*IOSR Journal of Computer Engineering (IOSR-JCE)*
*e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 22, Issue 2, Ser. I (Mar - Apr 2020), PP 14-20*
*www.iosrjournals.org*

# Development to speed in Internet Utilizationby Web Cache Replacement Model

Angkhana Janyaadisai[1], Prapai Sridama[1],
Nutdanai Singkhleewon[1]&Noppwan Wisedsindา[2]

[1]*Faculty of Science and Technology,BansomdejChaoprayaRajabhat University, Bangkok, Thailand*
[2]*Faculty of Management Science ,DhonburiRajabhat University, Bangkok, Thailand*

***Abstract:*** *To study the Internet usage speed, develop a high frequency chosen first (HF-CF) algorithm as an alternative to web caching, and compare the efficiency of this proposed HF-CF algorithm to a Least Recently Used (LRU) algorithm are been objectives in this research.This research uses a sample of 500,000 usage data sets from a university client machine to teach the system and subsequently develop a HF-CF algorithm based on the HF-CF technique to predict future web traffics. The efficiency of this proposed is compared between HF-CF algorithm andLRU. This research uses a test data set of 50,000 data points. The statistical analysis is based on means, standard deviations and t-tests. Based on the origin server's ratio between all download and request times, the average download time without web caching between 9 AM and 12 PM of the test data set was 4.01 seconds. Under the HF-CF algorithm, the mean of the hit rate was 40.42 percent while the mean of byte hit rate was 32.81 percent. Under the LRU algorithm the mean of the hit rate was 19.15 percent while the mean of byte hit rate was 17.64 percent. Under the HF-CF algorithm, the mean of the download time was 2.61 seconds while it was 3.32 seconds under the LRU algorithm.*

***Keywords***: *Web cache, High frequency algorithm chosen first, Hit rate*

---

Date of Submission: 17-03-2020 Date of Acceptance: 02-04-2020

---

## I. Introduction

The internet is important for many activities such as the education, economic, and multimedia in the present. There are many data from internet network while people accessing internet around the world are high quickly. The two problems for using online network are the network congestion and web server overloading that people have spent time for accessing network[1]. However, many researches solve this problem. There are using many techniques for data management such as web caching researches. The concept of web caching management is recording the web pages, which are used highest from people [2]. This concept can solve the bottle neck of web servers and can decrease traffic of internet network.

Web caching is a web-based management system to reduce time to access and load various websites from the source server. There are 3 levels of web caching (client level, proxy level and original server level). The proxy server is the center between servers and client. The proxy server will be requested from clients and the proxy server will manage data for loading to clients. Then the caching of proxy servers is important. The well efficiency of data management can decrease the response time and decrease using network bandwidth. If the information is being retrieved from the client and the required information is copied in the proxy serverthen the proxy server will send a copy of the data contained in the cache to the client. But there is no data requested by the client in the proxy server then the proxy server requests to the source server and forwards to the client. In addition, it will be copied for future use when someone requests this same information again. According to research on web caching, it is found that the general algorithm of web caching can increase webpage discovery rates up to 40% - 50% [3].

Most corporate internet connections today use the same access point for using the internet andthe internet speed will be reduced if there are many users. The amount of internet usage of the organization is high and is likely to increase at a double rate. The cause is due to various services occurring, such as network video and audio services, playing game online, and using various applications. Therefore, this research aims to improve the speed of the internet by using the web cache replacement model with high frequency algorithm

chosen first (HF-CF). The method is to load web pages that are expected in the future to be store in the web cache in advance to alleviate web server bottlenecks, reduce network congestion, and increase internet speed.

## II. Theories and Related Researches

In the development of the speed of the internet by using the web cache replacement model, the researchers have researched documents, academic articles, and information on the internet for studying principles, concepts, theories, and related researches with the following topic.

2.1 Web caching

Web caching is the process of copying information that has been requested by a user to a caching server, expecting that data to be requested in the future[4]. There is a copy of that webpage near to the user, such as at the client or proxy server. Web caching can help reduce 3 problems, which are 1) reducing user response time, 2) reducing network bandwidth usage, and 3) reducing loading of web pages from the source server.

1) Types of web caching

Caching keeps a copy of a webpage close to the user, where 3 caches are found: 1) the client calls the browser cache, 2) the proxy server is called the proxy cache, and 3) the source server is called the cache server.

2) Web cache replacement

The cache memory will fully copy the webpage that has been run from the user. If a new webpage is used then the copy of the new webpage is stored.Therefore, a decision must be made to remove the old webpage from the cache and to copy a new webpage that has just been run.

3) Replacement by considering the time of request

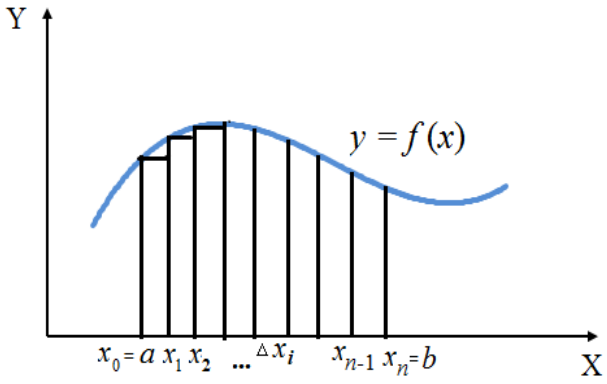
This method uses the modernization of data as a basis for replacing data. The least important webpageoccurs because the webpage is not being called for longer than other webpages. This webpage will be deleted when there is not enough memory cache to reproduce the data. This method is called the Least Recently Used: LRU). This method is high popular because LRU is high efficiency for many situations.

2.2 Mathematical and statistical methods

This research uses the following mathematical and statistical methods.

1) Finding the area under the curve

To find the area under the curve can be found by dividing the area under the curve into a small rectangle as much as possible. This method starts by dividing the closed range $[a,b]$ into n range. The dividing points are $x_0, x_1, x_2, ..., x_n$ , where $a = x_0 < x_1 < x_2 < ... < x_{n-1} < x_n = b$ If we divide close range of $[a,b]$ tosub close range of $\Delta x_i$ . The $\Delta x_i$ will be near zero. The area under the curve is $y = f(x)$ and itis above the X-axis. This example is shown at figure 1.




From figure 1, the A area is near add value of square area, which is under the curve. The equation 1 is shown below.

$$A = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i)\Delta x_i \qquad (1)$$

$$= \int_a^b f(x)dx$$

2) Random variable

Random variables are functions from sample spaces to real numbers. Random variable that have values in a countable set are called discrete random variables. A random variable that contains every value in a particular interval(or range) is called continuous random variable. Set A to be a discrete random variable. The probability function of A is $P(X \in A) = \int_A f(x)dx$. The expected value of random variable can use by $E(X)$ and we can solve $E(X)$ by using equation 2.

$$E(X) = \sum_{all\ x} xf(x) \tag{2}$$

Let $X$ is a continuous random variable. The density function of probability of x is $P(X \in A) = \int_A f(x)dx$. The expected value of random variable $X$ can write by $E(X)$ and we can solve by using equation 3.

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx \tag{3}$$

3) Technique for smoothing

Simple exponential smoothing is a forecasting method that is suitable for data that is stable and short-term forecasts. This method gives unequal weight or importance to each data. This method will give the most weight to the latest information and will reduce each other in an exponential way. The equation for forecasting can be show at equation 4.

$$\hat{Y}_{t+1} = \alpha Y_t + (1-\alpha)\hat{Y}_t \tag{4}$$

Let $\hat{Y}_{t+1}$ is forecasting value at t+1. $\hat{Y}_t$ is forecasting value at t. $Y_t$ is real value, which is made in t and $\alpha$ is weighted value of smooth adjustment.

4) Polynomial Regression Model

The polynomial regression model is a model with relatively simple function images. It is a function in the form of independent variable. This method is used if the interpretation of parameters in the model is not desired. This method requires regression to estimate or forecast. It can be used as a tool to check the different curves of regression between independent variables and dependent variables. The equation is shown below.

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + ... + \beta_p X_i^p + \varepsilon_i \ ; \ i = 1, 2, ..., n \ , \tag{5}$$

where $Y$ is dependent variable. $X$ is independent variable. $\beta_0$ is core cutting distance of $Y$. $\beta_1 - \beta_p$ is regression coefficient of number 1 to number p and $\varepsilon$ is random discrepancy.

5) Estimation of regression coefficient by least squares method

Ordinary least squares method is the estimation of the parameters that make the smallest squared error. From equation 5 and estimator $\hat{Y}_i = b_0 + b_1 X_i + b_2 X_i^2 + ... + b_p X_i^p \ ; \ i = 1, 2, ..., n$. We can find value of $b_0, b_1, ..., b_n$ that $\sum_{i=1}^{n} (Y_i - \hat{Y})^2$ has least value.

Let $Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$, $X = \begin{bmatrix} 1 & X_1 & X_1^2 \dots X_n^p \\ 1 & X_2 & X_2^2 \dots X_2^p \\ \vdots & \vdots & \vdots \ddots \vdots \\ 1 & X_n & X_n^2 \dots X_n^p \end{bmatrix}$ and $b = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{bmatrix}$.

We will get $b = (X^T X)^{-1} X^T Y$ .

6) Performance comparison

Performance measurement of cache memory replacement algorithm, performance measurement based on Hit Ratio (HR), data rate (Byte Hit Ration: BHR) found in cache, and average time to download data (Average Download Time: ADT). If the found rate and the rate of data size found in the cache are high then the web cache memory replacement algorithm is performing well.Because of when the encounter rate and the rate of data size found in the cache are high thus reducing the amount of internet congestion. Users can access data fastertherefore reducing the average download time (ADT). Equations can show below.

$$HR = \frac{\sum_{i=1}^{N} h_i}{N} \tag{6}$$

$$BHR = \frac{\sum_{i=1}^{N} b_i h_i}{\sum_{i=1}^{N} b_i} \tag{7}$$

$$ADT = \frac{\sum_{i=1}^{N} d_i (1 - h_i)}{N} \quad HR = \frac{\sum_{i=1}^{N} h_i}{N} \tag{8}$$

Let $N$ is then all numbers that request to use data. $h_i = \begin{cases} 1; \\ 0; \end{cases}$ is the data, which are used number and data are in the cache. $b_i$ is size of data, which request at number $i$. $d_i$ is the duration for downloading data, which request at number $i$.

**2.3 Related Researches**

Ibrahim and Xu studied the data retrieval method in the future. They found that the forecasting to request in the future by using content of web pages for linking HTML and using ANN for requesting in the future by the main word [5]. Other research, to studied about web cache replacement and advance information in client level for increasing efficiency in posting and getting data dynamic on network with parameters. The performance in this technique is the time of data access time [6]. A research studied about to increase the efficiency of the service on network system for decreasing the time for data access by intelligent web cache method and the pre-fetching data. This research found that LRU gave the average of hit rate at 68 percent, Neuro-fuzzy LRU gave the average of hit rate at 85 percent, and Bayesian-LRU gave the average of hit rate at 86.8 percent [7]. A study spent time for studying about the cache distribution by squid proxy server and LRU method. This study found that the hit rate and the byte hit rate of the cache distribution by increasing server for data backup was higher than LRU method [9]. In addition, many researches were interest to design new methods for increasing the efficiency of web caching and to compare these methods with LRU methods [8–10].

## III. Methodology

This research is a research and development aimed at improving the speed of the internet by using the web cache replacement model with a high frequency chosen first (HF-CF) algorithm.This research will load websites that have had a high frequency of launch in the past into a web cache memory. The data used in this research were websites that were used within a network of a university in Thailand between 9.00 a.m. and 12.00 a.m. from 1 December 2018 to 31 March 2019. The steps in researching are as follows.

**3.1) Data preparation**

The data in this research are information about websites usage between 9,00-12.00 a.m. between Monday and Friday for 30 days. The reason to choose to use this period is because of the high usage of internet. Before using data for research, data cleaning must be done to remove unwanted data. After that, this research will proceed as follows: 1) group the website data to select the 10 websites that are predicted to be the most used during the next period by considering the frequency of previous web browsing and single exponential smoothing technique. In addition, the research uses the frequency of website use in the past 6 weeks and forecast the frequency of website use of the following week of each website with a single exponential smoothing method. 2) The researcher enumerated the frequency of website use in every 1 minute for 30 days of all 10 websites. Next, to calculate the expected value of the number of times a website is used in each one minute period of time.

**3.2) Cache Feature Preparation**

For this step, 5 websites will be selected from 10 websites that are expected to be very popular in order to pre-load the webpage into the server cache. The method of selecting 5 websites was chosen every 10 minutes, using estimation method during the process of creating the 3rd polynomial regression.

**3.3) Cache Replacement**

Re-run the used website data sets every 10 minutes to test the found rate in the simulated web cache.

3.4) To compare efficiency between HF-CF and LRU

A comparison of the performance of the web cache replacement model using HF-CF algorithm and the web cache replacement model by the LRU by using the average of the rate of hit rate and the average of the size of data found in the web cache (Byte hit rate) is a variable in comparison. In addition, to compare internet speed efficiency by using average time to download data as a comparison variable.

## IV. Experimental Results

From the researchers developed the model to test the efficiency of HF-CF algorithm and used user data as input data to teach system (training set) and segment data at HF-CF not known to test the occurrence rate as shown in table 1.

**Table 1** Hit rate and byte hit rate of HF-CF

| Time | Number1 | Number 2 | Number3 | Number 4 | Number 5 | Hit Rate | Byte Hit Rate |
|---|---|---|---|---|---|---|---|
| 9.00-9.10 | youtube | ptvcdn | ytimg | atimemedia | kapook | 37.32% | 32.22% |
| 9.11-9.20 | youtube | ptvcdn | ytimg | atimemedia | fbcdn | 40.77% | 32.62% |
| 9.21-9.30 | youtube | ptvcdn | ytimg | atimemedia | kapook | 40.81% | 31.81% |
| 9.31-9.40 | youtube | ptvcdn | ytimg | atimemedia | kapook | 40.81% | 32.09% |
| 9.41-9.50 | youtube | ptvcdn | ytimg | atimemedia | fbcdn | 39.72% | 33.46% |
| 9.51-10.00 | youtube | ptvcdn | ytimg | atimemedia | fbcdn | 39.49% | 29.16% |
| 10.01-10.10 | youtube | ptvcdn | ytimg | atimemedia | fbcdn | 39.79% | 33.62% |
| 10.11-10.20 | youtube | ptvcdn | ytimg | atimemedia | fbcdn | 41.54% | 34.49% |
| 10.21-10.30 | youtube | ptvcdn | ytimg | atimemedia | kapook | 41.19% | 32.50% |
| 10.31-10.40 | youtube | ptvcdnt | ytimg | atimemedia | fbcdn | 40.27% | 33.98% |
| 10.41-10.50 | youtube | ptvcdn | ytimg | atimemedia | fbcdn | 40.97% | 32.85% |
| 10.51-11.00 | youtube | ptvcdn | ytimg | atimemedia | fbcdn | 38.60% | 29.46% |
| 11.01-11.10 | youtube | ptvcdn | ytimg | atimemedia | kapook | 39.78% | 35.68% |
| 11.11-11.20 | youtube | ptvcdn | ytimg | atimemedia | fbcdn | 41.98% | 33.41% |
| 11.21-11.30 | youtube | ptvcdn | ytimg | atimemedia | fbcdn | 42.53% | 36.15% |
| 11.31-11.40 | youtube | ptvcdn | ytimg | atimemedia | kapook | 39.71% | 32.30% |
| 11.41-11.50 | youtube | ptvcdn | ytimg | atimemedia | kapook | 40.78% | 34.18% |
| 11.51-12.00 | youtube | ptvcdn | ytimg | atimemedia | kapook | 41.42% | 30.66% |

From table 1, showing the site found rate values that the model has chosen for them by choosing from the points obtained from the area under the curve every 10 minutes. In addition, the average data found rate was 40.42 percent, the highest found rate was 42,53 percent, and the lowest found rate was 38.60 percent. Moreover, it was found that the average data size found was 32.81 percent, the highest data found rate was 36.15 percent, and the lowest data rate was 29.16 percent.

**Table 2** Hit rate and byte hit rate of LRU

| **Time** | **Hit Rate** | **Byte Hit Rate** | **Time 10.31-10.40** | **Hit Rate** | **Byte Hit Rate** |
|---|---|---|---|---|---|
| 9.00-9.10 | 27.29% | 31.71% | 10.31-10.40 | 7.40% | 5.88% |
| 9.11-9.20 | 28.29% | 32.19% | 10.41-10.50 | 8.37% | 6.32% |
| 9.21-9.30 | 9.62% | 5.18% | 10.51-11.00 | 6.79% | 4.41% |
| 9.31-9.40 | 33.71% | 28.22% | 11.01-11.10 | 8.42% | 6.84% |
| 9.41-9.50 | 29.04% | 31.22% | 11.11-11.20 | 15.94% | 13.57% |
| 9.51-10.00 | 13.39% | 9.79% | 11.21-11.30 | 40.58% | 35.86% |
| 10.01-10.10 | 6.95% | 6.24% | 11.31-11.40 | 33.57% | 28.92% |
| 10.11-10.20 | 14.26% | 11.11% | 11.41-12.50 | 5.52% | 6.39% |
| 10.21-10.30 | 25.14% | 29.55% | 11.51-12.00 | 30.49% | 24.18% |

From table 2 shows the hit rates and byte hit rates by LRU. The results show that it has an average of hit rate equal 19.15 percent. The highest of hit rate is 40.58 percent and the lowest of hit rate is 5.52 percent.

**Table 3** To compare average time (seconds) to download the data of the HF-CF algorithm and the LRU method

| เวลา | ADT (HF-CF) | ADT(LRU) | เวลา 10.31-10.40 | ADT(HF-CF) | ADT (LRU) |
|---|---|---|---|---|---|
| 9.00-9.10 | 2.86 2.662.66 | 3.17 | 10.31-10.40 | 3.01 | 4.00 |
| 9.11-9.20 | 2.66 | 2.94 | 10.41-10.50 | 2.69 | 3.75 |
| 9.21-9.30 | 2.47 | 3.56 | 10.51-11.00 | 2.54 | 3.68 |
| 9.31-9.40 | 2.29 | 2.52 | 11.01-11.10 | 2.87 | 4.04 |
| 9.41-9.50 | 2.70 | 2.82 | 11.11-11.20 | 2.57 | 3.51 |
| 9.51-10.00 | 2.57 | 3.56 | 11.21-11.30 | 2.62 | 2.68 |
| 10.01-10.10 | 2.70 | 3.74 | 11.31-11.40 | 2.67 | 2.91 |
| 10.11-10.20 | 2.43 | 3.50 | 11.41-12.50 | 2.57 | 3.84 |
| 10.21-10.30 | 2.47 | 2.86 | 11.51-12.00 | 2.28 | 2.59 |

From table3, the average download time of the HF-CF algorithm is 2.61 seconds. While the LRU gives the average download time equal 3.32 seconds. In addition, Moreover, every 10 minutes, the HF-CF algorithm takes less time to download the website than the LRU.

## V. Conclusions

The objective of this research is to improve the speed of the internet by using the web caching override model by a high frequency chosen first (HF-CF) algorithm. The loading of web pages that are expected to be requested in the future is stored in the web cache in advance, which can increase the rate of webpage encounter. This research has developed a model to test the efficiency of the encounter rate by comparing the HF-CF algorithm and LRU. The amount used as a training data sets is 500,000 data. The results showed that the HF-CF algorithm gave hit rate of 40.42 percent, byte hit rate average was 32.81 percent, and the average download time was 2.61 seconds.

## References


[1]. Manchanda and Gupta, N. (2013). Make Web Page Instant: By Integrating Web-Cache and Web-Prefetching. *Conference on Advances in Communication and Control Systems 2013 (CAC2S 2013*), pp.5-8.

[2]. Caceres, R. Douglis, Fred. Feldmann, A. Glass, G. and Rabinovich, M. (1998). *Web Proxy Caching:The Devil is in theDetails*. AT&T Labs–Research.

[3]. Anik, I. Arifa, A. and Abdul, H. M. (2017). HitBand: A Prefetching Model to Increase Hit Rate and Reduce Bandwidth Consumption.*I.J. Information Engineering and Electronic Business,* vol. 9(1), January 2017, pp. 36-46.

[4]. Ali, W. Shamsuddin, S. M. and Ismail, A. S. (2011). A Survey of Web Caching and Prefetching. *Int. J. Advance. Soft Comput. Appl*.,vol. 3(1), March 2011, pp. 1-27.

[5]. Ibrahim, T.I. & and Xu, C-Z. (2004). A keyword-based semantic prefetching approach in Internet newsservices. *IEEE Trans. Knowledge Management*, vol. 16(5), August 2004, pp. 601-611.

[6]. Balamash,A.Krunz, M. and Nain, P. (2007). Performance analysis of a client-side caching/prefetching system for web traffic. *Science Direct Trans. Computer Networks,* September 2007, vol. 51(13), pp. 3,673-3,692.

[7]. Vijilesh, V.and Nedunchezhian, R. (2016) Improving Access Latency by combining Intelligent Web caching and Web pre-fetching Techniques. *Journal of Convergence Information Technology(JCIT)*, June 2016, vol. 11(3), pp. 57-63.

[8]. Ghofir, A. and Ginanjar, R. (2017). Distributed Cache with Utilizing Squid Proxy Server and LRU Algorithm. *Indonesian Journal of Electrical Engineering and Computer Science,* August 2017, vol. 7(2), pp. 474-482.

[9]. Siriruangampai, S. and Nupairoj, N. (2014). Cost-based replacement policy for cloud storage. *The Tenth National Conference on Computing and Information Technology*, pp. 706- 712.

[10]. Sridama, P.Prakancharoen, S. and Porrawatpreyakorn, N. (2015). Web Usage Forecast With Adapted Least Recently Used Model. *ARPN Journal of Engineering and Applied Sciences*, August 2015, vol. 10(14), pp. 5719-5729.


Angkhana Janyaadisai,etal. "Development to speed in Internet Utilizationby Web Cache Replacement Model." *IOSR Journal of Computer Engineering (IOSR-JCE)*, 22.2 (2020), pp. 14-20.