

CSE310
Notes on
Redirection Schemes for Content Distribution on the Web

Samir R. Das

To reduce access latency, a popular method currently used in the Internet is *replication*. The content is replicated (copied) on multiple, geographically distributed servers on the Internet instead of keeping it on a single, designated server (called *origin server*). A client transparently accesses a *content server* close to itself. This not only reduces distance from client to server, but also distributes the load across the Internet. The *selection* of content server for a client relies on sophisticated algorithms and on-line measurements so that load is balanced and access latency is minimized.

A vital component of content distribution architecture is a method for redirecting clients to the content servers. This must be transparent to the client to be universally appealing. There are many methods possible with different user-perceived performance. For example, the origin server can reply the HTTP request from the client with appropriate reply code to redirect the HTTP request itself to an appropriate content server. Or, the origin server can rewrite the URLs in the web content for each incoming request so that the URLs of embedded objects are now different for different clients. The rewritten URLs point to appropriate content servers. However, a different DNS-based redirection method is widely used on the Internet today, promoted by companies such as Akamai. This is described below.

In the most general form of DNS-based redirection, only some objects are replicated; others are not. The non-replicated objects directly come from the original server. For the replicated objects, the hostname in the URL resolves to the IP address of an appropriate content server. In this scheme, the replicated object simply appears to come from another server (say, `http://a.foo.net`) instead of the origin server (say, `http://www.foo.com`). This alternate server (`a.foo.net`) is simply aliased to a hostname (say, `a.mirror.com`) maintained by a third party content distribution company (using the CNAME technique in DNS). Thus, the authoritative DNS server for the company (`mirror.com`) is contacted to resolve `a.mirror.com`. This authoritative DNS server *dynamically selects* the content server based on the IP address of the requesting client and then returns the IP address for this content server. (Multiple IP addresses could be returned as well; the client chooses one of them. DNS allows for this.) The *cache timeout* for this DNS resolution is set very low so that the authoritative DNS server retains the power of dynamic server selection for the most part. The advantage of such DNS-based redirection is that the scheme is largely transparent the original content provider (`foo.com`). They only decide which content needs to be replicated. The rest of the mechanism is handled by the content distribution company that is provisioned as a service.