Group 17
15CS10030, Nisarg
15CS10005, Animesh
15CS30019, Mayank
15CS30010, Prathamesh

Multimedia CDN

Implementation Plan

Features

- Features provided to End-users
    - Download multimedia content (jpg, mp3, mp4) using content ID
- Features provided to Content providers
    - Upload content for efficient and error free access by end-users
    - Receives billing information
- Features of Content Delivery Network
    - Storage and management of content through Data replication
    - Improved performance through better placement of data
    - Fault tolerance
    - Scalability
    - Performance measurement/ Billing

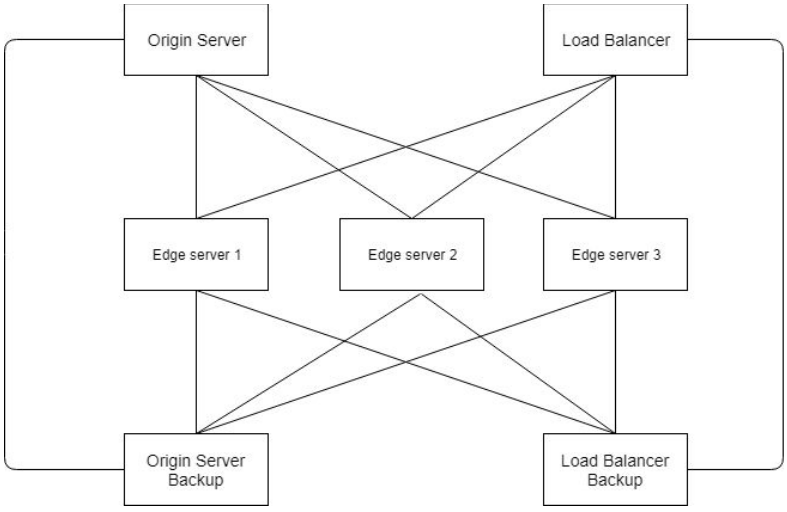
Design

Assumptions

- Reliable topology
- Asynchronous system
- Persistent, Read-only content
- Traffic is read-heavy (Frequency of upload is low)
- Each data item has a unique content_id

Entities involved.

- **Origin server** which holds all the data, (One backup to maintain fault tolerance)
- **Load balancer** to route requests, with a backup Load balancer to maintain fault tolerance
- Multiple **edge servers** to serve content requests
- **End-users** which request content



Request Routing. A client will send its content request, tagged with the content_id, IP address, and (location ID) to the Load balancer. The Load balancer will select an edge-server based on 3 factors - a) Load on server b) Location of server c) Presence of data item in server (In order of priority). The LB will then send a special response (IP+port of the selected server) to the client which will ask the client to resubmit its request to the selected server.

Content Distribution. When there is a cache miss, edge-server will pull the required data from the origin server.

Content Replacement. When there is a cache miss, and the cache is full, the edge server will replace the data using the LRU policy. A map (content_id⇒access time) will be maintained for the same.

Feedback. The edge servers will periodically send feedback to the load balancer to aid request routing. These special feedback messages will incorporate the number of requests served and the details of the data items present in the edge server.

Fault tolerance. (To add)

- **Edge server crash.** In case of an edge-server crash, Load balancer will reroute new or ongoing requests to different servers. After a timeout the end-user will again send a request for the same content to the load balancer. Timeout value will be configured such that within that time frame the load balancer would know if the edge-server has crashed and the current request will be served from different edge-server.
- **Load balancer crash.** To avoid single node crash failure in load balancer, two load balancers are configured in active-standby mode. The active load balancer has a floating IP address assigned to it (in addition to its private IP) which is exposed to the external clients. The load balancers use a heart-beat mechanism to monitor each other. If the load

balancer in the standby mode stops receiving the heart-beat from the active load balancer, it assumes that the active node is down and hence it takes over the floating IP and starts serving the user request

- **Origin server crash.**