SAND2013-7600P

Percival: Secure Data Archival Using Secret Sharing

Thomas M Kroeger, Joel C Frank, Ethan L Miller

Sandia National Laboratories and University of California, Santa Cruz

Sandia National Laboratories

Threats to Long Term Data Archives

- **Insider Threat:** *Insiders can be put into one of three categories:*
  1. *Users*--those with legitimate operational need for the data
  2. *System administrators*--system administrators with digital and usually physical access
  3. *Facilities personnel*--those with physical access (e.g. janitorial staff, security guards etc...)

Resilient to 2 & 3 and new capabilities to mitigate abuses by legitimate users.

- **Compromised Repositories:** The rate of security events these days necessitates that any wise system designer plan for parts of their system being compromised.
- **Central Point for Attack:** The central design of most archives provides a clear, single target for attackers.

Secure Foundation: Shamir Secret Sharing

Shamir's Secret Sharing provides a data protection that goes far beyond just splitting the data into parts. It is **information theoretically secure** and uses the fact that one can uniquely define a polynomial curve of order *O* with *O+1* points.

Example: Any 3 of 5 Secret $S = 1234$

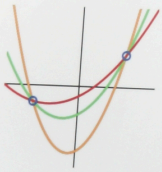
1. Create a polynomial of degree 2 by generating 2 random coefficients, e.g

$$f(x) = S + 166x + 94x^2$$

2. Generate 5 points along that curve:
   (1,1494); (2,1942); (3,2578);
   (4, 3402); (5,4414)

Any three enable me to solve for S.

All I know is S is a value on the Y-axis.
If I steal two points, I don't learn anything new.



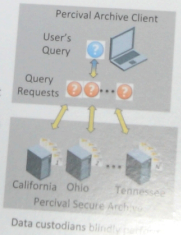
One can draw an infinite number of polynomials of degree 2 through 2 points.

Searching a Percival Archive

- Data is vulnerable when on the client.
- Create ability to operate on the data without needing to reassemble the shares.

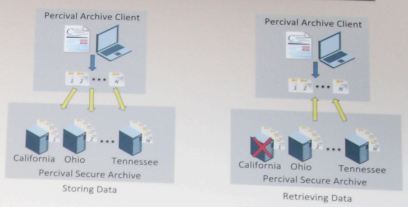
Keyword Search

- Recent research has explored protocols that enable keyword search across our archive.
- We use pre-indexing, key hashes, and bloom filters to provide a blinded search.
- The custodians do the searching but are blinded to the terms searched and the results of the query.
- Initial results with data from the Guttenberg project and Wikipedia have shown promise.



Data custodians blindly [illegible] the search

Percival: Using Secret Sharing for Resilient Data Archive



- Splits a 1K document into *N* shares of 1K that are stored by data custodians creating a distributed archive.
- To retrieve a document, the client gets and combines *T* shares.

- **Resilient to insider threat**. *T* different sites would have to collude.
- **Resilient to component compromise**. Able to operate with integrity and data protection in the presence of *T-1* compromised sites.
- **Resilient to crypto analytical attack**. *T-1* shares are information theoretic in their protections and are not susceptible to brute force decryption attacks.
- **Keyless**. Issues such as key rotation and revocation are moved to authentication.
- **Secure backup**. Beyond the data privacy protections, this system also provides a platform that, by design, securely protects from data loss due to equipment failure.

Potential Uses of Percival

Percival's architecture enables a broad range of capabilities for archive designers:

NNSA Data Archive

- Repositories at 9 sites, any 5 of which can be used to recover data (*N=9, T=5*).
- Secure operations with up to 4 sites compromised.
- Sites disconnecting from the internet because of internal compromise would not impact the archive's secure operations or data integrity.
- For OUO Archives, users could use a keychain for one sign-on to many repositories. For higher level data, multi-person access requirements could provide greater security.

Agency's Credit Card Data

- Many agencies (e.g. FEMA) have a large set of credit card records.
- All must be kept but most are never examined.
- Use Percival to split data into three shares (*N=3*).
- Outsource data repositories to 3 different cloud providers.
- Require two people to recover any credit card records; this inhibits insider abuse of credit card data. *(T=2)*

Percival Directions

Applied Research – Prototype development & Testing

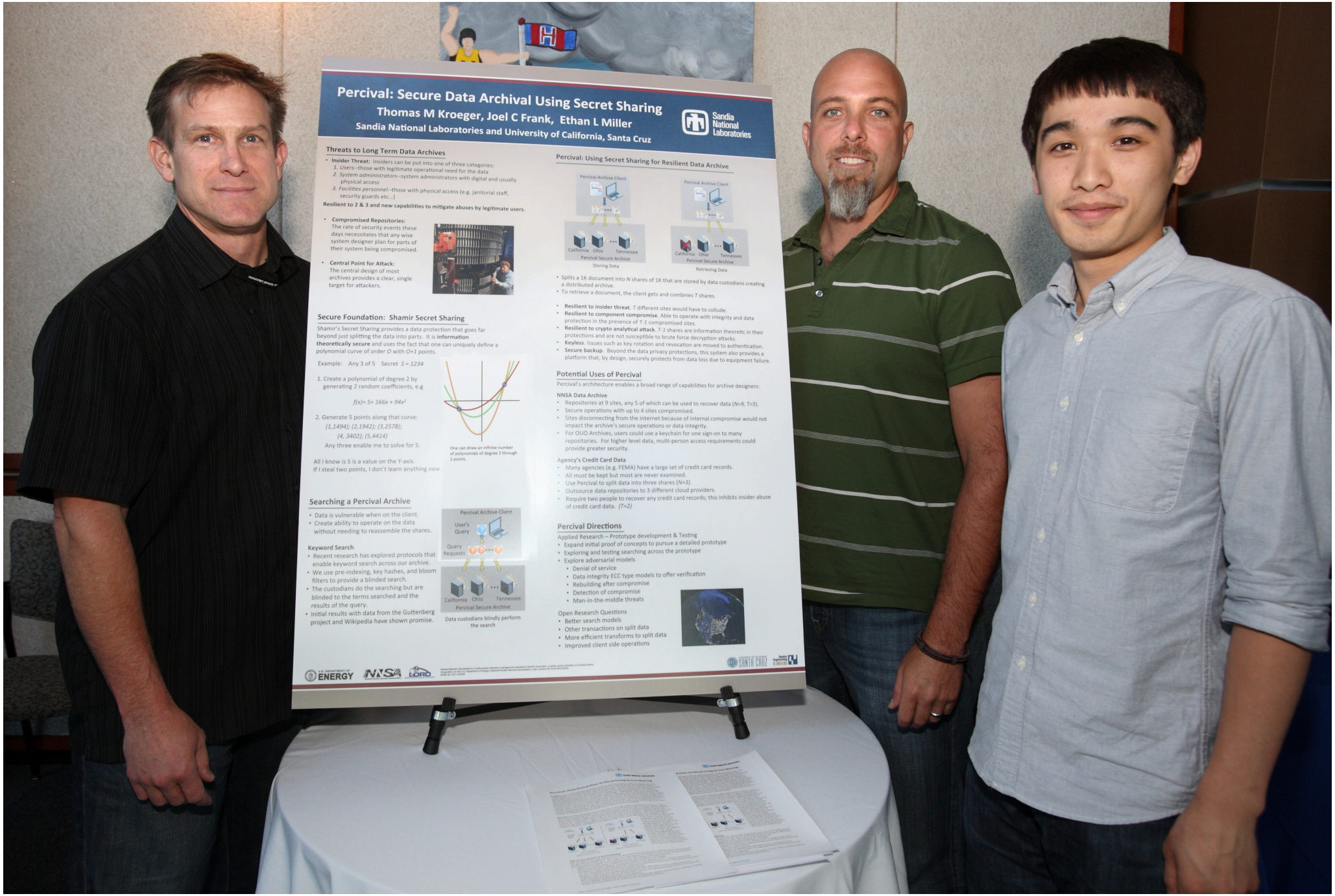
- Expand initial proof of concepts to pursue a detailed prototype
- Exploring and testing searching across the prototype
- Explore adversarial models
  - Denial of service
  - Data integrity ECC type models to offer verification
  - Rebuilding after compromise
  - Detection of compromise
  - Man-in-the-middle threats

Open Research Questions

- Better search models
- Other transactions on split data
- More efficient transforms to split data
- Improved client side operations

U.S. DEPARTMENT OF ENERGY · NNSA · DOD

UNIVERSITY OF CALIFORNIA SANTA CRUZ · Baskin Engineering


Percival: Secure Data Archival Using Secret Sharing
Thomas M Kroeger, Joel C Frank, Ethan L Miller
Sandia National Laboratories and University of California, Santa Cruz
Sandia National Laboratories
Threats to Long Term Data Archives
Insider Threat: Insiders can be put into one of three categories:
1. Users--those with legitimate operational need for the data
2. System administrators--system administrators with digital and usually physical access
3. Facilities personnel--those with physical access (e.g. janitorial staff, security guards etc...)
Resilient to 2 & 3 and new capabilities to mitigate abuses by legitimate users.
Compromised Repositories: The rate of security events these days necessitates that any wise system designer plan for parts of their system being compromised.
Central Point for Attack: The central design of most archives provides a clear, single target for attackers.
Secure Foundation: Shamir Secret Sharing
Shamir's Secret Sharing provides a data protection that goes far beyond just splitting the data into parts. It is information theoretically secure and uses the fact that one can uniquely define a polynomial curve of order O with O+1 points.
Example: Any 3 of 5 Secret S = 1234
1. Create a polynomial of degree 2 by generating 2 random coefficients, e.g
f(x)= S+ 166x + 94x²
2. Generate 5 points along that curve:
(1,1494); (2,1942); (3,2578);
(4, 3402); (5,4414)
Any three enable me to solve for S.
All I know is S is a value on the Y-axis.
If I steal two points, I don't learn anything new.
One can draw an infinite number of polynomials of degree 2 through 2 points.
Searching a Percival Archive
Data is vulnerable when on the client.
Create ability to operate on the data without needing to reassemble the shares.
Keyword Search
Recent research has explored protocols that enable keyword search across our archive.
We use pre-indexing, key hashes, and bloom filters to provide a blinded search.
The custodians do the searching but are blinded to the terms searched and the results of the query.
Initial results with data from the Guttenberg project and Wikipedia have shown promise.
Percival Archive Client
User's Query
Query Requests
California Ohio Tennessee
Percival Secure Archive
Data custodians blindly perform the search
Percival: Using Secret Sharing for Resilient Data Archive
Percival Archive Client
California Ohio Tennessee
Percival Secure Archive
Storing Data
Retrieving Data
Splits a 1K document into N shares of 1K that are stored by data custodians creating a distributed archive.
To retrieve a document, the client gets and combines T shares.
Resilient to insider threat. T different sites would have to collude.
Resilient to component compromise. Able to operate with integrity and data protection in the presence of T-1 compromised sites.
Resilient to crypto analytical attack. T-1 shares are information theoretic in their protections and are not susceptible to brute force decryption attacks.
Keyless. Issues such as key rotation and revocation are moved to authentication.
Secure backup. Beyond the data privacy protections, this system also provides a platform that, by design, securely protects from data loss due to equipment failure.
Potential Uses of Percival
Percival's architecture enables a broad range of capabilities for archive designers:
NNSA Data Archive
Repositories at 9 sites, any 5 of which can be used to recover data (N=9, T=5).
Secure operations with up to 4 sites compromised.
Sites disconnecting from the internet because of internal compromise would not impact the archive's secure operations or data integrity.
For OUO Archives, users could use a keychain for one sign-on to many repositories. For higher level data, multi-person access requirements could provide greater security.
Agency's Credit Card Data
Many agencies (e.g. FEMA) have a large set of credit card records.
All must be kept but most are never examined.
Use Percival to split data into three shares (N=3).
Outsource data repositories to 3 different cloud providers.
Require two people to recover any credit card records; this inhibits insider abuse of credit card data. (T=2)
Percival Directions
Applied Research – Prototype development & Testing
Expand initial proof of concepts to pursue a detailed prototype
Exploring and testing searching across the prototype
Explore adversarial models
Denial of service
Data integrity ECC type models to offer verification
Rebuilding after compromise
Detection of compromise
Man-in-the-middle threats
Open Research Questions
Better search models
Other transactions on split data
More efficient transforms to split data
Improved client side operations
U.S. DEPARTMENT OF ENERGY
NNSA
LLNL
UC SANTA CRUZ