



Caching Behavior of Web Browsers

Caching Behavior of Web Browsers

White Paper



WHITE PAPER

Caching Behavior of Web Browsers

When a user visits a web page, the contents of that page can be stored in the browser's cache so it doesn't need to be re-requested and re-downloaded. Efficiently using the browser cache can improve end user response times and reduce bandwidth utilization.

The cache-ability of an item on the browser is determined by:

- The response headers returned from the origin web server. If the headers indicate that content should not be cached then it won't be.
- A validator such as an ETag or Last-Modified header must be present in the response.

If an item is considered cacheable, the browser will retrieve the item from cache on repeat visits if it is considered "fresh." Freshness is determined by:

- A valid expiration time that is still within the fresh period.
- The browser settings as explained below.

If a representation is stale or does not have a valid expiration date, the browser will ask the web server of origin to validate the content to confirm that the copy it has can be served. The web server will then return a 304 to let the browser know that the local cached copy is still good to use. If the content has changed, the web server returns a 200 response code and delivers the new version.

How the browser cache is used is dependent on three main things:

- Browser settings
- The web site (HTML code and HTTP headers)
- How the user loads the page

Browser Settings

The user can configure how they want cached content to be stored and delivered from their local cache, or whether they want the content cached at all. Internet Explorer and Firefox classify these slightly different.

Every visit/view to the web page

When a user returns to a page that was previously visited, the browser checks with the origin web server to determine whether the page has changed since last viewed.

Every time I start the browser/Once Per Session



WHITE PAPER

Caching Behavior of Web Browsers

If a page is revisited within the same browser session the content will be delivered from the cache. When browser is closed and then reopened, a request will be sent to check whether the content has changed. If a page is visited during the same browser session, the cached files will be used instead of downloading content from the web server of origin.

Automatically/When the page is out of date

When the browser is closed and then reopened on repeat visits, it will use the lifetime settings of the cached content. If the same page is visited during a single browser session the cached files will be used. This is the default setting for both Internet Explorer and Firefox.

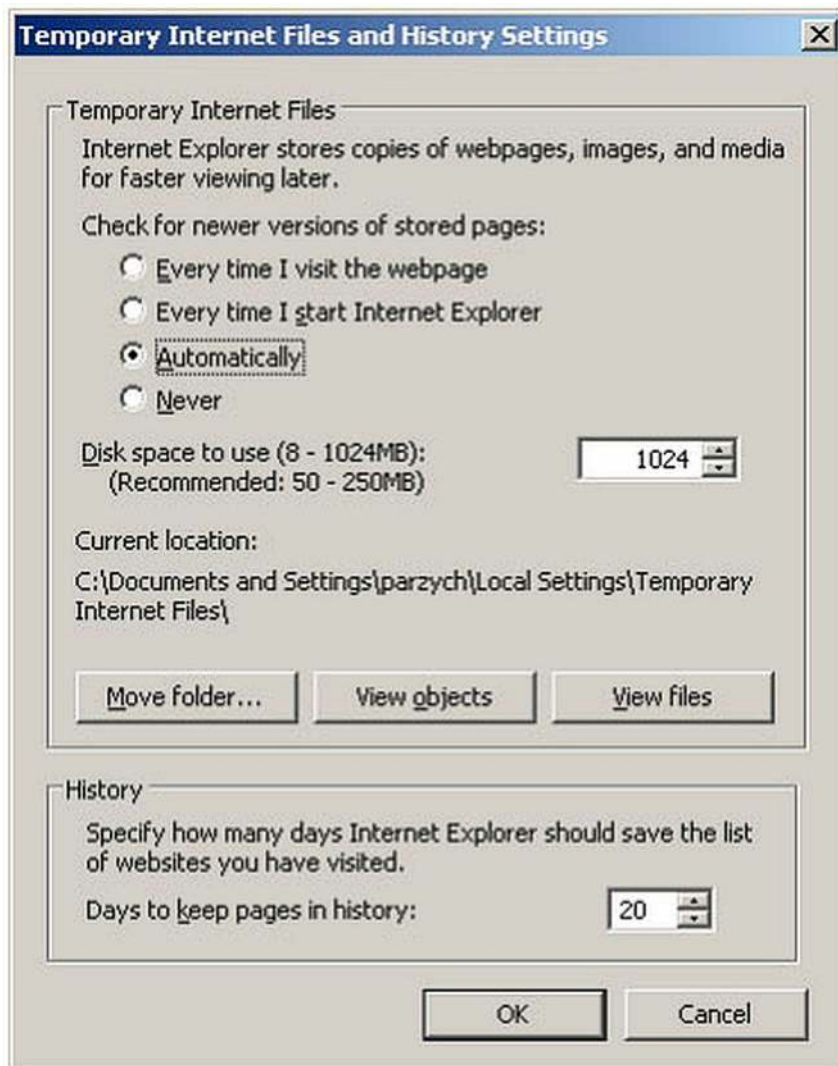
Never

The browser will not check with the origin web servers for newer content.

These settings can be configured in the following ways for IE and Firefox:

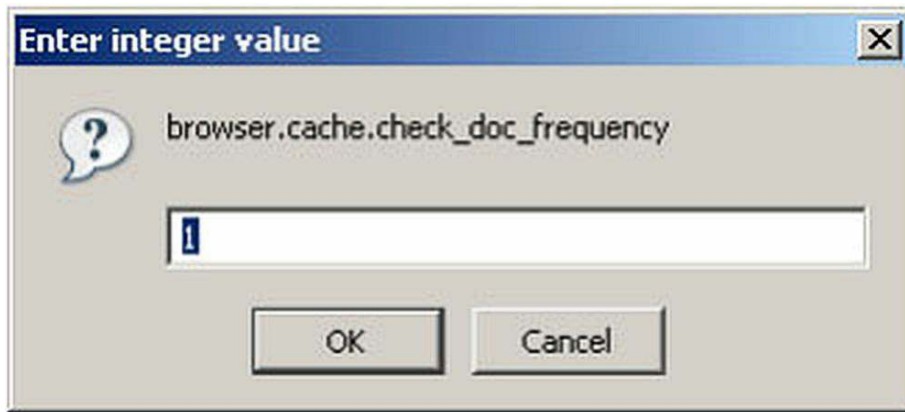
Internet Explorer

- Select **Tools**
- Select **Internet Options**
- IE 7 From the **General Tab** under **Browsing** history select settings
- IE 5 or 6 under **Temporary Internet Files**, click **Settings**



Firefox

- Type **about:config** in a Firefox browser
- Double-click the **browser.cache.check_doc_frequency** setting
- Enter the desired integer value in the dialog box



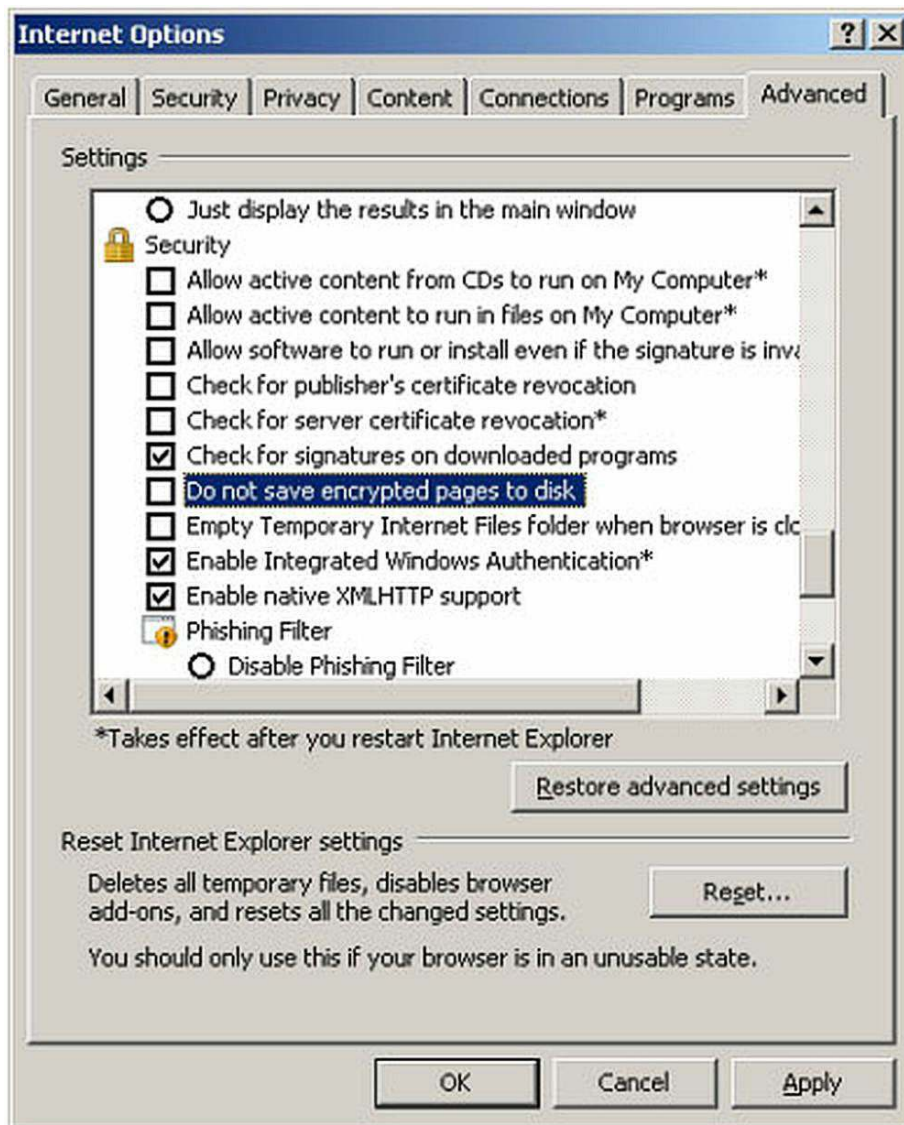
- 0 = Once per session
- 1 = Every time I view the page
- 3 = When the page is out of date (default)
- 2 = Never

In addition to configuring general cache settings, there are additional settings to configure that control whether SSL content is cached. When this option is enabled any SSL content is not stored to disk this includes the static images and includes forcing the browser to request the content on every visit to the page. Internet Explorer has this disabled by default, while Firefox has it enabled by default.

To enable/disable caching of SSL content:

Internet Explorer

- Select **Tools**
- Select **Internet Options**
- Select **Advanced**
- Under the **Security** section
 - Select the "Do not save encrypted pages to disk" option to not cache SSL content
 - De-select the "Do not save encrypted pages to disk" option to cache SSL content



Firefox

- Type **about:config** in a Firefox browser
- Double-click the **browser.cache.disk_cache_ssl** to change the setting
 - "True" indicates SSL content will be cached
 - "False" indicates SSL content will not be cached



WHITE PAPER

Caching Behavior of Web Browsers

The Web Site

In order for content to be served from the cache, the URL has to be an exact match to the content in the cache. Some web developers will add random numbers to part of the query string to ensure that the content is not cached and is always "fresh." When these random query strings are added to the URL the browser will not recognize the content as being the same as the item already in cache and a new GET request will be issued for the element.

In most instances the cache behavior of content is controlled by the Cache-Control and Expires HTTP headers. Cache-Control headers specify whether or not the content can be cached and for how long. The values can include:

- **no-cache** – Do not cache this content
- **private** – Can be cached by browsers, but not shared/public caches
- **max-age** – Set in seconds; specifies the maximum amount of time content is considered fresh

The inclusion of just an Expires header with no Cache-Control header indicates that the content can be cached by both browsers and public/shared caches and is considered stale after the specified date and time as shown below:

```
(Status-Line)  HTTP/1.1 200 OK
Content-Length  4722
Content-Type    image/gif
Date           Fri, 31 Aug 2007 10:20:29 GMT
Expires        Sun, 17 Jan 2038 19:14:07 GMT
Last-Modified   Wed, 07 Jun 2006 23:55:38 GMT

URL in cache?   Yes
Expires 19:14:07 Sun, 17 Jan 2038 GMT
Last Modification 23:55:38 Wed, 07 Jun 2006 GMT
Last Cache Update 10:20:32 Friday, August 31, 2007 GMT
Last Access      10:20:31 Friday, August 31, 2007 GMT
ETag
Hit Count       1
```

If no Cache-Control or Expires headers are present, the browser will cache the content with no expiration date as illustrated below:

```
Headers:
(Status-Line)  HTTP/1.1 200 OK
Accept-Ranges  bytes
```



WHITE PAPER

Caching Behavior of Web Browsers

```
Connection    keep-Alive
Content-Length 221
Content-Type   Image/gif
Date          Fri, 31 Aug 2007 10:27:06 GMT
Last-Modified  Fri, 02 Jun 2006 09:46:32 GMT

URL in cache?  Yes
Expires       (Not set)
Last Modification 09:46:32 Friday, June 02, 2006 GMT
Last Cache Update 10:26:32 Friday, August 31, 2007 GMT
Last Access     10:26:31 Friday, August 31, 2007 GMT
ETag
Hit Count      1
```

Some web developers have opted to use META Tags to control how content can be cached as opposed to setting cache parameters in the HTTP headers. Using the HTTP header is the preferred and recommended way of controlling the cache behavior.

Controlling Browser and Proxy Caches

```
<META HTTP-EQUIV="CACHE-CONTROL" CONTENT=" ">
```

There are four values that can be used for the content variable:

- **Private** – May only be cached in a private cache such as a browser
- **Public** – May be cached in shared caches or private caches
- **No-Cache** – Content cannot be cached
- **No-Store** – Content can be cached but not archived <META HTTP-EQUIV="EXPIRES" CONTENT="Mon, 22 Jul 2002 11:12:01 GMT"> The Expires tag should be used in conjunction with the Cache-Control tags to specify how long content can be stored.

Defeat Browser Cache

```
<META HTTP-EQUIV="PRAGMA" CONTENT="NO-CACHE">
```

When received, a browser will not cache the content locally; this is effectively the same as sending a Cache-Control=No-Cache header.

Refreshing Content or Redirecting Users to Another Page

```
<META HTTP-EQUIV="REFRESH" CONTENT="15;URL=http://www.example.com/index.html">
```




WHITE PAPER

Caching Behavior of Web Browsers

Refresh elements can be used to tell the browser to either redirect the user to another page or to refresh the page after a certain amount of time. The refresh tag works the same way as hitting the refresh button in the browser. Even if content has a valid expiration date, the browser will ask for validation that it has not changed from the server of origin. This essentially defeats the purpose of setting content expiration dates.

If a URL is specified in the META tag, that tells the browser to redirect to the specified URL after the time has elapsed. Redirecting users via the META tag as opposed to an HTTP-Response header is not recommended as META refreshes can be turned off by the user under the browser security settings.

How the User Loads the Page

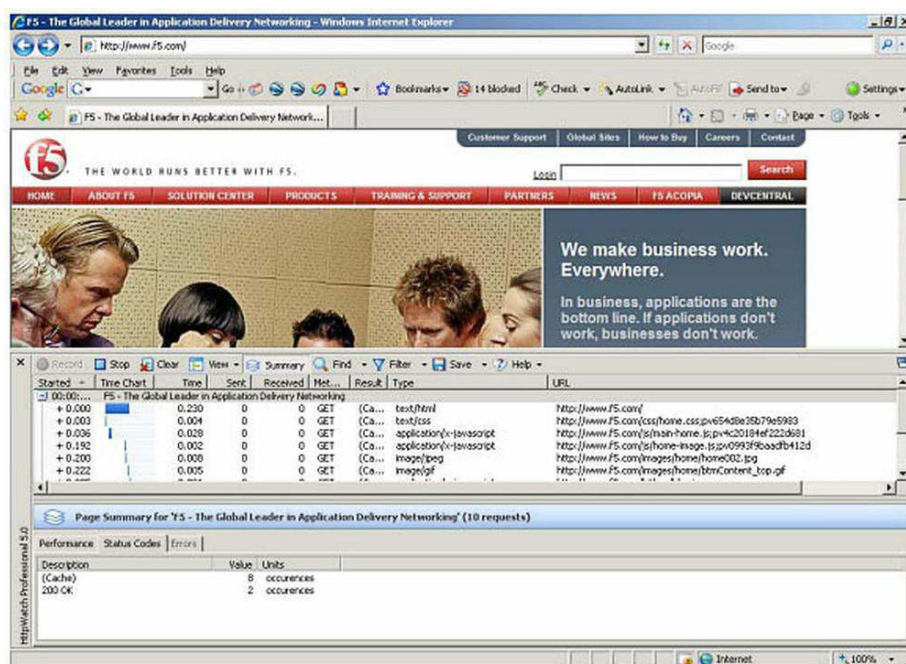
The use of how content is pulled from cache on repeat visits is impacted by the manner in which the request is issued.

WHITE PAPER

Caching Behavior of Web Browsers

Browsing Multiple Pages or Hitting the Back Button

While in the same browser session, all content for a site will be served from the local browser cache. If a user clicks through multiple pages of an application and the same graphics and elements are found on each page, the request will not be sent to the origin web server. Instead it will be served from the local cache. If the user re-visits a page during that session, all of the content—including the HTML—will be retrieved from the local cache, as shown in the image below (depending on the browser settings). As soon as the browser is closed, the session cache is cleared. For the next session, the only cache that will be used is the disk cache.



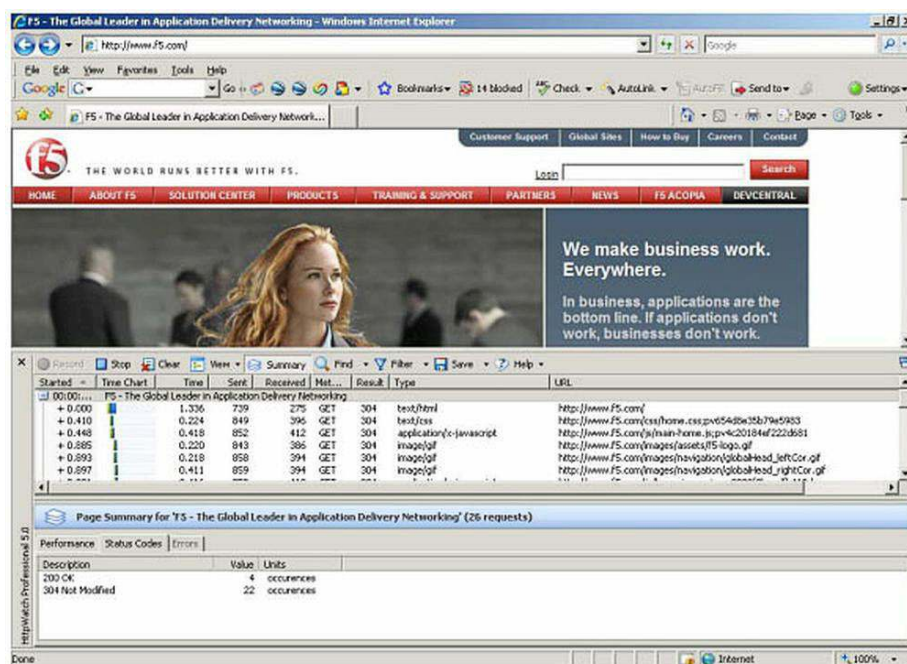


WHITE PAPER

Caching Behavior of Web Browsers

Refresh

Users might also hit refresh on a page to check for new content, such as an updated sports score or news article. Hitting refresh results in an "If-None-Match" header being sent to the origin web server for all content that is currently on the disk cache, independent of the expiration date of the cached content. This results in a 304 response code for each reusable item that is currently in the browser's cache, as illustrated in the picture below.



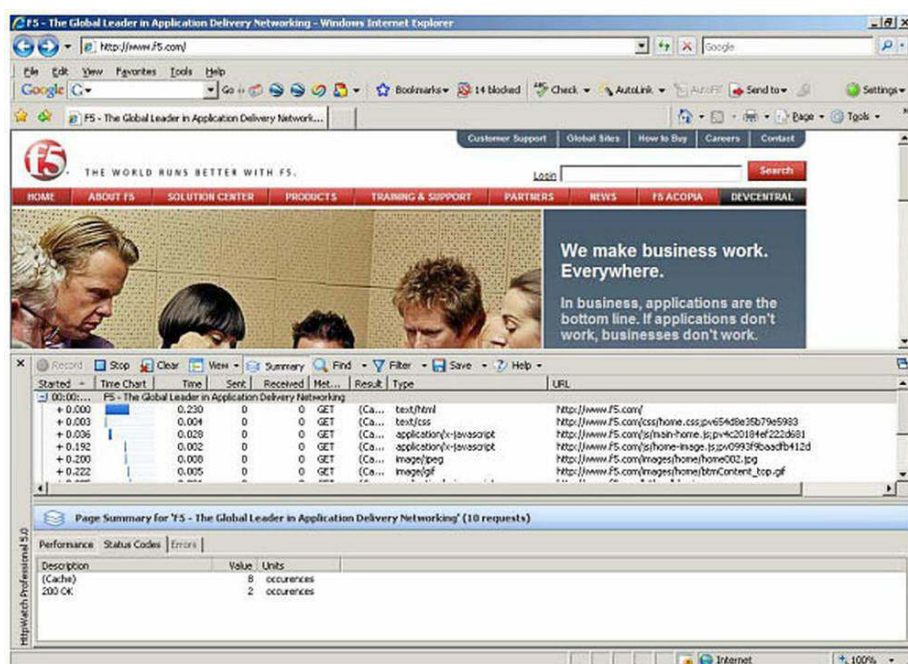


WHITE PAPER

Caching Behavior of Web Browsers

CTRL + Refresh or CTRL + F5

Hitting CTRL and refresh (in Internet Explorer only) or CTRL and F5 (Internet Explorer and Firefox) will insert a "Cache-Control=no-cache" header in the request, resulting in all of the content being served directly from the origin servers with no content being delivered from the local browser cache. All objects will contain a response code of 200, indicating that all were served directly from the servers as in the illustration below.

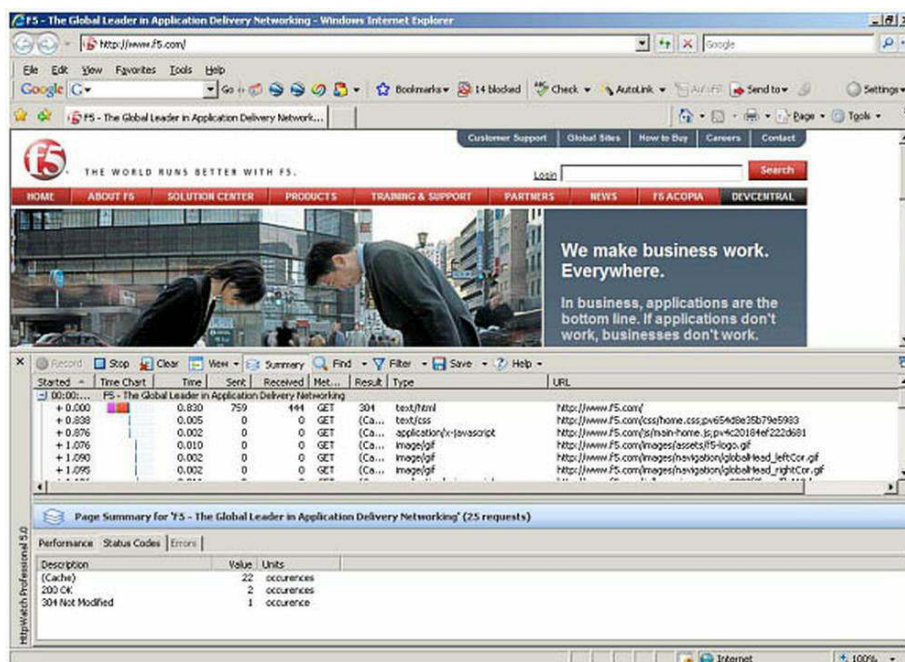


WHITE PAPER

Caching Behavior of Web Browsers

New Browser Session

If a new browser session is started and a user returns to a frequently visited site, the local browser cache will be used (based on the browser settings). If a valid expiration date exists for cached content, it will be delivered directly from the cache and no request will be issued to the origin web server. If content does not have a valid expiration date, the browser will insert an "If-modified-since" or "If-none-match" header into the request. If the content has not changed, then a 304 will be returned from the server and the content will be retrieved from cache. On the other hand, if the content has changed, the server will respond with a 200 and deliver the content to the user.



Recommended Settings

For repeat users BIG-IP® WebAccelerator™ can see great benefits, provided they use the following recommended settings. By using these settings, the user will get the most benefits from the Intelligent Browser Referencing features of WebAccelerator.

Browser Settings

- Automatically/When the page is out of date
- SSL content should be cached.



The Web site

If static content contains random query parameters to prevent caching an iRule can be used to remove these random parameters and enable caching.

As previously stated, using HTTP headers as opposed to META tags is the preferred way to control the cache behavior of an application. The use of META tags will potentially negate the end user benefits of acceleration. META tags can be eliminated through the use of iRules or custom rewrite scripts. With the elimination of META tags, end users will see the benefits of the Intelligent Browser Referencing.

Loading of the Page

To see the differences of the application with and without acceleration, a new browser session must be initiated; the other three ways of loading a page on repeat visits will show no differences with or without acceleration.

Conclusion

Eliminating the need for the browser to download content on repeat visits can greatly improve the performance of web applications. There are many factors that impact whether or not content can or will be retrieved from the local browser cache on repeat visits, including the browser settings, the web site, and the user's behavior. BIG-IP WebAccelerator can improve the utilization of the user's cache without needing to change the application.

F5 Networks, Inc.
401 Elliott Avenue West, Seattle, WA 98119
888-882-4447 f5.com

Americas
info@f5.com

Asia-Pacific
apacinfo@f5.com

Europe/Middle-East/Africa
emeainfo@f5.com

Japan
f5j-info@f5.com