Multi-Party Computation in Health Care

Parul Singh, Red Hat
Mayank Varia, Boston University
Secure Medical Image Processing

- At Red Hat we are involved in several research projects at universities and industries
  - Boston University & Boston Children’s Hospital
- **ChRIS Research Integration System (ChRIS)**
  - Web-based medical image processing platform with quicker feedback
  - Democratize Medical Image Processing application development
Medical Image Processing in ChRIS - Input
Medical Image Processing in ChRIS - Output
Conclave Cloud Dataverse (C2D)

SAIL
Software & Application
Innovation Lab

BU
Red Hat
MOC
Typical Workflow: Centralize Data Storage & Analysis

- Typically, transfer data to a common location
- All data contributors must trust the organization that performs the analysis
- Data may be vulnerable in transit, when stored, and during analysis
C2D Workflow: Compute Without Sharing Data

Company A

Company B

Data

Data

Secure and private computation

Results
C2D Workflow: Federate Data Storage and Analysis

- Data never leaves any organization in the clear
- MPC protects data in use during the analysis
- Inputs stay private as long as at least one computing entity can be trusted to behave as specified
• New analyses can be privacy-preserving because data not shared prematurely

• Without MPC, would need data transfer, so protections applied “too early” in pipeline
Selected MPC deployments

**Cybernetica**: VAT tax audits

**BU**: Pay equity in Boston

**Partisia**: Rate credit of farmers

**Google**: Federated machine learning

Employers agree to ... contribute data to a report compiled by a third party on the Compact’s success to date. Employer-level data would not be identified in the report.
MPC at the scale of a city

**COMPENSATION:**
The BLS samples approximately 120,000 workers each year to calculate average annual compensation. The average annual compensation of all workers was $59,062 in 2019. This includes wages, salaries, and other forms of compensation. The BLS also calculates the average annual compensation for workers in different industries and occupations.

**Average Annual Compensation by Gender**
- **Women:** $57,327
- **Men:** $61,880

**Average annual compensation by occupation:**
- **Women:** $46,890
- **Men:** $54,240

**TOTAL COMPENSATION (INCLUDING CASH BONUS OR PERFORMANCE PAY)**
The sample included data on performance pay or bonuses paid to employees. The average total compensation was $78,766 for men and $60,583 for women.

**ANNUAL COMPENSATION**
- **Men:** $75,786
- **Women:** $50,583

The similar usage gap also varied by job category. Among workers in the largest occupation categories, the difference in average annual earnings was $1,080 between women and men. This gap is larger for workers in the largest occupations, such as professionals and managers, who earn more than $100,000 on average.

**Earnings ratios by race, compared to white men:**
- **Women:** 0.75
- **Men:** 1.00

Consistent with other surveys, the gender wage gap varied by race. While white women earned 75% of the median wage of white men, black women earned 61%, and Hispanic women earned 57%.

The largest occupations were among Black/Alaskan Native women, earning 82% of the median wage of white men, and Asian women, earning 80% of the median wage of white men.

**Average annual earnings by occupation:**
- **Women:** $50,583
- **Men:** $75,786

The largest occupations were among Black/Alaskan Native men, earning 108% of the median annual earnings of white men, and Asian men, earning 106% of the median annual earnings of white men.

**Average annual earnings by industry:**
- **Women:** $50,583
- **Men:** $75,786

The largest occupations were among Black/Alaskan Native men, earning 108% of the median annual earnings of white men, and Asian men, earning 106% of the median annual earnings of white men.

**Average annual earnings by education level:**
- **Women:** $50,583
- **Men:** $75,786

The largest occupations were among Black/Alaskan Native men, earning 108% of the median annual earnings of white men, and Asian men, earning 106% of the median annual earnings of white men.
MPC gets...

Security from isolation

Performance from co-location

Scalability from ???
Conclave: Automating Secure Computation

**SQL-like programming language**
⇒ *No MPC experience necessary*

**Dispatcher** executes jobs on available backends
⇒ *No new infrastructure*

**Compiler** discerns boundaries of secure computing
⇒ *No need for privacy experts*

Software available at github.com/multiparty/conclave
Conclave’s Query Specification

# state where the data lives
data = cc.defineTable(schema, at=[org-1, org-2, org-3])

# compute over the data as if it resided in one place
rev = data.project(["companyID", "price"])
    .sum("local_rev", group=["companyID"], over="price")
    .project([0, "local_rev"])
market_size = rev.sum("total_rev", over="local_rev")
share = rev.join(market_size, left=["companyID"],
    right=["companyID"])
    .divide("m_share", "local_rev", by="total_rev")
hhi = share.multiply(share, "ms_squared", "m_share")
    .sum("hhi", on="ms_squared")
    .divide("hhi", by=10k)
Conclave’s Static Analysis: Calculate Relations in the Clear

# state where the data lives
data = cc.defineTable(schema, at=[partyA, partyB, partyC])

# compute over the data as if it resided in one place
rev = data.project(["companyID", "price"])
   .sum("local_rev", group=[“companyID"], over="price")
   .project([0, "local_rev"])
market_size = rev.sum("total_rev", over="local_rev")
share = rev.join(market_size, left=[“companyID"],
                right=[“companyID"])
   .divide(“m_share", "local_rev", by="total_rev")
hhi = share.multiply(share, "ms_squared", "m_share")
   .sum(“hhi", on="ms_squared")
   .divide("hhi", by=10k)
Integrating Conclave into the cloud

- Conclave runs in containers within each silo
- OpenShift/Kubernetes orchestrate the execution of Conclave jobs
- MPC jobs begin when data analysts make queries over aggregate data
- Benefit: improved performance of secure computing via co-location
How we arrived here

- Both projects have common technology stack
- OpenShift
  - Isolation Techniques
  - Scaled job framework
  - Resource Management
    - CPU/Memory/Network/GPU
How MPC fit in Health Care?

- When data is scarce
  - Boston Trauma Center
  - Understanding rare diseases
- Sharing patient data in the clear is restricted
  - Privacy laws
  - Hospital standard practices
Sharing Information w/o Sharing Data

- Augment ChRIS with cryptographically secure Multi Party Computation
- OpenShift enabled isolated computing environment
Isolated Computing Environment

- **What?**
  - Run individual computations on pre configured secure nodes
  - Container segregation on a host level
Isolated Computing Environment in Cloud

- **How?**
  - Machine like virtualisation
    - Namespace
    - SeLinux
  - Project Isolation
  - Network Isolation
MPC Application in Health Care - Example

- Analyse brain segment volume
- Extrapolates the patient's brain volume against the population mean
- Identify segments that have significant deviation from the population mean in terms of brain volume for the same age group
Medical Image Processor

- MPC (C2D)
  - Calculate population mean
  - Calculate population standard deviation

- Non MPC (ChRIS)
  - Project patient’s brain volume against the population mean
  - Number of standard deviations from the mean the patient datapoint is
MPC job in C2D

- Calculate population mean
- Calculate population standard deviation

Base64 encoded linear algebra query

```
protocol: {
  data: "eW1wb3J6IGVmbWxvXZU\nMnxhbmccgYXMgYzMKZnJvSB\nb25jbGF\n\nconfig: {
  ID: "brain-volume-density",
  backend: "swift"
}

swift: {
  endpoints:
  "partyId": "BCH",
  "containerName": "bch-swift-bucket",
  "fileName": "in1",
  "files": []
}

"partyId": "MGH",
  "containerName": "mgh-swift-bucket",
  "fileName": "in2",
  "files": []
```
MPC job in C2D

OpenShift enabled isolated computing environment.
Non MPC job in ChRIS

- Extrapolates the patient’s brain volume against the population mean of different age group
Non MPC job in ChRIS

- Number of standard deviations from the mean a datapoint is

\[ z = \frac{x - \mu}{\sigma} \]

- \( \mu \): Population Mean
- \( \sigma \): Standard Deviation
- \( x \): Data value/ Score
The Integration of ChRIS & C2D
The road ahead

- Collaboration when data is scarce
- MPC can make a huge impact in the medical landscape
- OpenShift provides the necessary features to build secure ecosystem for this collaboration
- MPC + OpenShift together can scale to big data
Questions?