Introducing...

Web browsing is too slow. Poor performing servers, network latencies, bottlenecks, sudden surges (flashing). Prone to errors of limited value.

Content Delivery Networks (CDNs) offer more effective solution. Replicate documents, control document consistency, analyze global access patterns, push replication, browsers need one-proxy setup, etc.

However, most CDNs are enterprise based, requiring control over all servers.

Alternative: end-users construct their own CDN using Globule, where all its users' machines (incl. desktop PCs) form the CDN.

- Provide (cheap) local resources;
- Gain access to (valuable) remote resources;
- Content owners remain in control of their documents.

Globule is a module for Apache.

- Easy to upgrade from no-replicated server;

Advantages of Globule replication:

- Accessing nearby replicas increases client perceived performance;
- Multiple replica servers help overloaded sites (scalability);
- Fault tolerance to keep site always available;
- Adaptive replication policies for optimal usage.

Adaptive replication

Documents have different access characteristics.

- Size, read and update frequency, spatial and temporal access;
- No one-size-fits-all approach to replication

Globule does not apply the same policy to all documents:

- Each document is replicated with the policy that suits it best;
- Choice of policy is done automatically.

Globule dynamically changes policies if access patterns change.

Evaluated by computing best performance of each possible strategy over past period, depending on several metrics (e.g., network latency, consistency tolerance, update traffic).

Dynamic document replication

Most large web sites use dynamic document technology (e.g. PHP, CGI, ASP) to generate web pages.

Stored data used by these scripts is being accessed from local databases.

Replication of dynamic documents requires replication of both code and data. In principle straightforward, however:

- Code modifies underlying data thus the system must maintain consistency among replicas;
- Scripts should transparently run on different hosts from which they originate;
- Replica servers should be protected from replicated scripts trying to access local state.

Globule will support incorporating on-demand replication of PHP scripts and application data.

Replica Placement

How do you select the right servers to host your content?

1. Locate where your clients are;
2. Identify hot regions where most clients reside;
3. Pick a server located inside the region;
4. Replicas are strategically placed nearby clients;
   - Minimizing client-to-replica latency;
   - Providing region relevant data distribution.

Internet latencies are modeled as an N-dimensional space using our SCOLE research (see browser redirection).

Replicated objects are stored in Globule’s Content Organiser. They may be replicated by any replica server, subject to the replication policy.

Security

How do you make sure that replica servers deliver your content?

Need mechanisms that verify the integrity from responses of untrusted replica servers.

Place replicas only at servers that you trust.

Arbitrary requests are checked by comparing hash value of content between replica and document owning server. Violation results in removal of the list from active replicas.