EFFICIENT ALGORITHM AND PROTOCOL FOR REMOTE DIFFERENTIAL COMPRESSION

Abstract

A method and system are related to updating objects over limited bandwidth networks. Objects are updated between two or more computing devices using remote differential compression (RDC) techniques such that required data transfers are minimized. In one aspect, efficient large object transfers are achieved by recursively applying the RDC algorithm to its own metadata; a single or multiple recursion step(s) may be used in this case to reduce the amount of metadata sent over the network by the RDC algorithm. Objects and/or signature and chunk length lists can be chunked by locating boundaries at dynamically determined locations. A mathematical function evaluates hash values associated within a horizon window relative to potential chunk boundary. The described method and system is useful in a variety of networked applications, such as peer-to-peer replicators, email clients and servers, client-side caching systems, general-purpose copy utilities, database replicators, portals, software update services, file/data synchronization, and others.

Figure 6