Efficient Consistency Control for Mobile Dynamic Contents Delivery Network

Zhou Su, Jiro Katto and Yasuhiko Yasuda

zhousu@asagi.waseda.jp

Graduate School of Science and Engineering, Waseda University, JAPAN

Abstract: This paper is to design an integrated consistency-control algorithm for mobile contents delivery network (M-CDN) to enhance the efficient utilization of network resource and support the client mobility. Firstly, by carrying out an analysis of mobile surrogates characteristics, for a given content which has been changed on its original node, only a limited number of its replicas instead of all replicas are updated. Then, if a replica has been selected for update, the latest version will be sent from an algorithm-decided site instead of from its original server.

Keywords: mobile contents, web consistency, wireless communication

1 Introduction

With the rapid developments of wireless communication technologies, a mobile contents delivery network has appeared to facilitate a set of moving nodes to communicate with each other via short-range wireless transmission protocols such as IEEE 802.11, Blue-tooth or Ultra Wide Band (UWB). In the mobile CDN, a moving node receives information from its neighbors, or from remote nodes by multi-hop transmission relayed by intermediated moving nodes [1][2].

This technology has attracted many attentions and applications: for example, the Nokia mobile server has been proposed to make a web-server running on a mobile phone accessible from the Internet using any web browsers. Furthermore, the Helix mobile server led by Real Networks is optimized for delivery of all major file formats to any wireless network, with the broadest support for mobile delivery standards. However, although the above companies advocate their wireless support for contents delivery, their technical details on how to keep the content consistency have not yet clarified nor verified.

In a mobile CDN, an important issue to be resolved is consistency management, which means the cached replicas on different sites must be updated if the originals change. If a requested object can be provided to the user with the latest version, the user will not need to contact the remote origin server to get the updated object. Therefore, the appropriate consistency control of server replicas benefits content providers by reducing latency for their clients, and benefits ISPs by reducing bandwidth consumption and transmission cost.

The conventional consistency methods for wired networks could not be efficiently carried out in mobile environment. It is because: in contrast to conventional CDN systems in wired networks that consist of surrogates, mobile CDN are subjected to the limitations of long-lived nature of typical multimedia sessions, wireless bandwidth, and the dynamically changed network topology. Challenges arise in how to improve the structure design and data consistency in a mobile CDN [3]-[5].

In this paper, we therefore propose an optimal algorithm for controlling Web consistency in mobile content delivery overlay.

2   Theory Analysis

3.1: Scalable Contents Selection

We consider an overlay network that has K mobile nodes. Each node i ($i=1,...,K$), has a λ_i (bytes/second) which denotes an aggregate request rate from clients to the server.

As for the contents, we assume that there are J different contents in our mobile CDN. A parameter P_j defines the request probability for content j.

According to the Zipf distribution which the distribution of many Web accesses follows, the probability that the content j is requested can be obtained as follows:

$$P_j = \frac{V}{w_j^{\alpha}} \qquad (1)$$

where V, α are parameters of the Zipf distribution, and w_j is the ranking of request times.

We assume that there is a Request Routing (RR) function in each node. The RR maintains the residence time value $r_{q,i}$ for each client q in node i's zone.

Then, the relative residence of a client with a node i can be

$$r_i = \frac{\sum_q r_{q,i}}{\sum_i \sum_q r_{q,i}} \qquad (2)$$

Assume that a client firstly enters in the area of node i and then requests contents j, we can get the probability that the content j is requested from node i to be:

$$\theta_{i,j} = \frac{V}{w_j^{\alpha}} \cdot \frac{\sum_q r_{q,i}}{\sum_i \sum_q r_{q,i}} \qquad (3)$$

3.2: Scalable Node Update

In a conventional method, if a surrogate decides to fetch the latest version of a modified document, it will contact the modified document's original server to fetch its latest version. However, this method causes a bad load balance if all requests going to the original server. Furthermore, fetching the contents also increased the extra network traffic.

Assume that the node i has a coverage area denoted by the radius of ca_i and its distance between node k is $D_{i,k}$.

For each node k, if it is in the coverage area of node i, then we can get:

$$D_{i,k} \leq ca_i \qquad (4)$$

$$k = \{1,2,...K\} \ \& \ k \neq i$$

Let Z_k denote the capacity (the number of simultaneous sessions) of node k, and Z is the number of total sessions happened in the whole network within a testing past period. Then the average sessions happened in node k during the past period can be:

$$(\lambda_k / \sum_k \lambda_k) \cdot Z) \qquad (5)$$

A free-degree is defined as follows to show the number of free sessions in node k.

$$f_k = (Z_k - (\lambda_k / \sum_k \lambda_k) \cdot Z) \Big/ Z_k \qquad (6)$$

Assume that the received signal strength (RSS) of the node k is RSS_k. The probability that its signal is better than the threshold is:

$$l_k = P\{RSS_k \geq Thresh\}$$
$$= P\{RSS_k | RSS_k \geq Thresh\}/T \qquad (7)$$

where $Thresh$ is the set threshold and T is the total detecting times within the testing period.

We define an index for scalable update as follows:

$$\Delta_k = \eta \cdot f_k + \varphi \cdot l_k \qquad (8)$$

$$\eta + \varphi = 1 \qquad (9)$$

Let $X_{i,j}$ be a parameter which takes a binary value of

$X_{i,j} = 1$ (if content j is stored in node i)
$X_{i,j} = 0$ (otherwise) $\qquad (10)$

Then, we can get a matrix X of which element is $X_{i,j}$, which represents a placement pattern of contents.

For the nodes, which are in the coverage of node i, and also have the replica of content j, we can select the node k with the maximum update index by

$$\begin{cases} i \neq k \quad \& \quad X_{k,j} = 1 \\ D_{i,k} \leq ca_i \qquad (11) \\ max(\Delta_k) \end{cases}$$

3 Proposal

We present our algorithm as follows:

1. Step1: Scalable Update   Selection

When a given content j changes at its original server $o(j)$, a consistency priority $\theta_{i,j}$ will be calculated according to Equa.3. For content j's each replica ($X_{i,j}$ =1) over the whole overlay network, only when its priority $\theta_{i,j}$ is beyond the threshold, the replica of content j at node i will be updated.

Otherwise, this replica will not be updated until a new request for content j happens at the site i next time.

Therefore, when a given content j is changed at its original server, not all its replicas ($X_{i,j} = 1$) over all overlay network will be updated according to the analysis of Web access distribution.

2. Step2: Lowest Delay Update

Assume that there are K ($X_{k,j} = 1$, $k = \{1, ..., K\}$) replicas of content j slected to be updated, for a replica at node k, an update priority Δ_k will be calculated according to Equa.11.

The latest version of content j will be sent to the node (which replica need to be replicated) from node k with the lowest Δ_k.

Therefore, the latest version will be sent from an algorithm-decided site instead of its original server resulting in the reduction of network traffic and user delay.

4 Conclusion

In this paper, we proposed a novel algorithm for consistency control in mobile contents delivery overlay. More simulation results will be carried out from the practical viewpoint. Furthermore, theoretical modeling for consistency control should be expanded and sophisticated.

Reference

[1] W.M.Aioffi, G.R.Mateus , J.M.Almedia , R.C.Melo . "Dynamic Content Placement for Mobile Content Distribution Networks". WCW 2004 , Berlin 2004

[2] H.Shen, M.S.Joseph, M.Kumar, and S.K. Das, "PReCinCt: A Scheme for Cooperative Caching in Mobile Peer-to-Peer Systems", Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium, 2005

3] G. Antoniol, G. Casazza, G. Di Lucca, M. Di Penta, and E. Merlo, "Predicting Web Site Access: An Application of Time Series", Proc. IEEE the third International Workshop on Web Site Evolution, Florence, Nov.2001.

4] M. Tariq, R. Jain and T. Kawahara: "Mobility Aware Server Selection for Mobile Streaming Multimedia Content Distribution Networks, "Proc.International Web Caching Workshop (IWCW), Sep. 2003.

5] L. Breslao, P. Cao, L. Fan, G. Phillips, and S. Shenker "Web Caching and Zip-like Distributions: Evidence and Implications" Proc. IEEE INFOCOM'99, New York, Apr. 1999.