## Project Name:
Percival – Secure and Searchable Archival Storage

### Summary as of Date:
Dec 2015

### Project Type (Proposed, New or Continuing):
Continuing

### Project Center (UCSC), Center Director:
Ethan L. Miller

### Project Investigators:
Ethan Miller, Darrell Long, Andy Hospodor,

### Affiliates/Collaborators:
Thomas Kroeger, Sandia National Laboratory

### Project Description:
Percival research focuses on developing new features and improvements for secret split datastores consisting of hundreds of millions of shares; enhancements include providing a mechanism for querying the datastore that does not rely on fixed-key encryption, and improved reconstruction algorithms. POTSHARDS’, the precursor to Percival, contributions included no single point of failure due to independent storage sites, minimizing insider threat, and the ability to operate securely while compromised. Percival builds upon these features by pre-indexing shares such in order to enable searching without the need for data reassembly, since doing so reintroduces a single point of failure. Current pre-indexing methods rely on fixed-key encryption; these methods are not well suited for long-term storage due to key management issues as well as introducing a single point of compromise. In addition to pre-indexing, Percival will greatly improve share reconstruction time thereby improving both search performance and reliability.

### Proposed Budget:
$45,000 per year (plus additional and student if funding is available)
Estimated (or Actual) Start Date: Dec 2013

Progress to Date:

Published Percival’s overall design at MSST 2015.

Due to the exponential nature of the reconstruction process, there exists a point at which the quantity of secret-split data causes reconstruction to be combinatorially prohibitive. Currently researching several novel methods that greatly reduce the search space required during reconstruction without compromising security; thereby enabling recovery of the original data, where previously recovery would have been combinatorially prohibitive. Furthermore, these reconstruction aids can be tuned based on the specific properties and needs of the implementing storage system, e.g. quantity of data, number of distributed storage sites, periodicity of reconstruction, and risk of targeted theft.

May 2015: Performed security analysis of Percival and concluded that flaws in original design greatly impacted its validity and restricted its application; ultimately it was discontinued. Completely redesigned Percival from the ground up, building upon the lessons learned from the previous design. Implemented and tested new design using the Digital Corpora. Work is currently being published at MSST 2015.

Nov 2014: Prepared a new test corpora based on Digital Corpora (http://digitalcorpora.org/). Post processing included stemming the files and performing a term frequency – inverse document frequency analysis of the roughly 1 million files. Implemented a base Percival system based on a JErasure frontend connected to a BerkeleyDB backend.

May 2014: Submitted overall design to MSST: “Blinded Searching a Secret Split Archive” Joel C. Frank, Shayna M. Frank, Thomas M. Kroeger, and Ethan L. Miller. Refined the long-term roadmap for the project, as well as investigated adding an access control system to Percival.

2013: Developed the base blinded searching design for Percival. Performed initial testing on real world corpora, which aided in adjusting the Bloom filter parameters to meet our security goals.

Publications (past 6 months)

MSST 2015: “Percival: A Searchable Secret-Split Datastore”
Milestones and Deliverables for Current Year

Months 1–6

- Develop an improved share reconstruction scheme that is a mesh-based variant of POTSHARDS’ approximate pointers
- Submit the improved reconstruction scheme for publication early 2016

Months 7–12

- Add support for additional access control systems (in addition to the currently supported RBAC)
- Improve query server recovery time in the event a salt is compromised.

Related Work within CRSS: N/A

Related Work Elsewhere:

Our work on blinded searching and Percival builds on two areas of related work: secret split archives and searching encrypted data sets while keeping the custodian blinded to the search. Secret-split storage was first developed practically in the PASIS project, and was later adapted for archival storage by Storer et al. in the POTSHARDS system. There has been a significant amount of work done in recent years regarding encrypted searching, but they rely on the inherent security of the encryption method itself since both the data and the search terms use fixed key encryption. Due to the single point of compromise in a fixed key system, as well as long-term key management issues, these solutions are not well suited for long-term storage.

How Our Proposal Is Different:

Our proposal differs from previous secure pre-indexing methods in that Percival does not rely on keys; nor does it rely on the inherent security of the encryption method used. Instead, it pre-indexes each share in the repository with a Bloom filter containing desired metadata ingested via salted hashing, and fills each filter to a desired capacity with random entries. Additionally, the Bloom filters design parameters are intentionally de-tuned to obfuscate their contents. The result is a secure and searchable data store, free from fixed key encryption, that is designed to operate while compromised.

Potential Benefits to CRSS Members: