# ✚IJESRT

## INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY
## PROXY BASED WEB CACHE CONSISTENCY FOR ENHANCING NETWORK EXECUTION

**Pushpendra Kumar\*, Ankur Goel**
M. Tech Scholar* Dept. of Computer Science Engineering, RCE, Roorkee, India
Asst. Prof., Dept. of Computer Science Engineering, RCE, Roorkee, India

## ABSTRACT
With the exponential growth of the Internet, the World Wide Web (WWW) has become the most widely used tool for access and dissemination of commercial educational, and news information on the internet. The success of the WWW is largely due to its cap ability of providing quick and easy accesses to a large variety of information form sites all over the world. The retrieval latency of the WWW is however sensitive to overload conditions of the network which agonizes web users. Caching data copies at client Sites (including web proxies) is regarded as a good technique for reducing both the network traffic and the server load, thus improving retrieval latency of web documents. Use of proxy caches in the World Wide Web is beneficial to the end user, network administrator, and server administrator since it reduces the amount of redundant traffic that circulates through the network. In addition, end users get quicker access to documents that are cached. However, the use of proxies introduces additional issues that need to be addressed.

The existing consistency protocols used in the web are proving to be insufficient to meet the growing needs of the Internet population. For example, too many messages sent over the network are due to caches guessing when their copy is inconsistent. Many decisions must be made when exploring World Wide Web coherency, such as whether to provide consistency at the proxy level (client pull) or to allow the sever to handle it (server push). What tradeoffs are inherent for each of these decisions? The relevant usage of any method strongly depends upon the conditions of the network (e.g., document types that are frequently requested or the state of the network load) and the resources available (e.g., disk space and type of cache available).

One goal of this dissertation is to study the characteristics of document retrieval and modification to determine their effect on proposed consistency mechanisms. A set of effective consistency policies is identified from the investigation. The main objective of this dissertation is to use these findings to design and implement a consistency algorithm that provide improved performance over the current mechanisms. Optimistically, we want an algorithm that provides strong consistency. However, we do not want to further degrade the network or cause undue burden on the server to gain this advantage. We propose a system based on the notion of soft-state and based on server push. In this system, the proxy would have some influence on what state information. Our results show that an average of 20% control message savings by limiting how much polling occurs with the current Web cache consistency mechanism, Adaptive Client Polling.

**KEYWORDS**: World Wide Web, Web Cache, Cache Consistency, Proxy Cache Server Invalidation, TTL (Time-to- Live) Client Polling

## 1. INTRODUCTION
The World Wide Web (WWW or, simply, the Web) has gained tremendous popularity due to the ease which a number of users can access information from all over the world. The increased population of users and the large size of files being transmitted have resulted in concerns for different types of users. Server administrators need to eliminate unnecessary traffic, thereby allowing more bandwidth for useful information. End users desire faster document retrieval. Hence caching is used.

**Web Caching:** Cache (memory) is memory that is stored very close to the CPU, say on the same chip as the CPU, to allow fast access. Similarly, a disk cache is memory that is used to store frequently accessed disk pages

for fast access. Web caching is the storage of web objects near the user to allow. Fast access, thus improving the user experience of the Web surfer.

To improve network performance, caches can be placed in various locations in the network: where requests originate (client or browser caches), within the network (proxy or network caches), and at servers (server caches). Of the three, proxy caches are the only ones that lessen storage requirements for client machines, reduce network traffic, and promote of documents between cooperating caches within a network.

Proxy caches decrease the number of messages that enter the network by satisfying requests before they reach the server. In addition, they act as intermediate agents to allow multiple clients to quickly access a group of popular Web pages. In such an approach, requests for documents are redirected to a proxy cache rather than being serviced directly by a server. When document are from a server, copies are kept in proxy caches throughout the Web. Without caching, clients must send request packets directly to a specified server and the server then respond with packets. This message exchange limits the potential growth of the Internet. As the number of client requests and server responses increase, performance in degraded.

1. There are several advantages of using web caching.
2. Web caching reduces bandwidth consumption which leads to network traffic and network congestion.
3. Web caching reduces access latency for following two reasons. First, frequently accessed documents are fetched from nearby proxy caches instead of remote servers: therefore, the transmission delay is minimized. Secondly, because of the reduction in network traffic, those documents not cached also can be retrieved faster than without caching due to less congestion along the path and less workload at the server.
4. Web caching reduces the workload of the remote Web server by disseminating documents among the proxy caches within a network.
5. If the remote server is not available due to a server "crash" or network partitioning, the client can obtain a cached copy at the proxy.
   A side effect of Web caching is that it provides a chance to analyze an organization's usage patterns.

However, web cache also has some disadvantages, which is showed as following: A major disadvantage is the return of stale documents to the client due to the lack of proper proxy updating. This issue is the focus of this research. The access latency may increase in the case of a cache miss due to the extra proxy processing. A single proxy may become a bottleneck. A limit has to be set for the number of clients a proxy    can serve. A single proxy is a single point of failure.

## 2.  MATERIALS AND METHODS
The current cache consistency scheme used on the Web, client polling, is bandwidth inefficient. Too many messages are due to caches guessing when their copy is inconsistent. If many requests that are sent to the server are for documents that are actually not modified, then bandwidth could have been saved by simply serving the request from a cache in closer proximity to the requesting client. To address this problem, server invalidation with leases techniques have been proposed in the literature. The algorithms assume that it is possible to estimate the time that documents will change, but it is known that Web traffic and access patterns change frequently. Therefore, the leases should be dynamically adjusted based upon the frequency of writes as well as the request rate to capitalize on the benefits of a lease invalidation scheme. Objects that change frequently, but are seldom requested, do not need to be invalidated often. Also, objects that are requested often, but are seldom changed, may benefit from a lease technique. These conditions lead to a need to determine the threshold of when client polling, invalidation, or both should be employed. Due to the server overhead and, in some cases, unnecessary traffic generated by object invalidation, there is minimal acceptance of strict server invalidation with lease techniques. However, it is shown that server invalidation with leases provides some benefit over client polling in conditions where the server makes many writes to popular objects and the tolerance for stale data is low.

The overall intent of our research is to define criteria for adapting Web consistency protocols to request and modification frequency and to specific conditions of the internet. This includes understanding how popular documents in the Web change, determining their request and write history, and identifying the network problems that can occur when employing existing mechanisms of coherency. A set of effective consistency policies is then identified from the investigation. We make recommendations based on how frequently documents change and are accessed. The main objective of our work is to use our findings to design and

implement a consistency algorithm that provides improved performance over the current mechanisms presented in the literature. We develop an algorithm that adapts based on requests and updates to documents. This is a server-dominated approach that changes to a proxy-dominated approach, introducing the notion of Split-Proxy-Cache (SPC). The approach to this research is twofold: to evaluate Web cache consistency in its current state and, then, to develop an improved approach to Web cache consistency based on the findings about the existing access pattern and network effect of the existing infrastructure.

This research provides possible alternatives to the cache consistency problem introduced in the Web when caching is employed and users are concerned about the response rate and bandwidth issues that arise due to the current approaches. By reducing the stress on the existing network, decreasing document latency, and creating a more efficient and reliable Internet environment for receiving Web documents, Web caches with an efficient cache consistency framework can significantly improve while minimizing network and server costs.

The objectives of the proposed research are the following.
1. Identify which performance measures are most useful for evaluating consistency of cached documents.
2. Evaluate the feasibility of server-based invalidation as compared to client based consistency.
3. Develop a new algorithm for consistency that results in low staleness and minimal cost to the proxy, network, and server.
4. Compare the new algorithm to the existing proposed algorithms.
5. Overall, this research provides alternatives to improve network bandwidth, reduce latency, and decrease server load while maintaining a consistent View of documents on the Web.
6. We examine each of the objectives in further detail below.

### Performance Metrics
This observational steps widely used individuals early experiments and then the capabilities overview about the modern algorithmic program contain note matter, note byte matter, deal with note matter, data file move matter, unwarranted revise matter, old cost, duration of triteness, along with remote computer status overhead. These kinds of steps happen to be concluded in line with the vital properties considered necessary to this benefits and drawbacks about feel methods.

Emails: Overall amount of communications dealt in order to gain feel for the duration of time. This is usually a matter of communications associated with a total deal concerning the placeholder along with remote computer which includes deal with communications along with data file transfers.

Bytes: Overall amount of bytes dealt in order to gain feel for the duration of time. This particular clearly shows irrespective of whether data file volume has a bearing on what amount visitors are brought in as a consequence of consistency.

Regulate Emails: Lots of deal with communications directed as a consequence of consistency. Regulate communications are classified as the situation communications vital to inform this placeholder about all changes. Some people have the Get-If-Modified-Since require directed by your placeholder and then the Definitely not Adjusted answer directed by your server.

Computer file Airport transfers: Lots of data file transactions transport to take care of consistency. Computer file transactions develop the main degree of targeted traffic, mainly for big files. This particular metric simply is liable for numerous communications that may contain this particular data.

Unwanted Posts: Lots of unwarranted communications transport to take care of consistency. This is usually a matter on the demands which will produce a Not-Modified response.

Triteness: Amount about papers demands produced by this person which will produce a old backup simply being came back in the user.

Duration of Triteness: General hours if the placeholder things usually aren't similar to the server. This particular metric comes about weight training reclaim time frame material for placeholder records of activity along with identify typical entire length about old documents.

Equipment Think Viewgraph: Level of dining room table open area (number about proxies along with documents) important to identify in the event that to inform customers that the point comes with changed.

Statistics can be actually maintained in the procurator and also server. Around the equipment, we have to deal with exactly how much spot can be employed each one procurator and in what ways enough time the particular equipment will keep which will information. Correspondence relating to the procurator as well as equipment is actually finish this. The particular equipment has to notify the particular procurator exactly how much spot it can be helped as well as for a way enough time, and also procurator has to notify the particular equipment precisely what data files to help keep using the restraints given.

Make it possible for pre-existing feeble regularity algorithms mustn't happen further general performance penalties. People end up with placed brand new lintel area which are forgotten with the modern equipment in case the Placeholder employs the present feeble regularity algorithm. Also, an already-established equipment disregards almost any brand new area employed an exciting new proxy. For this reason, most of us are yet to placed almost any'further general performance fees to get the present feeble regularity algorithms utilised within the World wide web today.

Make it possible for the proxies utilize the modern nodes as well as the modern proxies utilize the servers. It ought to be attainable to be able to incrementally use the modern system, i.e., it must be half-witted compatible. So the present proxies utilize the modern palette, you'll find mail messages (implied and explicit) which will notify the particular equipment not likely to keep your details for this proxy. This new equipment handles details and refrains right from keeping up express details in accordance with mail messages the fact that procurator sends. This guidelines is dependant on HTTP/ 1.1, so that to get user-defined exts contained in the protocol's inquire as well as resolution header. Head exts are employed make it easy for proxies to be able to at the start inquire as well as rekindle leases on a server. Hire inquire as well as resolution locations will be contained in the usual HTTP inquire as well as resolution messages. The particular reserve refilling as well as invalidation tickets will be mailed because HTTP inquire lintel extensions.

$$E\ (T) = \alpha\ (tst - tst\text{-}1) + (1 - \alpha)\ E\ (t\text{-}1)$$

Where (tst - tst-1) represents the last observed time between two consecutive significant updates. The value of $\alpha$ determines how much weight is given to the last observed value and how much to the previous estimates. The smaller the value of $\alpha$, the more weight is given to the older values in the series. Larger values of a result in more weight being given to more recent observations.

Choosing the right value for $\alpha$ requires some amount of knowledge of the usage patterns on the series. This implies that the value of $\alpha$ should be chosen dynamically using the observed time series characteristics. Every time the server estimates the next time to update, it uses the current time series of update timestamps and simulates the estimation on that time series for different Values of a. It then chooses that value of $\alpha$ that results in the least error. This mechanism ensures that the value of $\alpha$ adjusts to the usage pattern of the Current iteration of lease assignments.

## System Derivation
### Polling-Every-Read
Polling-Every-Read can be induced when a consumer demands some sort of record might be In the event Modified-Since coping with the Grab request. Numerous look and feel be in charge of information that will result can be K = 2R, where exactly 3rd thererrrs r can be the numerous requests. In the event hardly any adjustments developed, K = 2R are the superior destined with the numerous careless be in charge of information that will occur. Subsequently, the numerous further demands that will take place in the expression can be very low by simply the numerous register relocate information generating from a come up with possesses occurred. Numerous register geneva chamonix transfers can be F ree p = 3rd thererrrs rwatts.F ree p can be delimited on top of by simply F ree p = M, if perhaps there are many is currently writing compared to visits generated. F ree p = 3rd thererrrs r if perhaps further visits compared to is currently writing occur.

### Time-To-Live

Your Time-To-Live look and feel process can be induced when a memory cache situations out there it's duplicate of an document. It relates to any memory cache noticing than a cached duplicate can be moth-eaten together with the TTL cost for that document. At one time a replica is resolute for being moth-eaten its fetched of your equipment for a resulting request that will document. The top destined for the numerous be in charge of information, K, can be the numerous timeout time intervals in the trying with traffic. The fact that consequence usually is determined by even if there's a browse in each interval. Subsequently, K = T/big t, where exactly Testosterone levels can be time of this trying with site visitors and additionally big t is a timeout value. Should the period distance, Testosterone levels, can be a lot less than the majority of instance amongst visits, now the numerous demands that will bring in a note in the equipment can be 3rd thererrrs rbig t The cost connotes the most important demands that will happen from a timeout. Subsequently, K = 2Rbig t.Numerous register geneva chamonix transfers can be F ree p = M, where exactly M, can be the numerous is currently writing from a timeout, yet we simply count number the most important chance of these write.

### Invalidation

People check out the replace invalidation strategies in that analysis. In the invalidation procedure, all the device is normally initialized any time data files adjustments along the forum (due with a write). It requires all the forum notifying all the proxies if you have a cached piece happens to be stale. Should you make it easy for And announce the quantity of proxies to make contact with in the instance of a great invalidation and even N announce the quantity of publish customizations, therefore the quantity of invalidation communications is normally NW.For ones replace invalidation tactic, F ree p = 3rd thererrrs rm = A, that will reduce the quantity of regulate communications to assist you to D = And *W.The viewgraph involved with passing along all the document together with the invalidation happens simply because we're passing along all the document to proxies that need invalidation regardless if your next require was produced from all the procurator as well as not. Plus, we're passing along changes in order to those proxies for any publish whether really are wanted as well as not.

## 3. RESULTS AND DISCUSSION

The actual formula which will decides on papers pertaining to eradication is termed the Avaricious Twin Sizing Worldwide recognition Place algorithm. Papers can be put into the actual caches to make certain the most famous papers select invalidation. All of us utilize the GDSP renewal formula to make this happen purpose connected with recording preferred papers into your SR-Cache and fewer preferred components in the actual G-Cache. Avaricious Twin Sizing Worldwide recognition (GDSP) emanates from the actual Avaricious Twin Sizing (GDS) formula, which have the actual wherewithal to squeeze long-term admission wavelengths connected with net documents. Connection occurrence is certainly integrated into the actual GDS formula by way of minor to moderate application worth pertaining to the object. The actual reckoning in the application feature from the GDSP Protocol is

$$u(p) = f(p) * c(p)/s(p)$$

Where s(p) is the size of p, c(p) is the retrieval cost (miss penalty), and f(p) is the access frequency. The u( p) calculation represents the cost saved per byte of p as a result of all accesses to it in a given period of time. The pseudo-code for the GDSP Algorithm is:

GDSP Algorithm

L = 0.0

H(p)=0

For each request for object p do

    if p is in cache then

        H(p)=L+f(p)*c(p)/s(p)

    else

        while there is not enough free space for (p)

L = min{H(q) I q is in cache}

Evict q which satisfies H(q)=L

fetch p

$H(p) = L + f(p) * c(p)/s(p)$

### Validation and Verification

Within confirming that procedure on the reliability algorithms, we must validate the algorithms engage in just as expected. Wealth attraction calls for supply rule debugging, executing of assorted screening, and also video or graphic check involved with the end result to make certain that that envisioned metrics was calculated. Every these technique execution was approved through stages. Each and every cycle possibly engaged execution involved with pieces of software to make sure you speed up practise or possibly by using established applications to ensure operations.

Typically the placeholder and also forum was precise implementations, in no way simulations, and also needed even more video or graphic check plus handbook executing as compared with additional components. We certainly have proven that answers labeled earlier by a few actives and also tested that end result for these activities.

### 4. CONCLUSION

The sheer numbers of invalidations provides the invalidations for the purpose your rent appeared to be reasonable thus contributes to a new group message. That relies on your rent duration. Host Invalidation (INV) necessitates the greatest volume of invalidations. When the rent extent cuts down, the numerous invalidations falls. Full volume of management emails for brand new criteria is shown to turn out to be cheaper than intended for equipment invalidation.

Host Invalidation exclusively defines practically nonexistent triteness if the Revision minute rates are lesser compared to demand rate. SINC defines exclusively 10% more desirable triteness compared to real Invalidation if the revision oftenness and also demand oftenness seem to be comparable. Together with the prediction that there is typically a lot less update versions compared to asks for in your Cyberspace,

Purchaser Polling by using Adaptative TTL won't provide tough consistency. Seeing that revision oftenness cuts down with 50 percent of, Purchaser Polling by using Adaptative TTL gives you 20% more desirable consistency. Your particular study content over head is regular if the revision oftenness differences and also is there to normal at any rate 5% lesser compared to invalidation content over head when exercising on Host invalidation and brand new algorithm. Inside uses while persistence wants seem to be significantly less rigid, your Purchaser Polling by using Adaptative TTL criteria is a wonderful choice.

Your TTL appeal can help determine your arrive at with affirmation demand emails for the server. Your triteness with cached reports also increases because TTL prices increase. If the TTL appeal obtains big enough, both of those Invalidation and also brand new criteria attain lesser triteness rates.

Your invalidation content over head relies on your revision rate. In the event the revision minute rates are very low in accordance with your demand charge, then your invalidation content over head will also be lesser compared to demand content overhead.

### REFERENCES

[1] B. Krishnamurthy and C. E. Wills. Piggyback server invalidation for proxy cache coherency. In Proceeding of the Seventh International World Wide Web Conference, volume 3.
[2] J. Yin L. Alvish, M. Dahlin, and C. Lin. Using leases to support server-driven consistency in large scale systems. In Proceeding of the 18 th International Conference on Distributed Computing Systems, May 2007.
[3] Ninan, P. Kulkarni, P. Shenoy, K. Ramamritham, and R. Tewari. Scalable consistency maintenance in content distribution networks using cooperative leases. IEEE Transaction on Knowledge and Data Engineering.

[4] P. Cao and C. Liu. Maintaining strong cache consistency in the World Wide Web.

[5] H. Yu, L. Brelau, and S. Schenker. A scalable web cache consistency architecture. In Proceeding of the ACM SIGCOMM '98Conference on Application, Technologies, Architectures and Protocols for Computer Communication.

[6] K. Worrel. Invalidation in large scale network object cache. Master's thesis, University of Colorado, Boulder

[7] J. Gwertzman and M. Seltzer. World-wide web cache consistency, USENIX Symposium on Internet working Technology and System.

[8] D. Li, P. Cao, and M. Dahlin. Wcip : Web cache invalidation protocol. IEFT Internet Draft, November 2005.

[9] V. Duvvuri, P. Shenoy, and R. Tewari. Adaptive leases: A strong consistency mechanism for the World Wide Web.

[10] F. Douglis, A. Feldmann, B. Krishnamurthy, and J. Mogul. Rate of change and, other metrics: A live study of the World Wide Web. USENIX Symposium on InterneWorking Technologies and Systems, pages 147-158, December 2005.

[11] C. E. Wills and M. Mikhailov. Toward a better understanding of web resources and server responses for improved caching. pages l231 { l243, 2009.

[12] R. Sundaresan, M. Lauria, T. Kure, S. Parthasarathy, and J. Saltz. Adaptive polling of grid resource monitors using a slacker coherence model. ln Proceedings of the 12th IEEE International Symposium on High Performance Distributed Computing (HPDC'03), pages 260-269, June 2003.

[13] B. Davidson. A survey of proxy cache evaluation techniques. in Proceedings of the Fourth International WWW Caching Workshop, March 2007. '

[14] R. Wooster and M. Abrams. Proxy caching that estimates page load delay. in Proceedings of the 6th International World-wide Web Conference, pages 325-334, Boston.

[15] S. Jin and A. Bestavros. Popularity-aware greedy dual-size web proxy caching algorithm. In Proceedings of the 20th International Conference on Distributed Computing Systems.

**CITE AN ARTICLE**