**DATA DE-DUPLICATION**

D.Ravali T.Vignashree

B.E-3/4, IT B.E-3/4, IT

[deviravali1997@gmail.com](mailto:deviravali1997@gmail.com) vignashree7@gmail.com

**ABSTRACT:**

Data de-duplication is a technique for eliminating duplicate copies of data, and has been widely used in cloud storage to reduce storage space and upload bandwidth. However, there is only one copy for each file stored in cloud even if such a file is owned by a huge number of users. As a result, de-duplication system improves storage utilization while reducing reliability. Furthermore, the challenge of privacy for sensitive data also arises when they are outsourced by users to cloud. Aiming to address the above security challenges, it makes the first attempt to formalize the notion of distributed reliable de-duplication system. Here, propose new distributed de-duplication systems with higher reliability in which the data chunks are distributed across multiple cloud servers.

The security requirements of data confidentiality and tag consistency are also achieved by introducing a deterministic secret sharing scheme in distributed storage systems, instead of using convergent encryption as in previous de-duplication systems. Security analysis demonstrates that our de-duplication systems are secure in terms of the definitions specified in the proposed security model. As a proof of concept, an implement the proposed systems and demonstrate that the incurred overhead is very limited in realistic environments.