Secure, Archival Storage With POTSHARDS

Mark W. Storer Kevin M. Greenan Ethan L. Miller Kaladhar Voruganti

FAST WIP Session
February 14, 2007

Problem: Long-Term Encryption

- Keys may be lost
  - Effectively short-term data deletion
  - Not long-term deletion

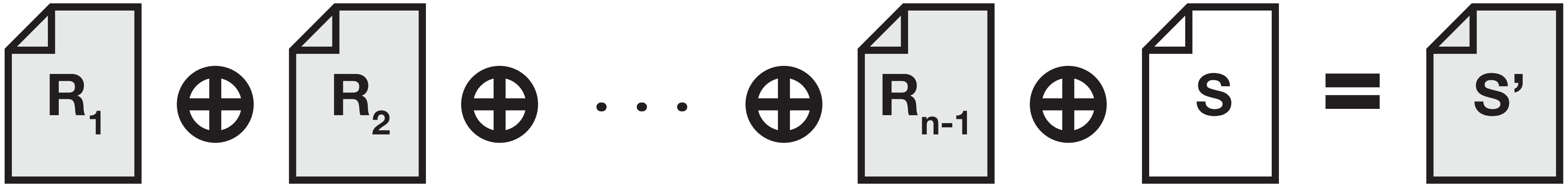
- Keys may be compromised
  - Affects the secrecy of *every* file encrypted with that key

- Even in the best case, is encryption ideal for long-term data?
  - Key management is complicated by long data lifetimes
  - Encryption is only *computationally* secure
    - Hard to predict the future of cryptanalysis
    - Future advances may compromise a *lot* of data in a short time
  - Difficult to re-encrypt petabytes of data

Possible Solution: Secret Splitting

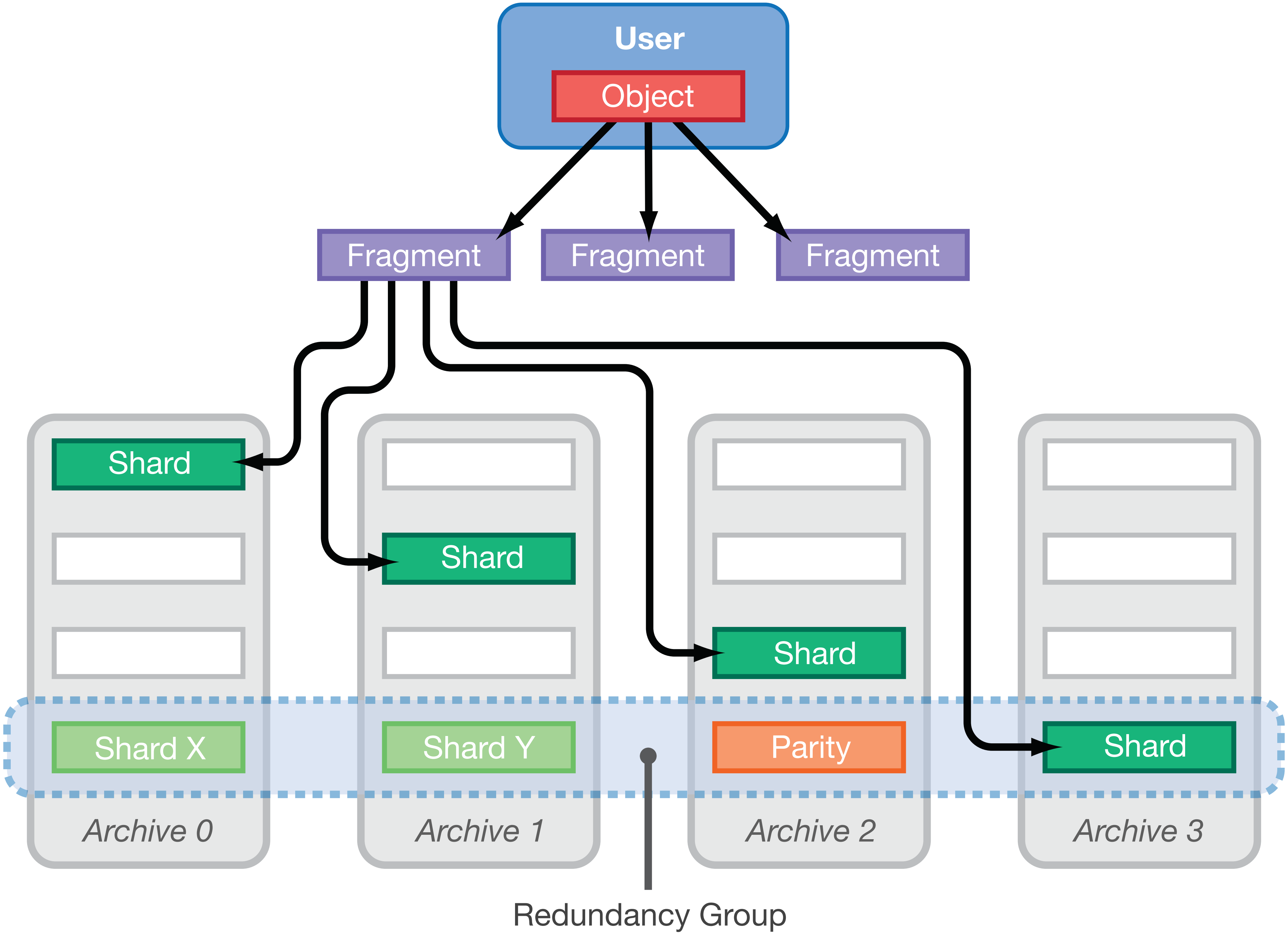
- Create n pieces of data, m of which are required to recovery the original data
  - XOR based (fast by typically n out of n)
  - Linear interpolation based (Shamir)
- Provably secure
  - Any less than m pieces reveals no information
  - Encryption is only computational secure
- Reconstruction not dependent on a single key
  - Encryption keys are a single point of failure
  - No secret share is more important than the others

A simple *n* of *n* secret sharing algorithm using XOR



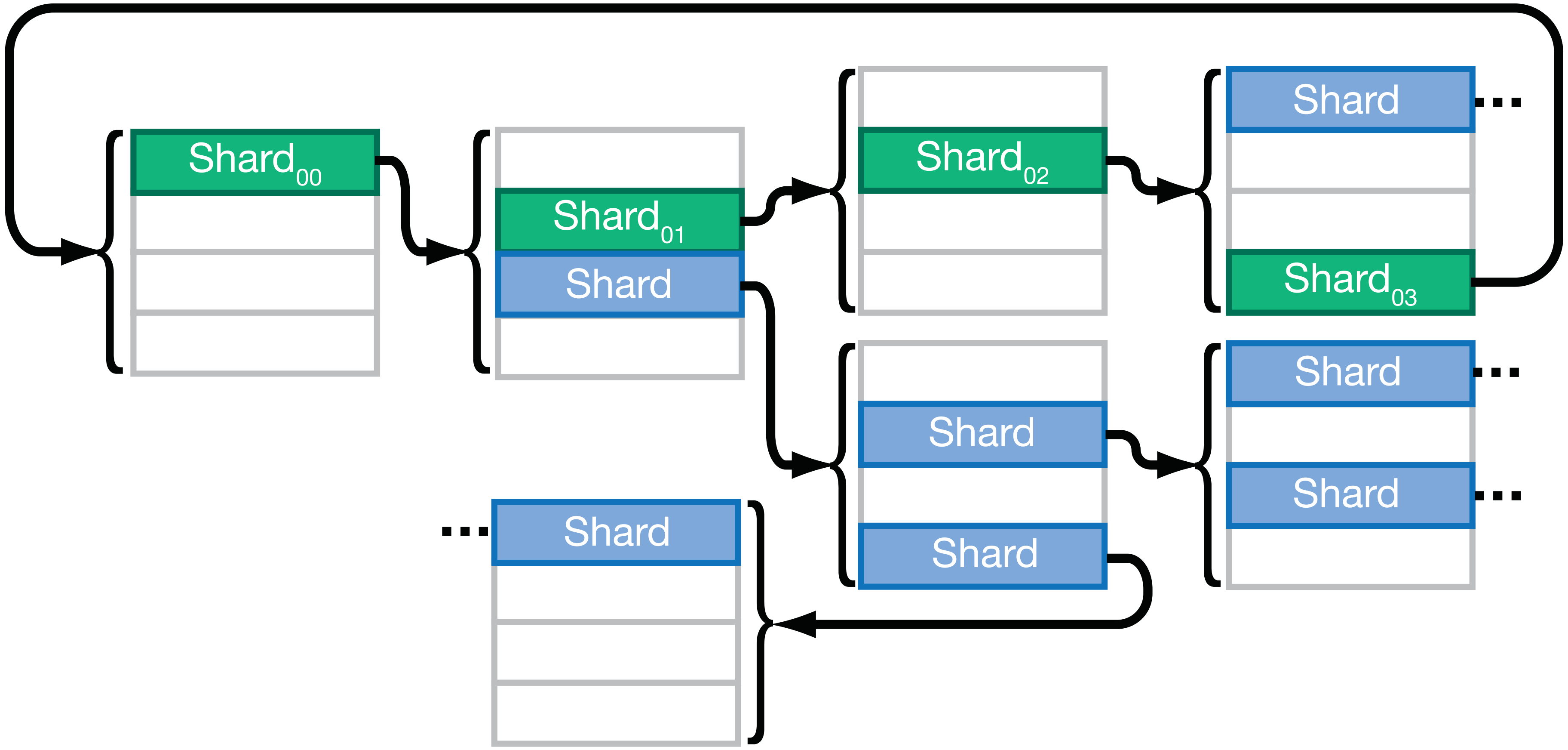
1) Generate *n-1* random pieces of data the same size as ***S*** (the secret to share)
2) XOR the *n-1* random pieces and ***S*** together to form ***S’***
3) Throw away the secret, ***S***, and distribute the *n-1* random pieces and ***S’***

POTSHARDS: Overview



POTSHARDS: Data Recovery

- Need to hide relationship between shards from intruders
- Need to provide sufficient hints to allow reconstruction from just the shards
- Solution: *approximate pointers*
  - Point to a *range* of shards rather than just one
  - No way to verify correctness of a tuple until all the shards have been gathered



Questions

> “The great secret that all old people share is that you really haven't changed in seventy or eighty years. Your body changes, but you don't change at all. And that, of course, causes great confusion.”

http://www.ssrc.ucsc.edu/proj/archive.html

- Thanks to our sponsors:
  - SSRC industrial sponsors
  - Los Alamos National Laboratory /
    Institute for Scalable Scientific Data Management
  - Petascale Data Storage Institute
- Thanks to POTSHARDS team members
  - Kevin M. Greenan
  - Professor Ethan L. Miller
  - Kaladhar Voruganti (IBM Almaden Research Lab)