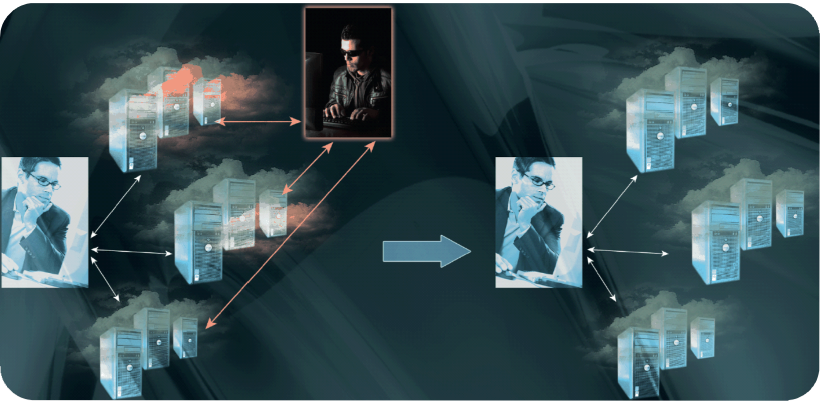
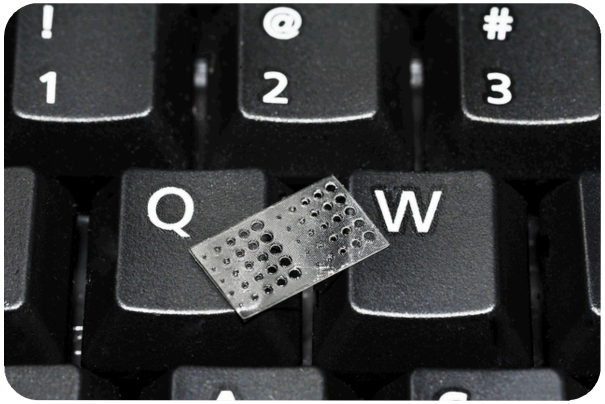
Exceptional service in the national interest

SAND2015-5517PE

Sandia National Laboratories





Security Implications of the Cloud

David Zage

Cyber Analysis R&D Solutions

Department 09526

U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration

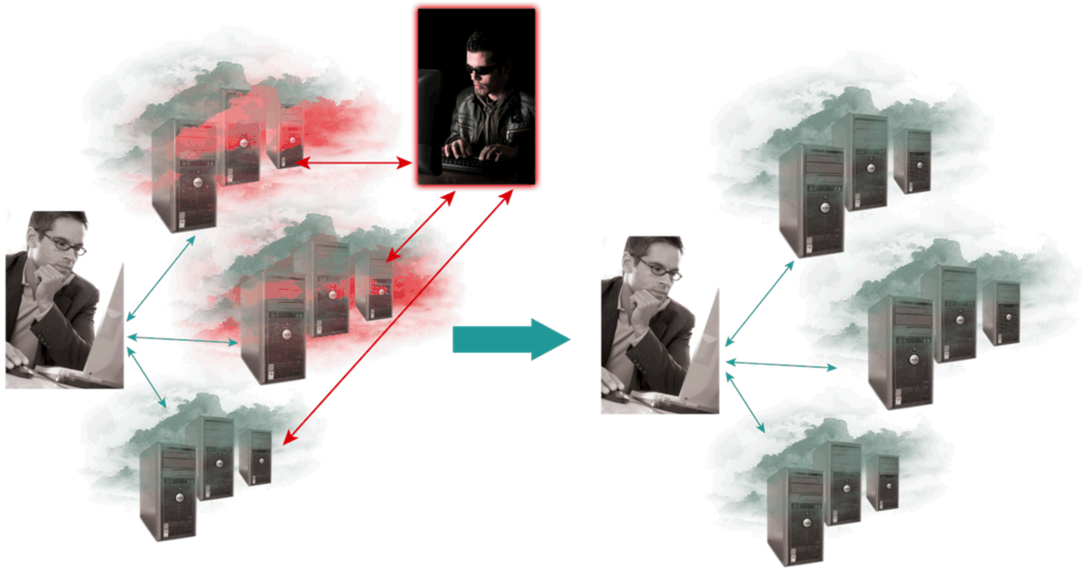
Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

What Is Cloud…and Why Do We Care?

- A model for "enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction"
- Cloud computing is wildly popular
- Security is still an afterthought
  - Most analysis is on ROI for companies not security

Research Goals

- Create secure storage protocols that are robust to malicious insiders that address the identified problems
- Allow for informed trade-offs between efficiency, performance, and security
- Understand implications of using the “cloud”



Storage Using Algebraic Subspaces

- Treat data as Matrix *D*
- Generate a random coding matrix *A*

$$E = A * D$$

- Compute the Singular Value Decomposition (SVD)

$$A = U\,\Sigma\,V^T$$

Example $\mathbb{R}^3 \rightarrow \mathbb{R}^4$:

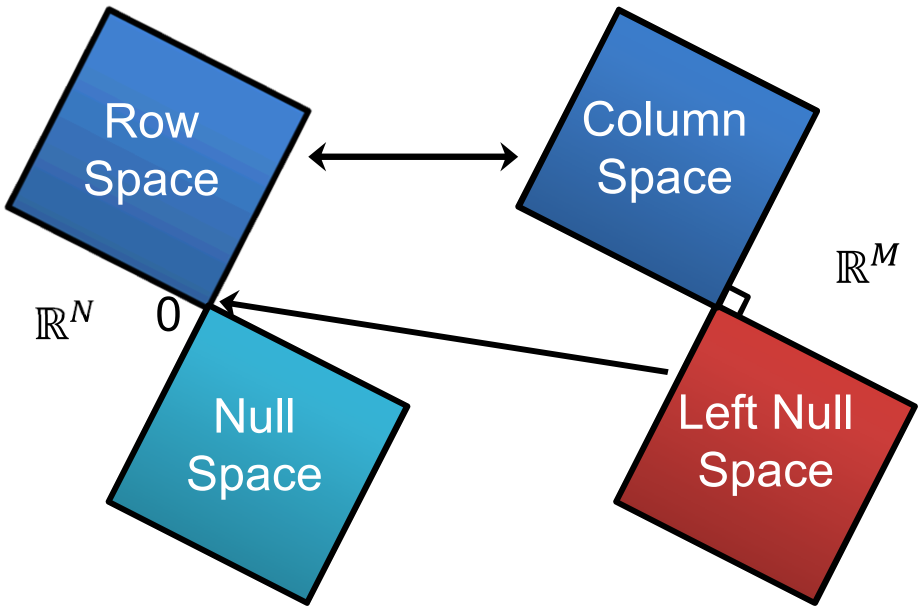
$$\mathrm{A} = \left(\overrightarrow{u_1} \quad \overrightarrow{u_2} \quad \overrightarrow{u_3} \quad \color{red}{\overrightarrow{u_4}}\right)\begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} \overrightarrow{v_1} \\ \overrightarrow{v_2} \\ \overrightarrow{v_3} \end{pmatrix}$$

Storage Using Algebraic Subspaces (2)

- "Append" linear combinations of the left null space ($\mathrm{N}(A^T)$) to $E = A * D$
  - *e.g.* $E = E \mid (c * \overrightarrow{u_4})$
- Data is recovered by multiplying by the (pseudo)inverse

$$D = A^{+} * E$$

$$\begin{pmatrix} d & d & d & 0 \\ d & d & d & 0 \\ d & d & d & 0 \end{pmatrix}$$



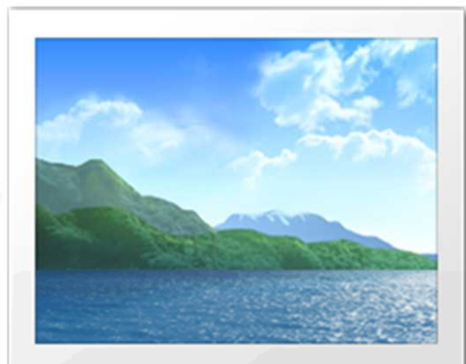
Visual Results

- Before

Encoded

- Decoded

(1)



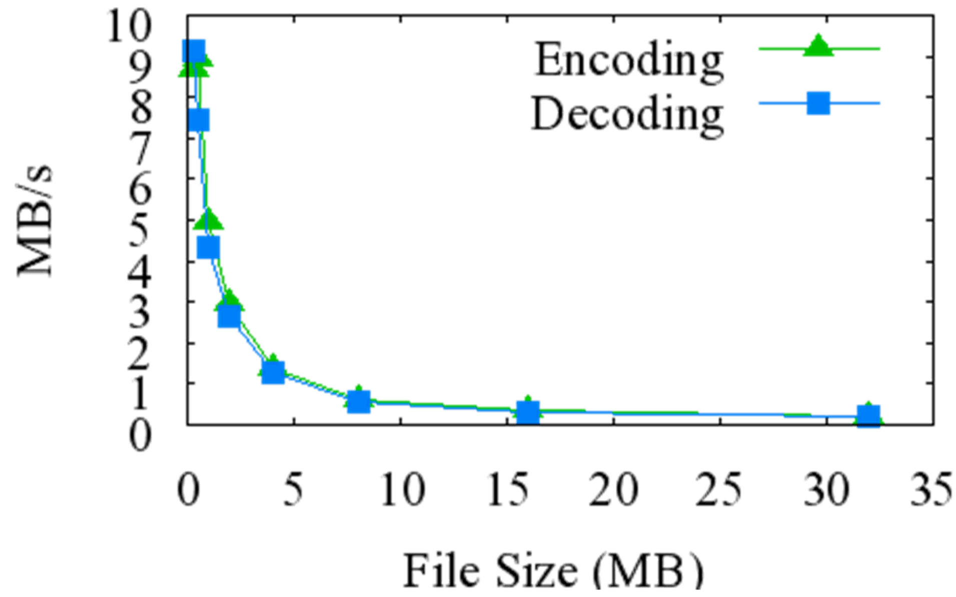
(2)

(3)

The image is not viewable!

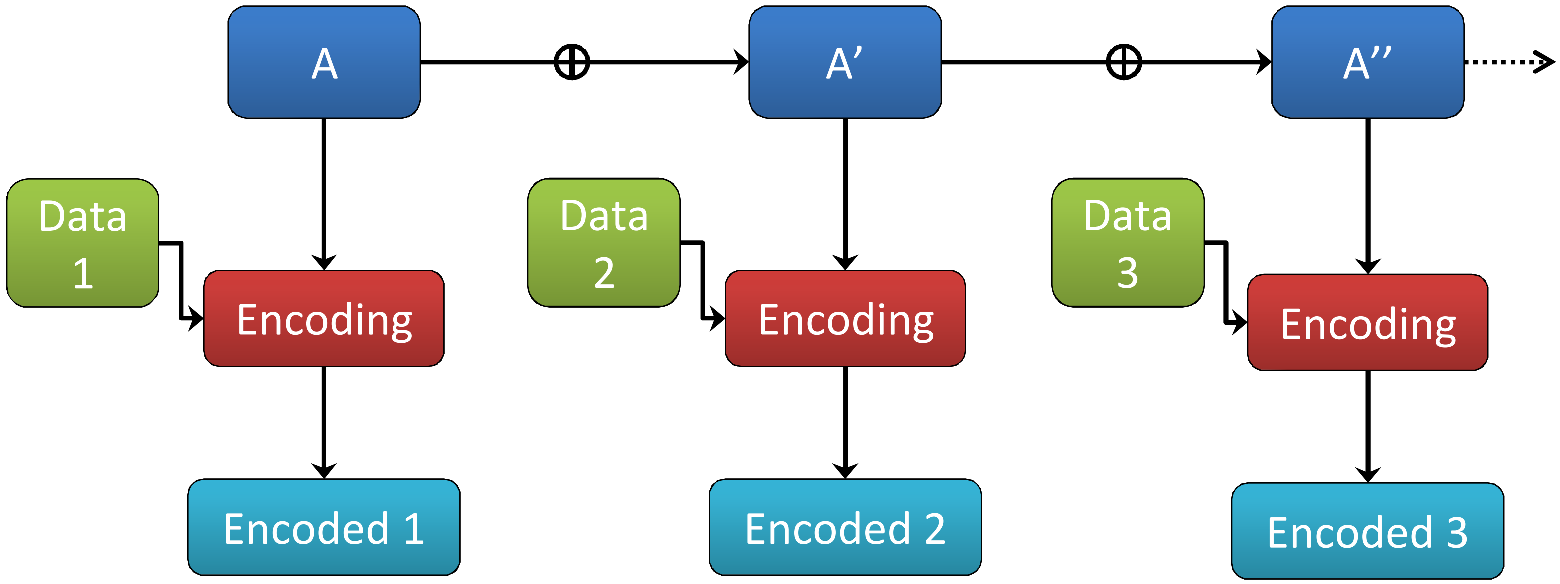
W&C Encoding & Decoding Results

- Encoding and decoding have similar performance
  - No efforts in parallelization
- Performance degrades as the file size increases becomes
  - single, large matrix is expensive to manipulate

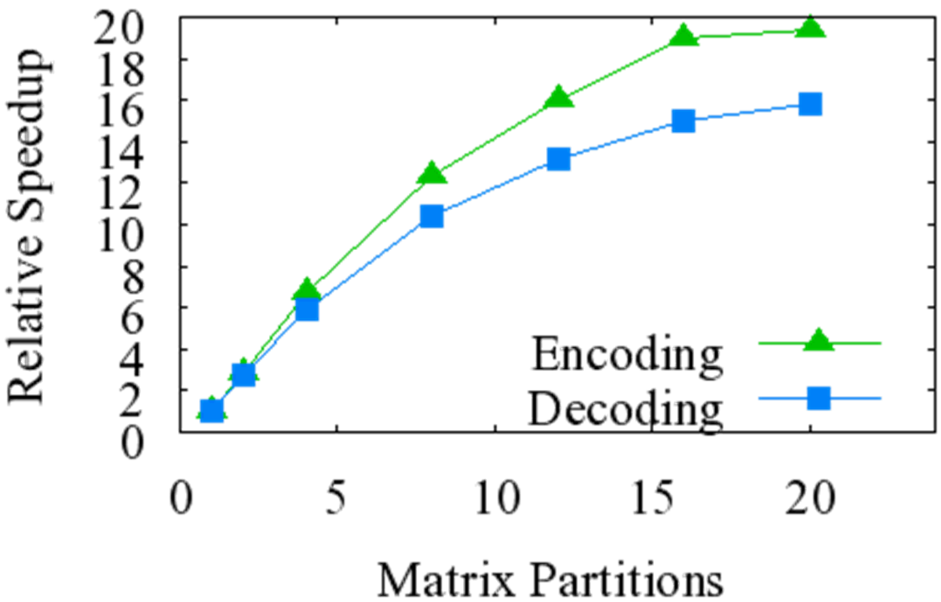


Improving Encoding Performance

- Large amounts of data can be costly to manipulate (time, memory, processing)
- *Matrix Block Chaining (MBC): Intelligently* partition the data into smaller chunks and encode
  - Avoids creating multiple encoding matrices



MBC Results



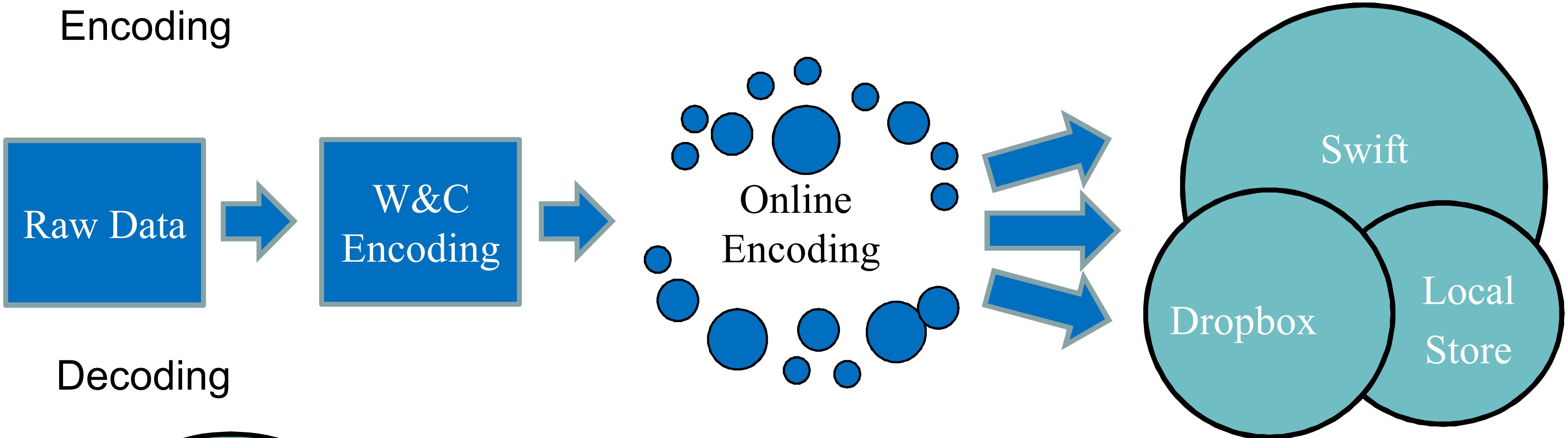
- Increasing the number of partitions can yield a significant performance boost
  - Too few = large matrices
  - Too many = excess processing overhead
- Optimal processing block size is ~2MB

Further Extensions

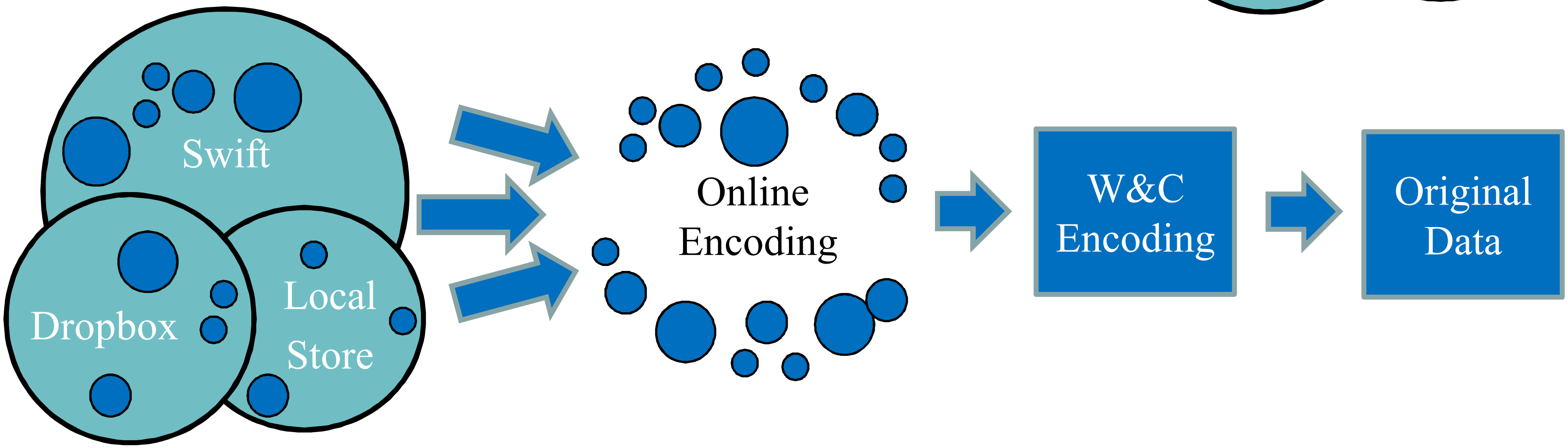
- Mathematical security proofs
- Multi-provider solution and security proofs
- Chaff obfuscation
- Encoding in finite fields
- Efficient parity verification
- Data resiliency via rateless codes

Storage and Retrieval

Encoding

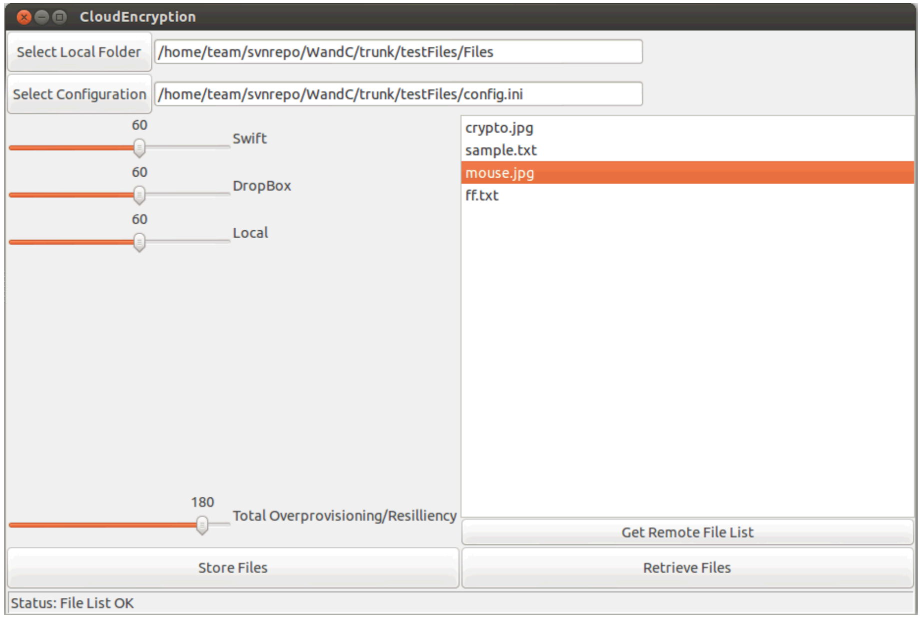


Decoding



Prototype

- GUI provides easy manipulation of available servers and settings
- Supports multiple “cloud” types: Swift, Dropbox, local storage. Etc.



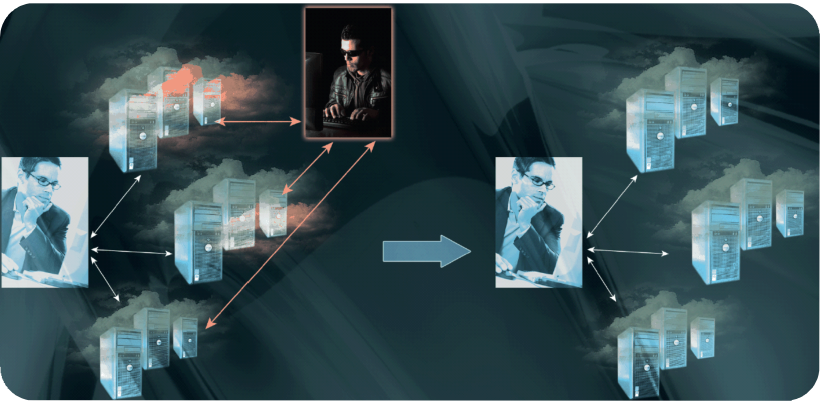
Other cloud-related work

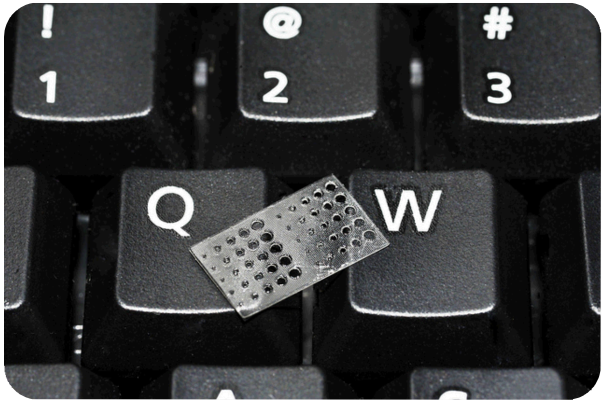
- Cloud security analysis
  - How to choose and vet a cloud provider
- Secure Distributed Set Membership
  - No single point of failure
  - Query data with decryption/reconstruction

BACKUP

Exceptional service in the national interest

Sandia National Laboratories





Protocol Properties

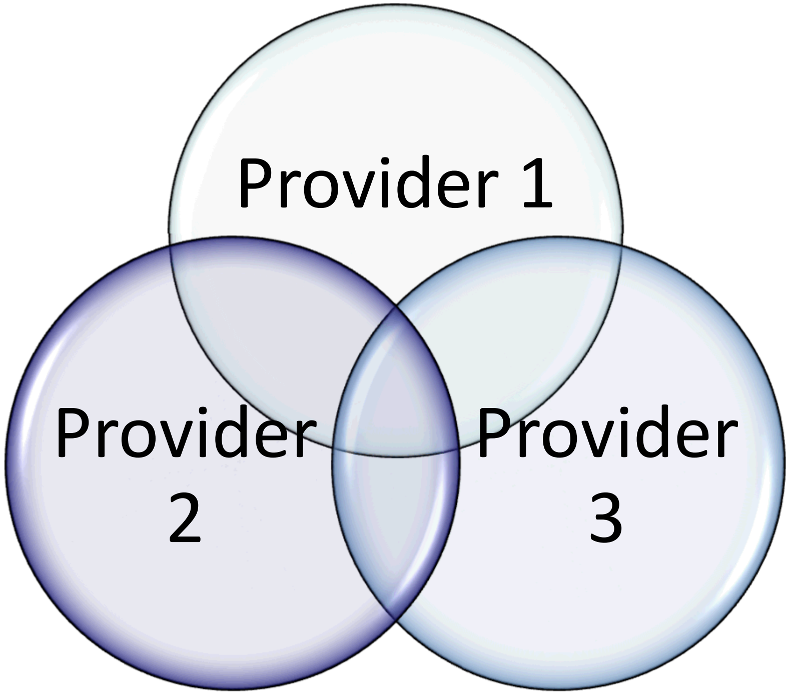
U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Future Avenues to Improve Cloud Security

- Evaluate maximization functions for cost/security
- Examine multiple providers and multiple user accounts to provide heightened privacy for users
- Evaluate efficient metadata structures
- Optimize parallelism and algebraic calculations
- Optimize chaff placement
- Evaluate security in new paradigms (e.g., quantum)



Pictorial example of data stored at three service provider

Multi-Provider Storage Security

Proof Sketch:

Assume an attacker has access n-k rows of E and all of the encoding matrix A

- Create a set of linear equations

$$\begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{k,1} & \cdots & a_{k,n} \end{pmatrix} \begin{pmatrix} d_{1,1} \\ \vdots \\ d_{n,1} \end{pmatrix} = \begin{pmatrix} e^*_{1,1} \\ \vdots \\ e^*_{k,1} \end{pmatrix}$$

- Attacker guesses n-k-1 valid elements of d, and can rewrite the equations with fewer unknowns

$$\begin{pmatrix} a'_{1,1} & \cdots & a'_{1,k+1} \\ \vdots & \ddots & \vdots \\ a'_{k,1} & \cdots & a'_{k,k+1} \end{pmatrix} \begin{pmatrix} d'_{1,1} \\ \vdots \\ d'_{k+1,1} \end{pmatrix} = \begin{pmatrix} e'^*_{1,1} \\ \vdots \\ e'^*_{k,1} \end{pmatrix}$$

Multi-Provider Storage Security (2)

- Attacker performs Gaussian elimination

$$RREF(\mathbf{A'}) = \begin{pmatrix} & s_1 \\ \boldsymbol{I_k} & \vdots \\ & s_k \end{pmatrix}$$

- Attacker is left with more unconstrained variables than linear equations and infinitely many solutions
  - *RREF(A')* is not a one to one mapping

Protocol Overhead

- Computation Complexity
  - ***SVD → $O(4m^2n + 8mn^2 + 9n^3)$ for a m by n matrix***
  - Matrix-matrix multiplication → (≤ $O(n^{2.807})$)
  - matrix-vector multiplication → ($O(n^2)$)
- Storage Overhead

$$SO = \frac{(\lceil(\lceil\sqrt{(\text{Fsize}/\text{Parts})}\rceil \times (1 + \text{Chaff}))\rceil)^2}{\text{Fsize}/\text{Parts}}$$

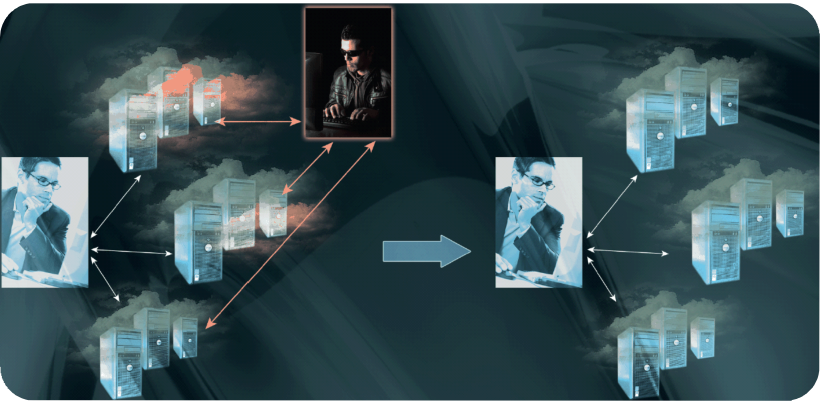
  - Example: MBC breaks a 1MB file into four partitions, the storage overhead is 4.3% greater than that of the unencoded 1MB file for 2% chaff

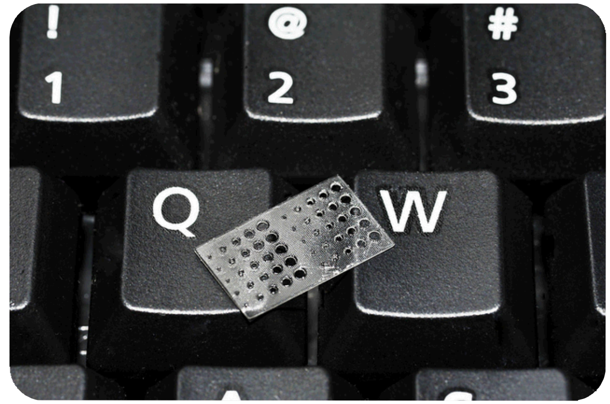
Efficient Chaff Verification

- Chaff is derived from the left null space of the encoding matrix
  - Chaff is orthogonal to vectors derived from column space
- Subspaces of the column space can be used to check the chaff present in a file

$$\boldsymbol{E}^{*\boldsymbol{T}} \left(\sum_{x=1}^{n} (\boldsymbol{u_x} \times r_x) \right) = (v_1, \ldots, v_m)^T$$

- Any resulting zeros in the multiplication of the encoded data by the column vector correspond to columns of chaff





Protocol Extensions

Encoding in Finite Fields

- Problem: Floating point arithmetic has rounding error(s)
  - Limits storage capability
  - Can cause issues with data reconstruction
- Solution: Perform calculations over finite fields
  - Exact calculations over large, prime fields (no rounding error)
  - Changes security proofs, but does not weaken the security of the system

Parity Through Chaff Scaling

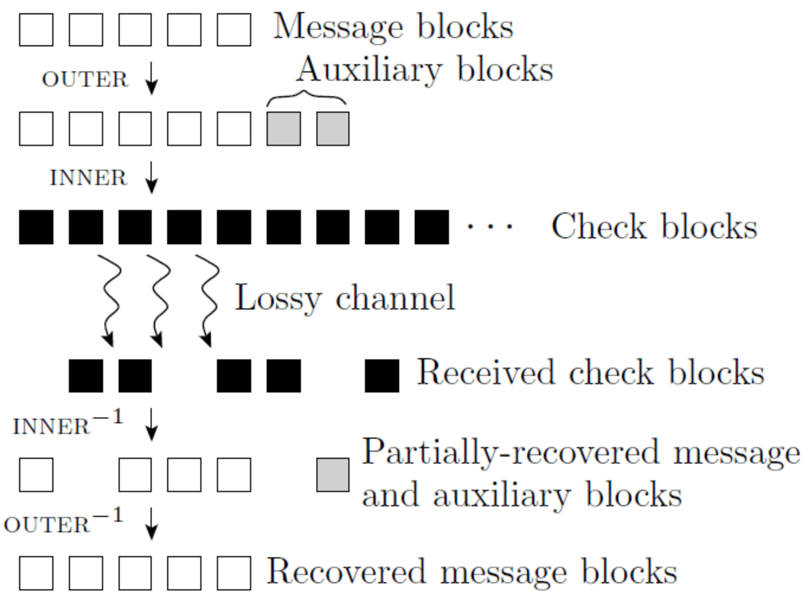
- Problem: How do we achieve fine-grained data integrity?
  - Currently, the solution determines if the file is correct or not
- Solution: Use chaff columns as data integrity checks
  - Chaff can be freely scaled with no adverse effect on decoding
  - Each column acts as the checksum for preceding data columns

Parity Through Chaff Scaling (2)

- Since each chaff column is in the left null space, they can be independently scaled without affecting recovery
- Scale each of the m columns of chaff by a parity value calculated from n/m columns of original data E
- Parity can be through xor, summation, multiplication, etc.
- After data is recovered, finding correct parity values effectively guarantees correctness in associated data columns

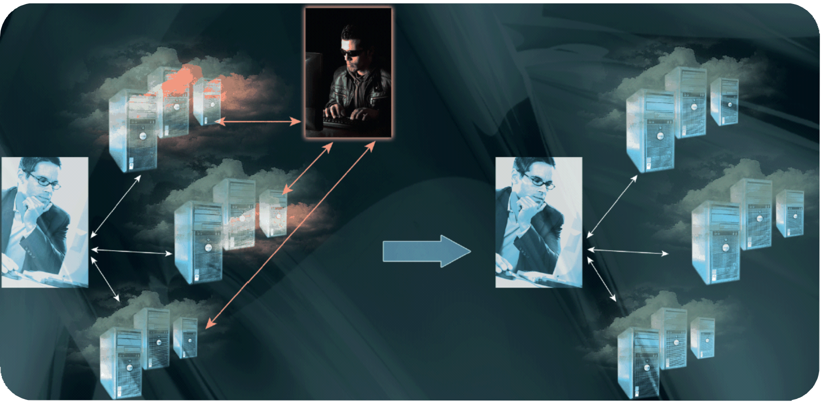
Resiliency via Rateless Coding

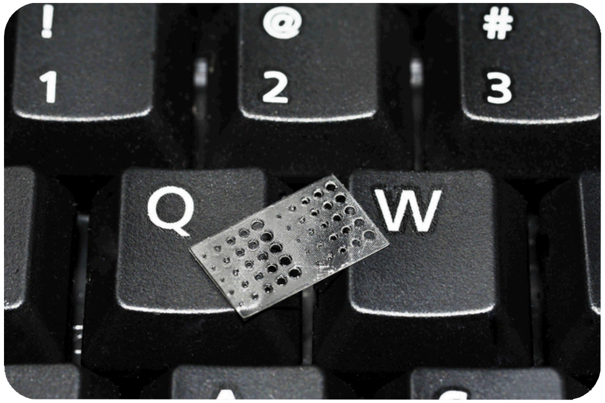
- Problem: How do we get data back when there are network issues?
- Solution: Leverage rateless, locally-encodable codes
  - each file of size has practically infinite encoding possibility
  - encoding blocks can be computed quickly and independently of othe blocks
- Allows for reconstruction if networking or providers fail
- Allows for informed trade-offs between cost, resiliency, and information security
- Initial testing is being conducted



Exceptional service in the national interest

Sandia National Laboratories





Secure Distributed Set Membership

U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Secure Distributed Set Membership

- **Problem:** Long-term Encryption is fragile
- **Solution:** Secret Sharing for Provably Secure Systems
- Archives that distribute data with secret sharing can provide information theoretic data protections and a resilience to:
  - malicious insiders,
  - compromised systems, and
  - untrusted components.
- We are developing ways to functionally use secret shares without reassembly