Customer-oriented Policy for Proxy Management System

Pattarasinee Bhattarakosol
Department of Mathematics,
Faculty of Science,
Chulalongkorn University, Thailand.
bhattara@sc.chula.ac.th

Wijak Srisujjalertwaja
Department of Computer Science,
Faculty of Science,
Chiang Mai University, Thailand.
wss@cs.science.cmu.ac.th

Abstract- *A technology that has developed dually with the Internet is the proxy technology. Its main advantages are reducing network bandwidth usage, increasing availability of content providers, and reducing users' response time. However, the implementation of the proxy still cannot serve well for a large number of customers since the ability of services is relied on the physical size of proxy itself. Customers might have to wait for a long time before receiving their requests or have to resubmit the requests repeatedly according that the requested information cannot be found in the cache. This paper is focused in the problem of increasing the hit rate of the proxy without using the very large volume of proxy cache, a transparent policy-based proxy system is proposed. The Customer-oriented policy is considered as a strategy to increase the object hit rate without high investment on the physical equipment.*

Keywords: Proxy, Web Caching.

1. Introduction

It is the fact that the Cyber world is a part of human's lives that we cannot deny. Additionally the learning system is adjusting itself using the popular information technology – the world of webs. Therefore, the web connection system can be called as the vein of information. The important issue that every Internet Service Providers (ISP) is concerned is the quality of service to their customer. One significant factor that can be measured the quality of services of any ISP is the Internet speed that a customer obtains for each connection.

Generally, the good quality of services is the result of a good network management policy. S.H. and F.H. Bakry [1] had proposed a technique called "comprehensive information profile" where variety of network profiles will be applied to setup the management policy in order to gain the better network. The other technique that was proposed by Hemmen [2] is the list of five models used for network organisation and management by considering all important key factors for successful network management. While some researches are focused on the organizational models [1][2], another research groups have considered the technical factors such as Hassan [3] who proposed the bandwidth management system so that the users' satisfactions can be obtained, and D.W-K & C.S. Hong [4] that proposed a quality of service management framework which is applied to an ATM network. Although many solutions are proposed, problems of quality of services are still incompletely solved. Bhattarakosol et al. [5] indicated that each user group has different traffic requirement. One of the proposed solutions is to install separated proxy server for each user group which can reduce the users' time and core network bandwidth.

In this paper, we will consider the situation of an organization that has to support various Internet web requested, such as ISP functions. One normal solution for manage web browsing processes is to implement a proxy system. Most organizations usually implement a shared proxy for any local requests. Unlike other organizations, ISP has to support various groups of customers, and customers hardly called the same web sites in the same/small interval of time. Thus implementing a share proxy for customers is not a good solution under this circumstance.

Section 2 is the major problems that will be taken into the consideration of improving the proxy service of the organization like ISP. Section 3 is the proposed solution for organizations that need to serve various types of requests from various customers. Section 4 is the system evaluation where the advantages and disadvantages of the proposed policy are described in three dimensions: the delay of the services cost of maintenance, and the scalability of the services that ISPs can provide. Section 5 and 6 are discussion and conclusion, respectively. For the future works are elaborated in Section 7.

2. Problems

Since the ISP has to services various types of customers so that their proxy system must be implemented using a high performance computer.

Moreover, the system administrators must be high skill persons so that they can maintain the system effectively. In order to maintain a good quality of the proxy services, the proxy software must be able to manage general, large, and complicated requirements from the ISP's customers. Therefore, the efficient proxy system may not be a freeware, and may be costly. Thus, the expenses that the ISP must spend on these requirements are high, for all expenses in hardware, software and peopleware.

Considering a high traffic from a large number of requests, the proxy system may not be able to guarantee the services to every customer. Thus, some customers must be waiting for their requests is the proxy queue for minuets, or have to resubmit their requests to the proxy from time to time. In this situation, it indicates that the availability of the system for customers does not exist for those customers, and the hit rate may be low.

According to the proxy's function, if a request URL's content does not available in the proxy's cache, the Internet retrieval process will be started. Therefore, if there is a large traffic sending out from the ISP, the large amount of Internet usages must be paid. Therefore, this paper will consider a new proxy architecture that solves the problem of high requests of customers, and high expenses of ISP that mentioned previously.

3. Customer-Oriented Policy

One standard service that any ISPs provide to their customers is the ability to browse webs around the world. Since every customer must register to the ISP, so this registration information such as user's age, occupation, and user's interest should be applied to identify the customer's group. Therefore, the user's profile is the important information to classify the customer's group. However, the correctness to classify a user is to consider the profile of browsing webs.

Although the interested areas are identified by the user's registration information at first, some users might browse web differently from what they registered. Since the users' interests can be related to many factors such as age, occupation, hobby, study, etc. Therefore, the registration information cannot correctly identify the right group for each user.

Since every proxy system will have its logged file to keep information of each web browsing. Therefore, the ISP can use the logged information to identify the characteristics of each user easily and correctly. The main reason to use the information from the logged file to identify the right group for a user is that every user will have their main interested. For example, teenagers normally browse webs for games, and entertainments, whereas the researchers or academia use web for information searching.

Considering the search engines functions, we can see that every search engine will classify the applications in groups such as mail, computers, educations, movies, etc. These techniques can be applied to the proxy implementation policy that ISP should implement proxy systems for each group of applications that available over the search engines, or implement proxy based on the standard functions which are educations, entertainments, technical solutions, news and miscellaneous.

This paper proposes a transparent policy-based proxy system as shown in Figure 1. Transparent proxy caching is a solution to the browser configuration problem. The policy of the system is to distribute user requests to the corresponding proxy server, which categorized into group of interest. The objective of the proposed system is to increase object hit rate. Setting up the proper groups of interest will increase the hit rate, which improves response time and saves wide-area bandwidth. The system is scalability by adding more proper groups of interest.

The proxy system is located on an ISP site. It consists of a proxy manager and group of proxy servers. The proxy manager matches users to a corresponding proxy server by using user profiles that users originally registered their groups of interest to the ISP. Each proxy server responds on a group of users according to their groups of interest. Suppose the ISP has five groups of interest: educations, entertainments, technical solutions, news and miscellaneous, then five proxy servers are established to support each groups of interest.

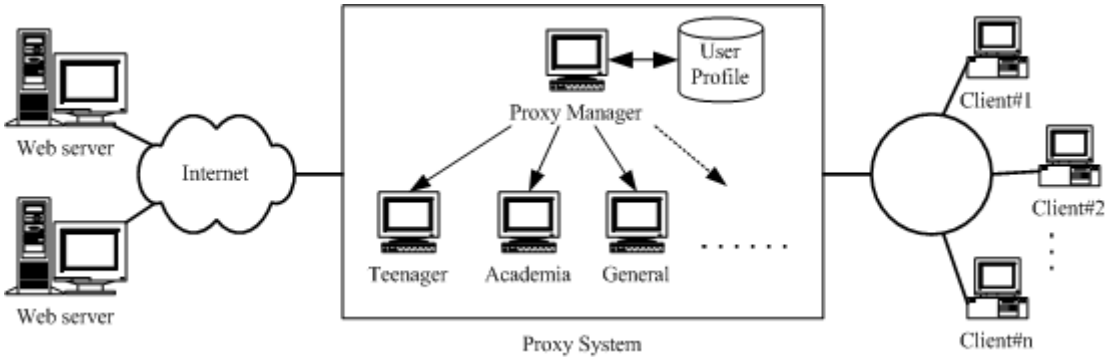

Figure 1. Transparent policy-based proxy system

System workflow is shown in Figure 2. The proxy manager intercepts a user request, then classifies the user into group of interest, and finally pushes the user request to a corresponding proxy server. After that, the assigned proxy server handles the user request. It searches the requested object on its cache and returns the object to the user. If it is not available, then the server will get the requested object from the request Web server, then, not only return that object to the user, but also keep the object in its cache.

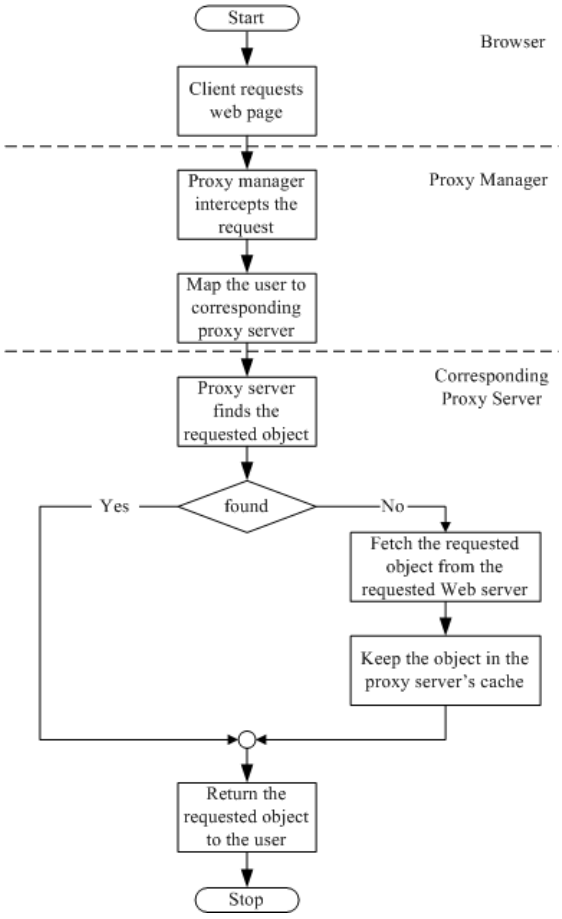

Figure 2. System workflow

Considering the responsibility of the proxy manager in Figure 2, it is similar to the function of the DNS where the internet destination is identified before sending to the Internet router, which is similar to the individual proxy system. Therefore, the system configuration of the proxy manager should be a high performance computer system that can manage all users' requests simultaneously.

According that the main function of the proxy manager is to identify the request into the classified groups, identified by the ISP, therefore the proxy manager needs to implement a URL database system (UDS) where all URLs are stored. Once a user requests a URL, the requested URL will be searched in the UDS to identify the service and the right proxy system will be assigned.

Referring to the proxy architecture, a logged file is assigned to store all accesses to the proxy. Thus, applying the same function to the proxy manager, the logged file of proxy manager stores information of each request such as user login name, requested URL, class of the URL, and time of the usage. These information are used to classify users based on their requests. So, the class of users can be identified correctly.

Based on the user classification above, the ISP can offer a new service policy for any users that can identify on only one class where the users' requests can be managed directly to the identified proxy server without passing the control of the proxy manager. This service is called as *customer-oriented policy*. The customer-oriented policy is the policy that a customer is assigned as a member of a particular proxy service only. Figure 3 presents the proposed architecture of the customer-oriented policy.

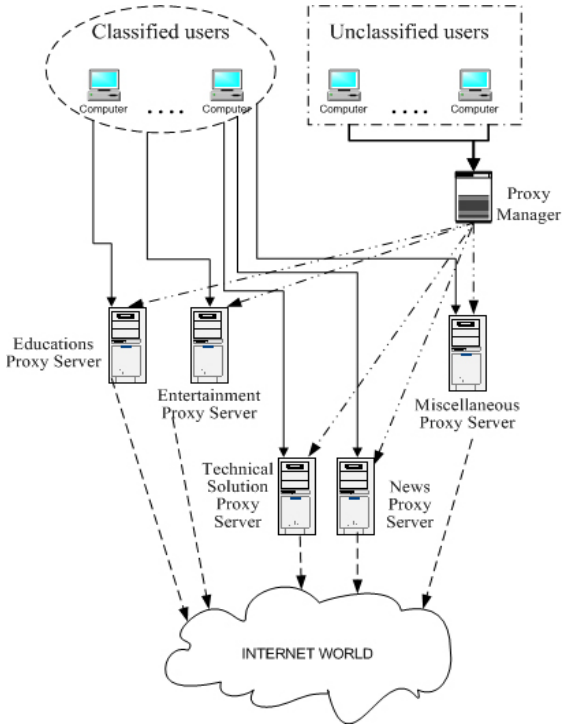

Figure 3. Customer-oriented Architecture

From Figure 3 all transactions from the unclassified users are sent to the proxy manager while transactions from classified customers are sent directly to the suitable proxy servers. In this circumstance, the classified customers will obtain required information faster than the unclassified customers. On the other hand, the unclassified customers have chances to retrieve various types of information.

4. System Evaluation

The proposed architecture will be evaluated in 3 aspects: reducing of the delay time, scalability of services, and cost effectiveness. The significant outcome of this architecture is that the quality of services of the ISP can be maintained without increasing cost of operations.

4.1 Reducing of the Delay Time

According to Figure 3, the proxy manager is the main system that responsible to classify users' needs (of unclassified users group) and then switches the requests to the suitable proxy system. We can see that although the request from unclassified customers must pass the classification process of the proxy manager, but the number of requests is smaller than the total number of customers that ISP usually services. This is because some customers are already assigned to the particular servers. Therefore, the requests of the assigned groups are not longer managed by the proxy manager. Thus, time to provide services to the undefined customers will not be too large, and definitely be smaller than the time to retrieve information directly from Internet.

Since each proxy is assigned to a particular service area such as entertainment, the probability that the requested data is stored in the cache is higher than using only one proxy to handle many classes of requests, the HIT rate is increased and the MISS rate is decreased. Therefore, the waiting time of users can be small because the information may be retrieved from the proxy cache. Thus, the delay time of each user is shortened.

On the other hand, if the required information cannot be found in the cache, the Internet retrieval process begins. Since we had shown that the MISS rate is decreased, then the number of Internet retrieval from the proposed system is smaller than the current system. Thus, the available bandwidth for Internet retrieval will be larger than the current system. Thus, the speed of Internet retrieval is increased. Therefore, the user's waiting time remains small.

Since there are some customers that belong to the certain classified groups. Thus, their requests will directly send to the required servers without passing the classification routine of the proxy manager. Then applying the same concept of searching for information from the cache of unclassified group, these classified groups will obtain their information quicker than the unclassified persons because their requests do not waste the overhead for proxy classification. Consequently, the quality of service for these classified groups can be guaranteed.

4.2 Scalability of Services

As the fact that the ISP business must expand their target groups to increase their profits while the quality of services must be maintained and the expenses must be in controlled. Since the current services are implemented using single proxy to serve various types of requests. Therefore, when the requests are increased, the size of the proxy might have to be increased. Thus, the ISP has to make a new investment on the new-enlarged proxy system otherwise the quality of services cannot be maintained.

Considering the objective of each proxy assigned in the proposed model, each proxy's cache will store the information of the same area of interest. Thus, every user who is interested in the same area can connect to the assigned proxy server, with or without passing through the proxy manager. Therefore, when the number of users who are interested in the same area is expanding, the assigned proxy is still able to serve these needs because the required domain remains the same. This indicates that the proposed architecture support the scalability of the ISP's customer without changing or expanding the size of the proxy server.

Although the proposed architecture still needs a high performance computer system, but only one system is needed for performing the request classification, and the other proxy systems that work under the control of the proxy manager can be in middle range computers that the cost of each one is low, and the maintaining system can be managed easily and quickly. Therefore, the expense for hiring the system administrator can be reduced because the system is not complicated.

4.3 Cost Effectiveness

As a result of increasing the HIT rate, the MISS rate automatically reduced. Therefore, the amount of traffics sending out to Internet from the ISP will also be decreased. Consequently, the expense of Internet usage that the ISP has to pay will be remitted, while the customers' satisfaction can be maintained and expanding.

The other advantage of this proposed architecture is that when a proxy is down, or disconnected from the network, other proxy still be able to serve other users. Therefore, the availability of the service exists, and the quality of services can be achieved.

The unavailable circumstance can be considered as the cost effectiveness of this new architecture because the customers still are able to obtain their information although a proxy is down. In the business competition world, if a customer cannot link to the requested web site often because the proxy is down, the probability that the customer

will change to the new ISP is large. Therefore, the ISP who lost their customers is also lost their benefits while the business expenses are still the same. So, this new architecture can obviously guarantee the services' availability to the ISP customers.

5. Discussion

In this paper we proposed a new proxy architecture system for ISP so that the customers can obtain their requests is a short time period. However, this architecture will work well in the situation that the ISP can classify their customers into classes. The main idea that is presented for customer classification is to perform the customers' usages profiles analysis, or allows users to choose their interested areas. If the users' profiles are analyzed, the result of the analysis must guarantee that the class assigned to each customer must be suitable for their interests.

The main concept for performing the customers' profiles analysis is obtained from the applying the search engine architecture like google, yahoo, and msn. However, this concept does not guarantee that all customers can be classified to a certain suitable group. In the case of customers browse Webs randomly; these customers cannot be identified to any groups. So, without high performance proxy manager, the customers may obtain their requests slowly because of the bottleneck problem mentioned above.

Bottleneck of the proposed system at the proxy manager is concerned. Because it works as a junction of the system, intercepts every request, then searches the user's group of interest, and finally pushes the request to a corresponding proxy server. Queuing requests and searching data that related to the request are main tasks. Therefore, a computer with high processing power is needed in order to handle these tasks. However, the bottleneck will be reduced after some customers are identified to some particular proxy servers to serve their needs directly. Nevertheless, these particular groups are scoped into the particular proxy servers and cannot call on other URLs of other services.

In this system, many proxy servers are provided to support users' requests. The requests will be spread to each proxy server depends on the group of interest which can reduce the response time. With assumption that users in the same group of interest have the same target set of Web usages, the proposed proxy system should perform at high HIT rate. The main factor that should be considered is how to categorize the groups of interest in order to have number of users with equal distribution in each group. The number of users in each group impact to performance of the proxy server. For example, an ISP has two groups of interest, teenager and academia. If the ISP estimates that their users almost in the teenager group, then the ISP should split the teenager group into other groups such as entertainment, sport, travel, etc.

The development of information technology is rapidly changed, many new information services are offered over the Internet whilst many previous information services are changed or abrogated. Thus, the UDS must be updated from time to time otherwise the new services cannot be reached by users. Furthermore, the categories of users may have to be reviewed at least once in 6 months so the services' structure can match to the real world. However, the changes of the available services do not necessary cause the change of the ISP services' structure, or the criterion to identify users' groups unless there are a large gap among the current groups and the available services.

6. Conclusions

In this paper, a transparent policy-based proxy system, named the customer-oriented policy, is proposed. Transparent technique is applied to the system to eliminate browser configuration problem. A proxy manager is established to manage the system policy, matches user to a corresponding proxy server based on a specific group of interest in the user profile.

The significant point in this paper is to partition services into classes and offers customers to apply to a particular class. Since the proxy servers are assigned for a specific service area, the HIT rate is increased. Once the HIT rate is raised, the response time and the delay time of each customer is small because information may be retrieved directly from the proxy's cache.

The consequent of reducing MISS rate affects to the number of traffics sending out from ISP. When there is a small amount of transactions delivered to Internet, the available bandwidth will be shared with small amount of users. Therefore, the retrieving speed from Internet will be large and customer will receive their information in a short period of time.

The availability of the services is another issue that is considered for this architecture. The separation of services is the key to maintain services when a connection is not available. If a connection is down, the other connections to other services still available. Thus, the satisfactions of customers can easily be maintained.

According that the services are separately managed by individual proxy servers, the increasing of customers in the same domain of interest – the same proxy server – is still able to grant services without expanding the server size. Therefore, this architecture also supports the scalability of the Internet services of ISP.

As a result of supporting scalability of the services and the ability to reduce the traffic flows to

the Internet, the maintenance fee of the ISP can be reduced while their customers are satisfied with the services.

7. Future Works

Our remaining works are as follows.

- Classify the ratio of users for their interested area. The data in the logged file of an ISP will be analyzed to obtain the ratio of users in each area of interest.
- Set the hardware specification of proxy system based on the ratio of interest described above.
- Implement and evaluate the performance of the proposed architecture, comparing with the previous system.
- Perform the theoretical approval for the propose system.

Acknowledgement

We are grateful to Professor Dr. Morris Chang from Department of Electrical and Computer Engineering, Iowa State University, who sparkled the idea of the project and gave his valuable time for his valuable advice.

References

[1] S. H. and F. H. Bakry, "Identifying Information Network Profiles for Planning Management", *International Journal of Network Management*, vol. 10, no. 6, pp. 315-322, Nov/Dec 2000.

[2] L. J. G. T. van Hemmen, "Models Supporting the Network Management Organization", *International Journal of Network Management*, vol. 10, no. 6, pp. 299-314, Nov/Dec 2000.

[3] S. Jha and M. Hassan, "Java Implementation of Policy-Based Bandwidth Management", *International Journal of Network Management*, vol. 13, no. 4, pp. 249-258, Jul/Aug 2003.

[4] D. W.K. and C. S. Hong, "A QoS Management Framework for Distributed Multimedia Systems", *International Journal of Network Management*, vol. 13, no. 2, pp. 115-127, Mar/Apr 2003.

[5] P. Bhattarakosol and V. Ngamaramvaranggul, "An Internet Web Management Policy for Government Organisation", *Proceedings of Network Research Workshop*, pp. 249-255, The 18th APAN Meeting 2004.7.2-7.7, Cairns, Australia.