SAND2015-2912D

Secure Distributed Set Membership through Secret Sharing

Sandia National Laboratories
T. M. Kroeger, D. J. Zage and C. A. Phillips
Livermore, California & Albuquerque, New Mexico

University of New Mexico
J. Saia
Albuquerque, New Mexico

Tufts University
T. R. Benson
Medford, Massachusetts

Problem

Data Security and Availability Conflict

Data security and availability for operational use are frequently seen as fundamentally opposing forces.

Encryption is fragile:
Tools like homomorphic encryption are a start but most encryption algorithms are at best assumed to be secure, and often in reality just delayed release.

Secret Sharing Provides Provably Secure Systems
Archives that distribute data with secret sharing can provide information theoretic data protections and a resilience to:

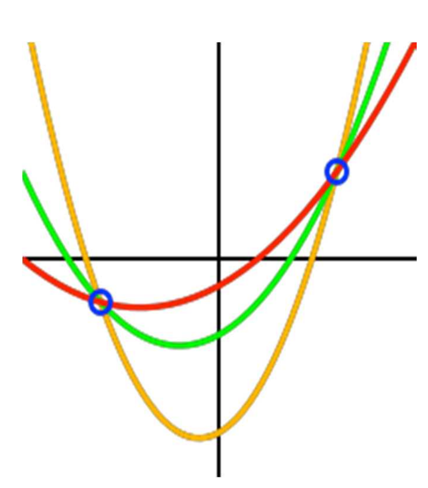
1. malicious insiders,
2. compromised systems, and
3. untrusted components.

We are developing ways to functionally use secret shares without reassembly.

Approach

Secure Foundation: Shamir Secret Splitting

Shamir's Secret Sharing (SSS) provides a data protection that goes far beyond just splitting the data into parts. It is **information theoretically secure** and uses points on a polynomial curve to securely encode sensitive data.



An infinite number of polynomials of degree 2 exist through 2 points.

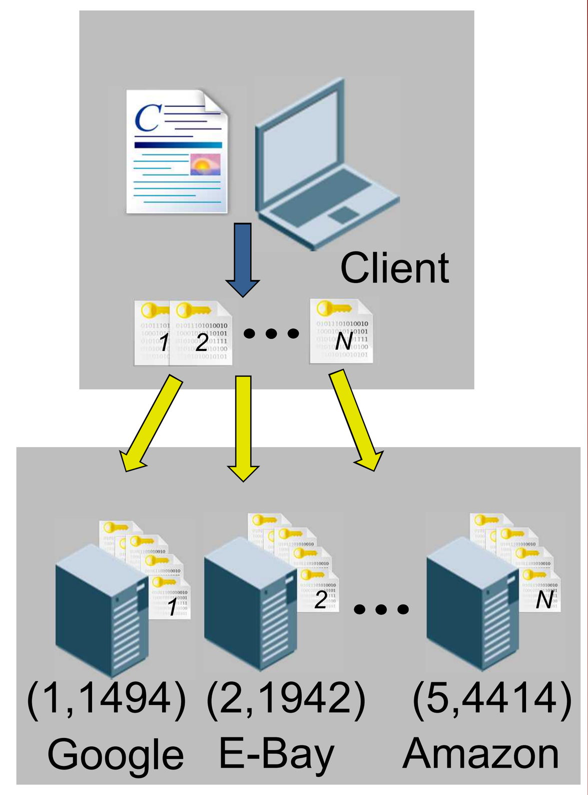
Example:
Any 3 of 5 Secret *S = 1234*

1. Create a polynomial of degree 2 by generating 2 random coefficients
   $f(x) = S + 166x + 94x^2$

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

Any three points enables a user to solve for S. With one or two points you know nothing more than when you had none; S is one of infinite possible Y-intercepts.

Distributed Secure Archive



Results

Serial Interpolation Filter (SIF)

By using Lagrange Interpolation serially across the archive, we can query our data ***without*** exposing the data.

Lagrange Interpolation allows for the recovery of the original function via the stored points:

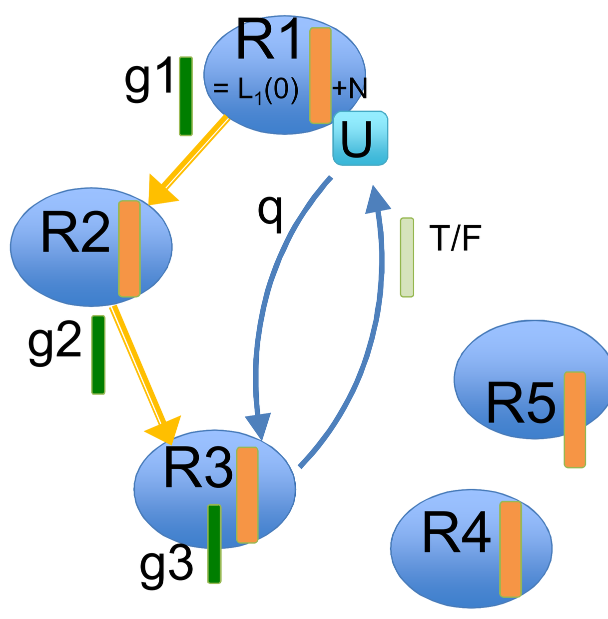
$$f(x) = \sum_{i=1}^{n} L_i(x) f(i)$$

where $L_i(x)$ is the Lagrange basis polynomial and f(i) is the function evaluated at i. Reconstructed, f(0) = S.

Example:
5 companies are willing to share their list of sensitive security data (e.g., list of known bad address). They encode each entry with a 2nd order polynomial. Now we can use a SIF to determine if a given address is in the list.

User U, who works at company *R1,* does the following to test if an address *Z* is on the list.

1) U creates random private random number called a nonce N.
2) U & R1 calculate $g1 = L_1(x)f(1) + N$ and send g1 to R2
3) R2 calculates $g2 = L_2(x)f(2) + g1$ sends g2 to R3
4) R3 calculates $g3 = L_3(x)f(3) + g2$
5) U sends q=Z+N to R3
6) R3 For all entries where g3==q
   send TRUE to U
   else
   send FALSE to U



Significance

With data breaches plaguing both governement and corporate America, the need for better data security architectures is painfully obvious.

Our approach:

- provides information theoretic protections,
- avoids difficult key management, and
- enables interacting with the data without reassembly, preserving the protections of SSS.

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.

SAND 2010-xxxP