Distributed Hosting of Web Content with Erasure Coding and Unequal Weight Assignment

Jin Li[+] and Cha Zhang*

[+]*Microsoft Research*
One Microsoft Way, Bld. 113, Redmond, WA 98052
jinl@microsoft.com

**Carnegie Mellon University*
Porter Hall B42, Pittsburgh, PA 15213
czhang@andrew.cmu.edu

Abstract

By sharing and hosting the personal web content in a peer-to-peer (P2P) fashion, we may increase the bandwidth of retrieval and improve the reliability of retrieval. The cost is that the web content has to be replicated to and stored in the peers, consuming valuable network bandwidth and peer storage space. In this work, we develop two technologies, namely hierarchical content organization with unequal weight assignment and erasure coding, to reduce the amount of content to be distributed, yet still maintain the retrieval speed up and reliability. Significant improvement over the ordinary web server is demonstrated.

1. Introduction

The Internet empowers everyone to be a potential publisher. More and more people are producing their own web pages, sharing things such as the diaries, weblogs (Blogs), personal photo/video collections, and personal experience/knowledge/advice. We generally refer to the above as the web content. Unlike big publishers, which may rely on the expensive server arrays and dedicated network links to deliver the content, the server employed by the individual consumer is usually no more than a home computer, and the network capacity is no more than a single Internet connection to the ISP. Both the server and the network link of the consumer can be unreliable and insufficient in capacity (serving bandwidth) to respond to the content access request of the client.

To improve the capacity and/or the reliability of the server, one possible solution is to upgrade the server hardware/software and the network link/contract with the ISP. However, such upgrade solution can be costly. An alternative cost effective solution is to build a peer-to-peer (P2P) network. The server replicates the to-be hosted content in its entirety to the peers. When the content is accessed by the client, it can be either accessed from the original server through the server's network link, or accessed from the peers that hold a duplicated copy of the content, or even from both the server and the peers. The content access becomes more reliable, as it is unlikely that all the computers and the associated network links that hold the content are down. Moreover, the serving bandwidth of the content is increased as well, as the content can now be retrieved from multiple computers and multiple network links.

In such a P2P web hosting network, the content has to be replicated to the peers prior to its access. The replication process uses up valuable bandwidth resource and storage space. In this work, we adopt two strategies to reduce the amount of content to be replicated to and stored in the peers. First, we apply erasure coding on the web content, so that each peer may choose to host a partial copy of the content, and the client may mix and match the partial content hosted by the peers and assemble the wanted web page. Second, we organize the web content into a hierarchical structure and assign different weights to different portion/category of it. The web pages that are frequently accessed, the text portion of the web page, and the base layer of the photo/video collection may be assigned with a larger weight, and be replicated more extensively. On the contrary, the web pages that are less frequently visited, the decorative portion of the web page, and the enhancement layer of the photo/video collection may be assigned with a smaller weight, and be distributed more restrictively. Through organizing the web pages, assigning different weights to different web files, and applying the erasure coding, we greatly reduce the amount of content to be replicated to and stored in peers, whereas maintain the key benefits of the distributed hosting solution, i.e., retrieval speedup and reliability improvement.

The rest of the paper is organized as follows. We briefly review the framework of the P2P web hosting and the related work in Section 2. The erasure coded content replication and the hierarchical content organization with unequal weight assignment technologies are described in Section 3 and 4, respectively. Experimental results are shown in Section 5. Conclusions are given in Section 6.

2. P2P Web Hosting and Related works

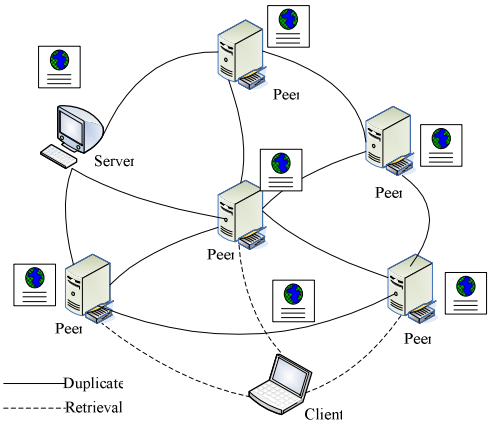


Figure 1 P2P Web Hosting.

The framework of the proposed P2P web hosting system can be shown with Figure 1. First, the server publishes the web content, be it web pages, Blogs, or photo/video collections. The web content is then distributed to a number of peers, often in a partial, erasure coded form. At the time of retrieval, the client locates the server and/or a number of peers that hold the partially replicated content. The peer list

0-7803-8603-5/04/$20.00 ©2004 IEEE.

may be provided by the server, or it can be provided by certain distributed hash table (DHT) technology, e.g., the PNRP protocol of the PeerNet SDK[8]. It then retrieves the web pages from both the server and the peers.

According to an eMarketer survey [7], the number of broadband household worldwide reaches 86.2 millions in 2003. The combined computation and network resource of this consumer P2P network, if linked, exceeds the resource of any corporation on earth. Consumer P2P applications, such as Napster, Gnutella and KaZaa, are highly popular in this network. At first, it seems rather straightforward to extend the sharing of files to the sharing of web site. However, the access of web pages has special characteristics that present unique challenges.

First, a web site consists of a collection of files that are tightly knit together. As a result, the access of a single web page usually results in the access of a collection of files, with predictable follow-up accesses pointed by hyperlinks that often lead to the access of another collection of files (web pages) in the same web site. Second, a web site is much larger than a single file, and consumes more bandwidth to replicate and more storage space to host. With the proliferation of the digital cameras/camcorders, more and more photo/video collections appear on the web page, resulting in larger and larger web sites. Third, a web page must be retrieved with sufficient speed (serving bandwidth) during the access of the client. This is rather different from the P2P sharing of files, where the client may afford to retrieve the file slowly, often hours and even days. The web page must also be readily available, as the client is unlikely to wait for hours for the proper peers that hold the content to come online.

Thus, in addition to the challenges we face in a typical P2P application, e.g., providing the proper incentives for the client to host the web site [4], and locating the distributed content in the P2P network for retrieval [5], a P2P web hosting application must deal with unique challenges of the web hosting application. In this work, we focus on one special issue of P2P web hosting, i.e., to distribute and store the web content among the peers in an efficient fashion.

Commercial corporations have been using multiple servers to improve the capacity and the reliability of the web hosting for quite some time. The web servers can be either tightly coupled in a local area network [1], or be loosely coupled and distributed at different geographical locations, such as the content distribution networks (CDN)[2]. Akamai, one of the successful commercial corporations that use CDN solutions, has used multiple servers to host the web sites of some major corporations, such as IBM and FedEx. In the commercial web hosting, since the hosting corporation usually owns all the servers that host the content and the network links between them, the bandwidth required to duplicate the web content and the storage overhead needed to hold the web pages are usually not the primary concerns. This is also true for certain restrictive web hosting applications, such as YouServ [3], which is a solution to share files and web pages of individual users through standard web protocols on the intranet of a corporation. Existing research on distributive web hosting usually focuses on improving the response time of the server, such as the server placement strategy and direction of the web request to the proper server.

However, this is not the case with a consumer P2P network, where both the network bandwidth and the storage capacity is at a premium for the peers, and the P2P web hosting application is competing with other applications for such resources. Therefore, it is necessary to develop technologies that may improve the web hosting reliability and serving bandwidth while reducing the network bandwidth and storage capacity used to host the web site. In the following, we develop two technologies for this, namely erasure coded content replication and hierarchical content organization with unequal weight assignment.

3. Erasure Coded Content Replication

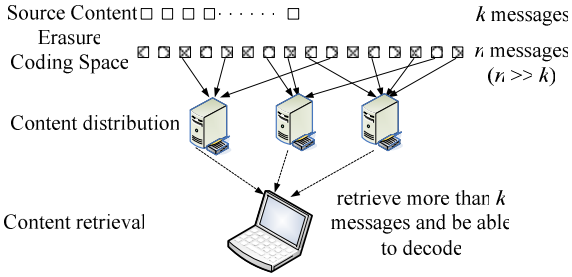


Figure 2 Erasure coding of content.

The technology of erasure coded content replication can be shown in Figure 2. We split the web content into a number of blocks, B_m, m=0,1,…,M-1, each of which is the smallest unit for the web access and retrieval. The block B_m is further split into k original messages. Through an (n,k) erasure codec, we form an erasure coding space of n coded messages. When the web site is distributed to the peers, each peer i picks p_i distinct messages out of the n messages of the erasure coding space according to certain peer replication ratio $w_{m,i}$ of the block B_m and the peer i:

$$p_i = w_{m,i} \cdot k \tag{1}$$

The number of messages p_i to be replicated can be a fractional value, which is simply interpreted as being $\lfloor p_i \rfloor$ with probability $(1+\lfloor p_i \rfloor - p_i)$, and being $\lfloor p_i \rfloor$+1 with probability $(p_i - \lfloor p_i \rfloor)$, where $\lfloor x \rfloor$ is the floor function. To make sure that the messages distributed to the peers are unique, we may assign a different erasure coding key space for each peer. The aggregated content replication ratio of the block B_m is denoted as C_m, which is the total amount of copies of the block B_m replicated in the P2P network:

$$C_m = \sum_i w_{m,i}, \tag{2}$$

At the time of retrieval, as long as the client can find k distinct messages on the on-line peers, it can retrieve and decode the original block. Apparently, with erasure coded content distribution, the retrieval reliability and the serving bandwidth grow in proportional to the aggregated content replication ratio. With the peer replication ratio $w_{m,i}$ being constant, by doubling the aggregated content replication ratio, the average serving bandwidth almost doubles, as there are twice as many peers holding and serving the content. The P2P network may also retrieve the content with double the

reliability, i.e., to retrieve the content successfully with the peers being online with half the probability.

In this work, we use the Cauchy-based Reed-Solomon erasure codes [6]. We choose Reed-Solomon codes over the other erasure coding technologies, such as Tornado codes and LDPC codes because of the maximum distance separable (MDS) property of the Reed-Solomon codes, which guarantees decoding as long as k distinctive coded messages are received. Compared to the Vandermonde matrices based Reed-Solomon codes, Cauchy-based Reed-Solomon codes have low decoding complexity for erasure coding in exchange for a higher complexity for error correction coding. This suites the P2P web hosting application well, as the primary form of the error is the loss of the coded messages caused by the drop of peer connection or the loss of the packets during network transmission.

Parameter k of the erasure codes determines both the granularity of the block as well as the size of the erasure coding space. The original block is broken into k equal sized messages. The larger the parameter k, the more pieces that the block is broken into, which leads to an increase of both the granularity of the access and the overhead of the erasure coding. On the other hand, the maximum size of the erasure coding space n is capped by 2^k-1. Therefore, k must be sufficiently large to make sure that every peer may have unique keys in the erasure coding space. In this work, we use the parameter k=16 and n=2^k-1=65535. This may accommodate at least 4095 peers. Furthermore, each message piece and the resultant coded message are currently at or below 1 KByte. Therefore, each coded message can be sent via a single network packet.

4. Hierarchical Content Organization with Unequal Weight Assignment

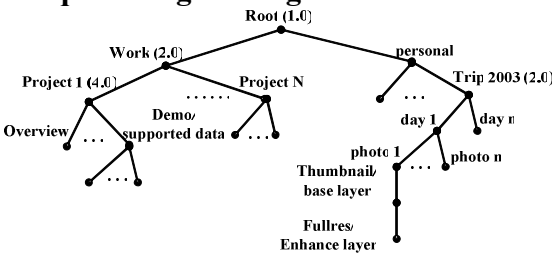


Figure 3 Sample web page structure and weight assignment.

A web site is naturally hierarchical. We show a sample web page in Figure 3. The site can be organized by topics, e.g., by splitting between the work related portion and the personal related portion, by time/event, by interest, etc.. The files that form the web site are not of the same importance, or weight. That is, they may not be visited with the same frequency, and their missing may lead to different level of annoyance. For example, recently created web pages are usually visited more often than the stale ones, and the web site of certain topics may be more popular than the others. Even the files of the same web page may vary in importance as well. For example, pictures and icons may be less important than the web page text, which conveys the basic information of the web page; and the thumbnails of the pictures may be viewed more often than the large full resolution pictures.

Recognizing the differences in the web files, we assign each web file a weight value that is associated with its replication ratio, which in turn governs the reliability of the retrieval and the serving bandwidth. In this work, the weight of the web file consists of two parts:

$$R_m = S_m \cdot T_m, \tag{3}$$

where R_m is the weight of the web file/blocks, S_m is the site weight, and T_m is the type weight. In general, doubling the weight leads to the double of the aggregated content replication ratio, which causes the resultant web file to be retrieved with twice the serving bandwidth and retrieval reliability. The site weight S_m reflects the differences in weights by topics/themes. It is manually assigned by the web site owner. A sample assignment of the site weight is shown in Figure 3. To reduce the manual labor of assigning the weight, the owner only needs to manually assign the site weight for a few nodes, and the rest unassigned nodes will simply inherit their weight values from the parent nodes. For example in Figure 3, only the site weights of the root, the work, the project 1, and the Trip 2003 nodes are manually assigned. The site weights of the rest of the nodes are inherited from the weights of the parent nodes through the hierarchical structure. The type weight T_m reflects the differences in weights by the types and attributes of the files. We may, for example, assign the full resolution pictures with a type weight that is half of that of the thumbnail pictures. With the weights of all the web files assigned, we now establish the relationship between the weight and the peer replication ratio. Assuming that a certain peer node i agrees to host no more than K_i bytes of the web site, the key is to find a relative peer replication ratio λ_i, such that the peer replication ratio of the block B_m is calculated to:

$$w_{m,i} = \max\{1, R_m \cdot \lambda_i\}, \tag{4}$$

The total amount of content to be replicated to the peer node i can be represented as:

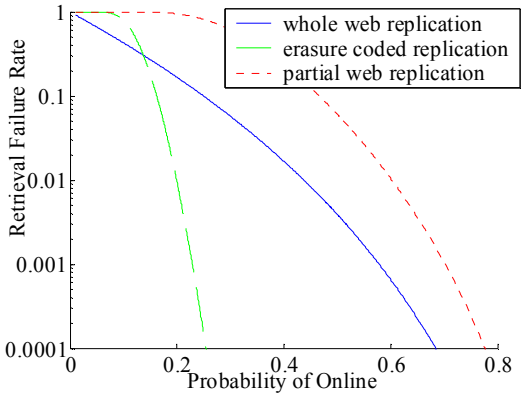
$$D(\lambda_i) = \sum_m |B_m| \max\{1, R_m \cdot \lambda_i\}, \tag{5}$$

Since $D(\lambda_i)$ is monotonically increasing with the increase of λ_i, we may find the largest λ_i given the amount of content (K_i) allowed to be replicated to the peer node. We may then use (1) and (4) to calculate the number of coded messages to be replicated to the peer for each web file.

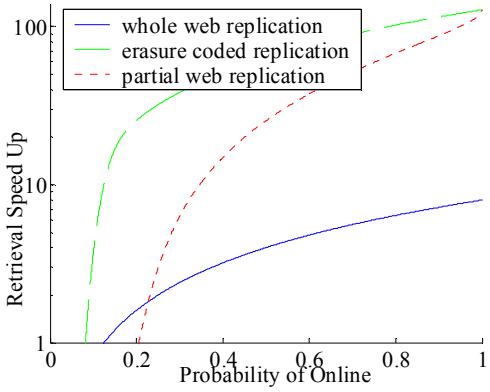
5. Experimental Results

In the first experiment, we compare the erasure coded content distribution with two alternative strategies: 1) replicating the web site in its entirety, 2) partially replicating the web site without erasure coding. In the 2nd scheme, each block is split into k pieces, however, during the replication stage, the original message piece is sent to the peer without erasure coding. We assume that the same amount of network and storage resources is used to distribute and host the content, i.e., the aggregated content replication ratio C is the same for all schemes. We further assume that the original server is off line, and each of the peers has an identical serving bandwidth, and has an independent probability (p) of being online to serve the client. We may then calculate the probability of successfully retrieval of the web site in the P2P

network, as shown in Figure 4(a). We may also calculate the average speed up of retrieval for the various content replication schemes, and show it in Figure 4(b). In both figures, the horizontal axis is the probability (p) of a peer node to be online. The parameter of the content distribution is C=8, k=16.



(a)



(b)

Figure 4 Comparison of the whole web replication (solid curve), partial web replication (dotted curve) and the erasure coded replication (dashed curve) with (a) the retrieval failure rate and (b) the retrieval speed up [C=8,k=16].

We observe that with the same amount of network and storage resource consumed, the erasure coded content distribution provides a great improvement in the reliability of retrieval and a substantial increase in the serving bandwidth. In fact, in the example of Figure 4, once the probability that a peer is online is greater than 0.13, the rate of failure to retrieve the web site for the erasure coded content representation is thousands times smaller than that of the whole web replication and the partial web replication without erasure coding. Moreover, the erasure coded content replication always has a lower retrieval failure rate compared with the scheme without erasure coding. Compared to the whole web replication and the partial web replication without erasure coding, the erasure coded content replication also speeds up the retrieval around 16 times and 1-10 times, respectively.

We have built a P2P web hosting system with erasure coded content distribution and hierarchical content organization with unequal weight assignment. In Figure 5, we show an example of a test web site. In the test setup, the web site is replicated to and hosted on 7 peers. The original web site occupies 228 megabytes. During the replication, each peer agrees to host 60 megabytes of the web site, results in an average replication ratio of 0.26. Since the web files are unequally weighted, the peer replication ratio for the actual web files varies, ranges from 0.25 to 1.0. During the web page retrieval, the client retrieves the web from the 7 peers simultaneously, erasure decodes the web page, and renders the web. In this test set up, the client is still able to retrieve the web page as long as there are more than 4 peers online. A running scene is shown in Figure 5.

Figure 5 Screen capture of a running P2P web client.

6. Conclusions

We have developed a P2P web hosting application, in which the client may retrieve the web page simultaneously from the server or the peers that host a partially replicated copy of the web site. We use erasure coded content replication and hierarchical content organization with unequal weight assignment to reduce the amount of content to be distributed. The P2P web hosting increases the serving bandwidth and improves the reliability of the web retrieval.

7. References

[1] V. Cardellini, E. Casalicchio, M. Colajanni and P. S. Yu, "The state of the art in locally distributed web-server systems", *ACM Computing Surveys (CSUR),* vol. 34, no. 2, Jun. 2002, pp.263-311.

[2] D. C. Verma, *Content distribution networks: an engineering approach,* Wiley Publishers, 2002.

[3] R. J. Bayardo Jr., R. Agrawal, D. Gruhl and A. Somani, "YouServ: A web-hosting and content sharing tool for the masses", *WWW 2002,* May. 7-11, 2002. Honolulu, Hawaii.

[4] T. Ngan, D. Wallach, P. Druschel, "Enforcing fair sharing of peer-to-peer resources", In Proc. IPTPS '03, Berkeley, CA, 2003.

[5] S. Ratnasamy, S. Shenker, I. Stoica, "Routing algorithms for DHTs: some open questions", In Proc. IPTPS'02, Berkeley, CA 2003.

[6] S. B. Wicker and V. K. Bhargava, Reed-Solomon Codes and their applications, IEEE Press, New York, 1994.

[7] http://www.info-edge.com/samples/EM-2094sam.pdf

[8] Windows XP Peer-to-Peer Software Development Kit.