CHAPTER 7
CONCLUSION AND FUTURE ENHANCEMENTS

7.1 Objectives Revisited

The main objectives of this thesis work, as stated in previous chapter 1 are achieved as mentioned below.

- **Reducing amount of Information to be transmitted:** The amount of information to be transmitted through Internet is reduced by using compression, optimization techniques or keeping the frequently accessed cacheable documents locally. Browser aware clustered architecture is developed for keeping the remote data as close as possible and this technique survives from failure also. Compression and optimization and web caching techniques are incorporated into the adaptive traffic/latency reduction technique in order to reduce the data to be transmitted through Internet.

- **Efficient utilization of existing bandwidth:** For achieving efficient utilization of the existing bandwidth of current network, dynamic pre-fetching technique is proposed and it is also incorporated into the adaptive traffic/latency reduction technique.

- **Avoiding network bottlenecks:** For avoiding network bottleneck such as congestion, traffic threshold values are used in the client side adaptive traffic/latency reduction algorithm. Pre-fetching occurs only if the current network bandwidth usage is less than all these threshold values. Thus the possibility of occurrence of congestion due to pre-fetching is avoided.
• **Increasing cache hit ratio:** Browser aware clustered cache sharing architecture, dynamic pre-fetching technique and adaptive traffic/latency reduction algorithms are proposed for increasing cache hit ratio. Browser aware clustered cache sharing architecture increases cache hit ratio up to 60%, dynamic pre-fetching technique increases it up to 75% and adaptive web traffic/latency reduction technique increases it up to 82%.

• **Reducing latency:** Browser aware clustered cache sharing architecture, dynamic pre-fetching technique and adaptive traffic/latency reduction algorithms are proposed for reducing latency. Browser aware clustered cache sharing architecture reduces latency up to 25%, dynamic pre-fetching technique reduces it up to 63% and adaptive web traffic/latency reduction technique increases it up to 65%.

• **Permitting the disconnected operation of web pages:** Since the frequently accessed cacheable documents are stored in cache, the access to the desired web pages in cache is permitted even if the connection to the web server is lost. Browser aware clustered cache sharing survives from failures and provides documents if it is available in any of the cooperating caches in the local network.

### 7.2 Conclusions and Major Contributions

In this research work, client side web caching and pre-fetching techniques were studied and new techniques were reported for improving the efficient access of web sites. The consolidation of novel techniques reported in the present work is given below.
• In **browser aware clustered cache sharing architecture**, a coordinator is elected for each cluster. Remote summaries, cluster summaries and browser index file are stored in the coordinator. The coordinator forwards the requests to the web servers only if, it could not find cache hit in these summaries of cache directories and browser index file.

• In **dynamic pre-fetching technique**, bandwidth usage is dynamically controlled for avoiding congestion and maximizing the utilization of the existing capacity of the network. In this technique, the number of URLs to be pre-fetched depends upon the maximum capacity of the network, existing bandwidth usage, user preferences and weights of URLs obtained from the user's access pattern.

• In **adaptive traffic/latency reduction technique**, client side and web server side algorithms are developed for achieving web traffic/latency reduction and accessing the web sites efficiently. Existing traffic/latency reduction techniques such as compression, optimization, web caching, dynamic pre-fetching and traffic dispersion are incorporated into the adaptive traffic/latency reduction technique for increasing cache hit ratio and reducing latency.

Table 7.1 shows the comparison between the maximum benefits achieved by existing web caching and pre-fetching techniques with newly reported techniques. Among the various caching techniques found in literature, the maximum cache hit ratio achieved is 50% [34, 35] and latency reduction achieved is 28% [37]. New browser aware clustered cache sharing technique increases cache hit ratio up
to 60% and reduces latency up to 25%. Since it is a trade off method between memory requirements and inter-proxy traffic, it is to be preferred over the existing web caching techniques and architectures.

Among the various web pre-fetching techniques found in the literature, the maximum cache hit rate achieved is 30-75% [36] and latency reduction is 60% [38, 39] if the pre-fetching decision is not left to the user. Dynamic pre-fetching technique makes use of the idle time of the network for pre-fetching and reducing the delay in accessing the web pages. The dynamic pre-fetching technique avoids congestion by maintaining almost constant traffic and guarantees the maximum utilization of the existing bandwidth capacity. It also increases cache hit ratio up to 75%, reduces latency up to 63%. So this technique is to be preferred to the existing pre-fetching techniques.

<table>
<thead>
<tr>
<th>Table 7.1 Comparison of Caching and Pre-fetching Techniques</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Web caching methods found in literatures</strong></td>
</tr>
<tr>
<td>Maximum increase of cache hit ratio</td>
</tr>
<tr>
<td>Web caching methods found in literatures</td>
</tr>
<tr>
<td>Browser aware clustered cache sharing architecture</td>
</tr>
<tr>
<td>Pre-fetching techniques found in literature</td>
</tr>
<tr>
<td>Dynamic pre-fetching technique</td>
</tr>
<tr>
<td>Adaptive web traffic/latency reduction technique</td>
</tr>
</tbody>
</table>

* Some methods such as model based predictive pre-fetching [211], adaptive pre-fetching [212, 213, 214] and non-interfering pre-fetching [216] have bandwidth monitoring and traffic control facility, whereas most of all other pre-fetching techniques do not have such facility.
The adaptive traffic/latency reduction technique is a hybrid technique by incorporating most of all existing traffic/latency reduction techniques. Using adaptive traffic/latency reduction technique, cache hit ratio increased up to 82% and the latency is reduced up to 65%. The average bandwidth utilization using this technique is 80%. Since adaptive traffic/latency reduction technique avoids congestion, maintains constant bandwidth usage, increases cache hit ratio and reduces latency, it is to be preferred to any other existing traffic/latency reduction techniques.

7.3 Scope for Future Enhancements

Although researchers have given constant attention to caching and prefetching research, maximum traffic/latency reduction has not been achieved. Further research work is required for accessing the web documents in real time. Even if the newly reported techniques in this thesis increase cache hit ratio and reduce latency, there is scope for future extension of these techniques to enhance web performance. The following enhancements of this thesis work have been identified.

- Load balancing [6, 7] can be incorporated into the browser aware clustered cache sharing architecture. Whenever a proxy in the cluster is overloaded, it can inform the coordinator and the coordinator can distribute the work to other proxies in that cluster.
- In the browser aware clustered cache sharing architecture, cluster identities and member proxies within each cluster are decided by the
administrator by configuring an applet. Instead, cluster formation and adding member proxies to each cluster can be done automatically by measuring the proximity distance between proxy servers and current work load.

- In dynamic pre-fetching technique, priorities of the users are not considered. The access patterns of URLs are represented by weights of URL's in hash table. There is no provision for identifying the last accessed user of these URL's. The hash table can be extended for storing meta information such as date, time and user of each URL accessed last, last modification date and highest priority of the user accessed. This will help the intelligent agent to find the sites to be pre-fetched more accurately.

- Dynamic pre-fetching technique is useful only if the user always surfing the web in the same local area network using the same proxy server or the user is surfing the web using a single computer always. This technique can be extended to access the web from anywhere.

- In adaptive web traffic/latency reduction technique, intelligent agents can be used for controlling the compression and quality of online transactions and video conferencing applications. In video compression technique, agents can be used at both client and web server sides for identifying key frames and subsequent frames can be compressed based on these key frames.

- In adaptive traffic/latency reduction technique, intelligent agents can be maintained at server side for finding the auto summarization of text and audio-video presentations [232, 233].