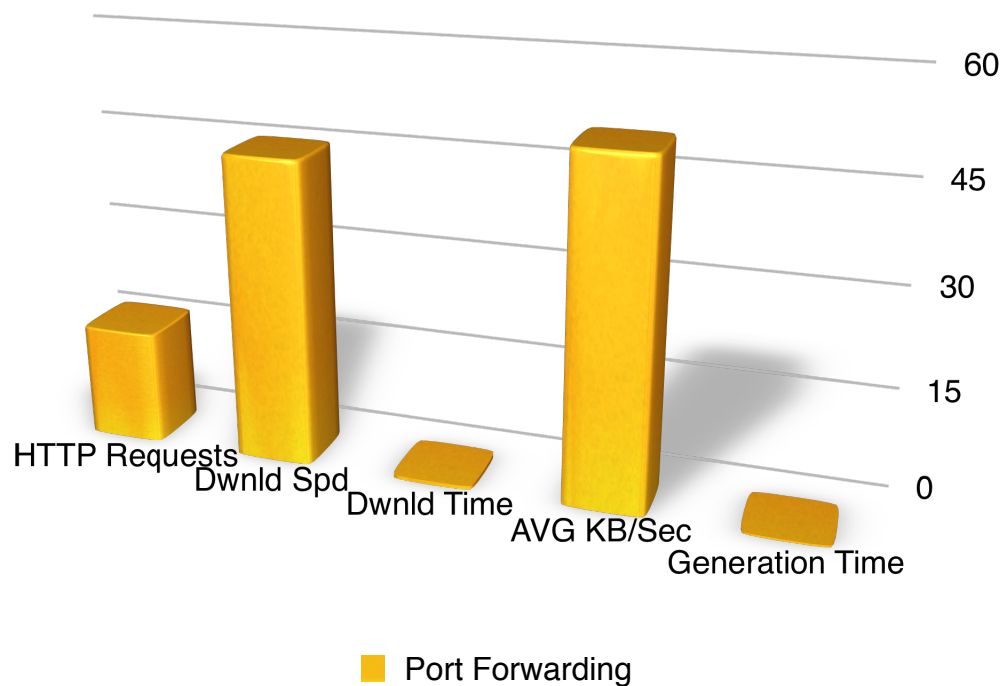
I was very surprised to find that the HTTP requests had actually gone up. After further investigation, it was due to the commenting system loading all of the different images with the page. The download speed still did variate greatly with a 61% difference between the highest and lowest download times. This could be causing a large performance difference between the trials. So with that in mind, I made an attempt to take the server out of the picture.

4.3 Private Server Tests

I was able to install a program called BitNami [Bit10] on my personal computer and I made the site look just like the live website. With this set up on my own machine, I would be able to see how the site would perform on a private server. Since

the live site is on a shared server, everyone on that server has to share processor and ram resources. By running the site off of my own laptop, I could see how it would perform with sole access to the ram and processor. I still used the same website to run a performance test on it but first I had to forward the ports from my machine to the router. See table 4.7 for exact values.

Figure 4.9- Port Forwarding (Simulated Private Server) Tests

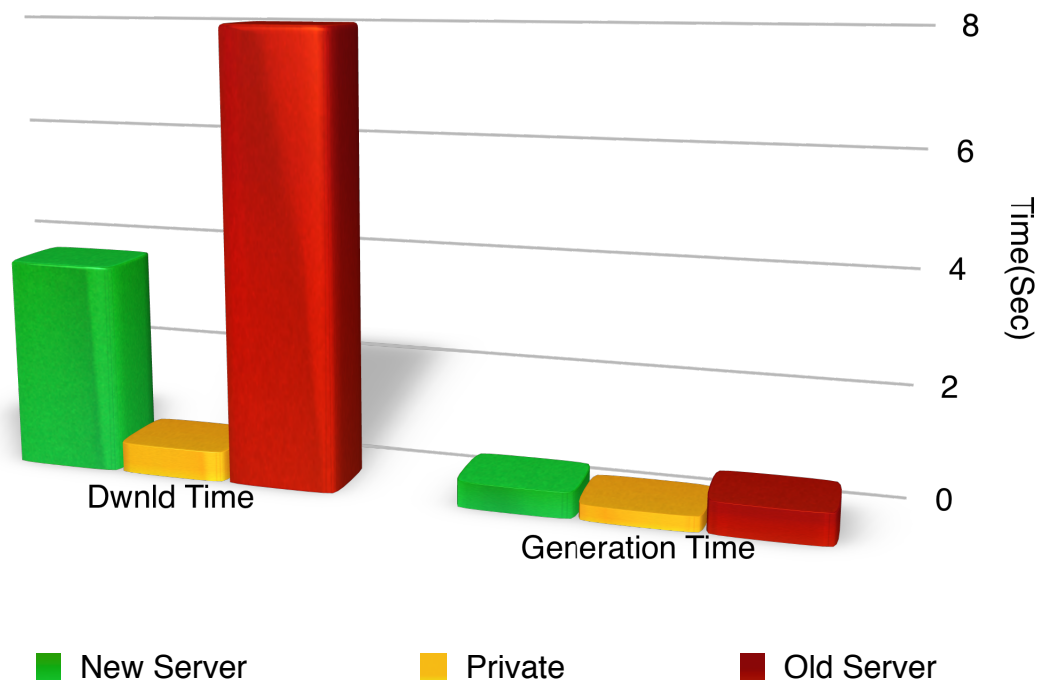


4.4 Final Server Tests

From figure 4.8, I was able to conclude that the site itself was operating at peak efficiency! The only major cause of site slowdown is now limited to server performance because I was still connecting to the performance testing site over a network. With this in mind, I started to look into upgrading the software on the server. Upon contacting my hosting company regarding my server issues, they had offered to move my site to a new

server that had the latest PHP installed on a 64 bit environment. When the site had completed its move to the new and upgraded server, I did another performance test on it.

Figure 4.10 - New Server VS Private VS Old Server



	Download Time Seconds	Generation Time Seconds
New	3.74	0.457
Private	0.536	0.27
Old	8.39	0.561

Figure 4.11 Average Kilobytes Per Second

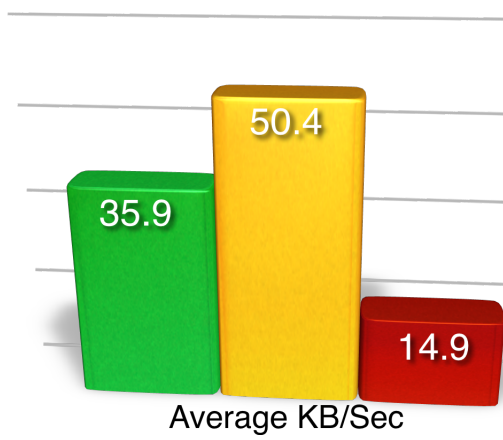
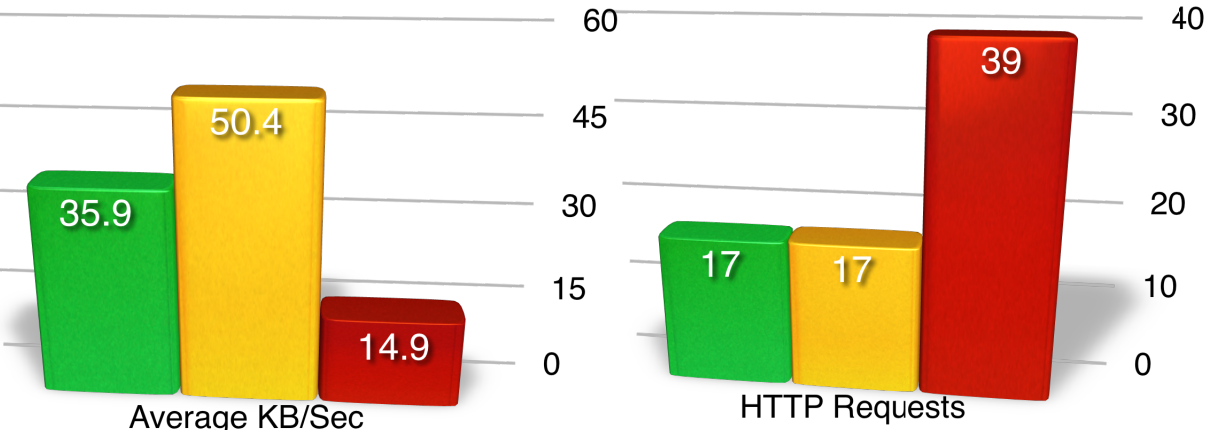


Figure 4.12- Number of HTTP Requests

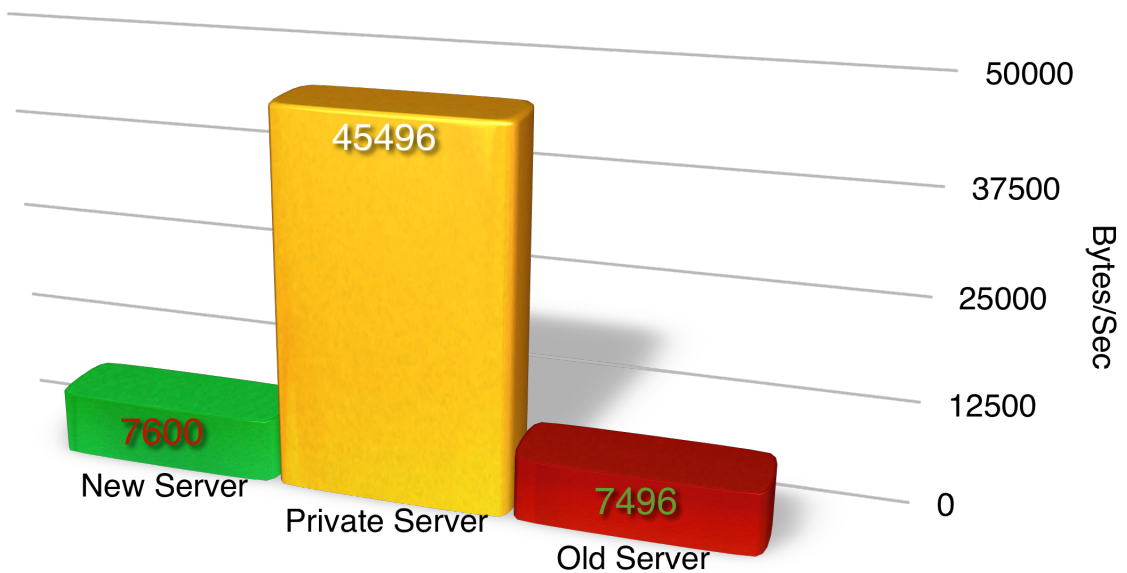


■ New Server

■ Private Server

■ Old Server

Figure 4.13 - Download Speeds



With this new server, the website is running very well! It now has the same amount of HTTP requests as the simulated private server and a download speed slightly greater to that of the best download speed recorded on the old server. The download time had also gone down from the old to the new server by 55% while the average bytes per second per downloaded item went up by an amazing 140%! The generation time had only gone down by 18% but the overall performance of the site has improved greatly and it now stays that way! Before the site had issues where it would perform poorly at times and at other times, it would perform very well. Now the site is performing very well and consistently too!

5 Conclusion

In conclusion, there are many factors that effect a website's performance. The webmaster must make sure that their programing is using only the files needed so they reduce HTTP requests to have less items downloaded and to make sure their code is following W3C validation standards. In addition to having lean, valid code, the webmaster must make sure to keep up with the proper software upgrades and server/database maintenance. With all of that in mind, the webmaster should not have any issues with performance and would then just have to concentrate on generating content and making sure that visitors can see their site by providing the proper meta tags and descriptions.

6 Bibliography

- [Ban07] Bangeman, Eric. "Opera tries to force IE into W3C compliance with EU complaint." 13 Dec. 2007. <<http://arstechnica.com/tech-policy/news/2007/12/opera-tries-to-force-ie-into-w3c-compliance-with-eu-complaint-firefoxs-success-may-work-against-it.ars>>
- [Bit10] Bitnami. 2010. <<http://bitnami.org/>>
- [Cla05] Claiborne, Scottie. "Threads of the Web - Linking for Success." 25 May 2005 <<http://www.searchengineguide.com/scottie-claiborne/threads-of-the-web-linking-for-success.php>>
- [Fai09] Fairley, John. "SEO Social Media Analytics & Paid Search." Drawing Web Traffic Workshop. 500 W. Madison, Chicago. 14 Dec. 2009.
- [Fra00] Franklin, Curt. "How Internet Search Engines Work." 27 Sept. 2000. <<http://www.howstuffworks.com/search-engine1.htm>>
- [Ope10] Open Source Matters, Inc. 2010. <<http://www.joomla.org/>>
- [W3c10] W3C. 2010. <<http://www.w3.org/>>