Cerebro: A Platform for Cryptographic Collaborative Learning

Ryan Deng
Joint work with Wenting Zheng, Weikeng Chen, Aurojit Panda, Raluca Ada Popa, Ion Stoica
CMU, UC Berkeley, NYU
Anti Money Laundering (AML)
Anti Money Laundering (AML)

• Bank wants to detect and prevent money laundering using machine learning.
Anti Money Laundering (AML)

• Bank wants to detect and prevent money laundering using machine learning.

• Scotiabank CRO: “getting AML right is of **critical strategic importance** to our bank.”
Anti Money Laundering (AML)

• Bank wants to detect and prevent money laundering using machine learning.

• Scotiabank CRO: “getting AML right is of critical strategic importance to our bank.”

• Criminals conceal illegal activities across many banks.
Anti Money Laundering (AML)

• Bank wants to detect and prevent money laundering using machine learning.

• Scotiabank CRO: “getting AML right is of critical strategic importance to our bank.”

• Criminals conceal illegal activities across many banks.
Anti Money Laundering (AML)

• Bank wants to detect and prevent money laundering using machine learning.

• Scotiabank CRO: “getting AML right is of critical strategic importance to our bank.”

• Criminals conceal illegal activities across many banks.
Anti Money Laundering
Anti Money Laundering

• Want to jointly compute a model for detecting money laundering across many banks.
Anti Money Laundering

• Want to jointly compute a model for detecting money laundering across many banks.

• Scotiabank CRO: “So in the future, collaboration will be vital … The ability to put together our data sets and collaborate … will enhance yields by orders of magnitude.”
Anti Money Laundering

• Want to jointly compute a model for detecting money laundering across many banks.

• Scotiabank CRO: “So in the future, collaboration will be vital … The ability to put together our data sets and collaborate … will enhance yields by orders of magnitude.”
• Want to jointly compute a model for detecting money laundering across many banks.

• Scotiabank CRO: “So in the future, collaboration will be vital … The ability to put together our data sets and collaborate … will enhance yields by orders of magnitude.”

• Cannot share data due to privacy concerns, regulatory policies, and business competition.
Anti Money-Laundering

Flu prediction

Cloud-based video analysis for home intrusion
Anti Money-Laundering

Flu prediction

Cloud-based video analysis for home intrusion
People are unwilling or unable to share sensitive data for collaborative computation
How to allow organizations to share data for collaboration without showing the plaintext data?
Secure multiparty computation

(MPC\cite{Yao82,GMW87,BGW88})
Secure multiparty computation (MPC)\cite{Yao82,GMW87,BGW88}

- Parties emulate a trusted third party via cryptography
Secure multiparty computation (MPC\cite{Yao82,GMW87,BGW88})

- Parties emulate a trusted third party via cryptography
Secure multiparty computation (MPC) [Yao82, GMW87, BGW88]

- Parties emulate a trusted third party via cryptography
- No party learns any information beyond the final result
Challenges in Deploying MPC End-to-End
Challenges in Deploying MPC End-to-End

• Generality vs. Performance
Challenges in Deploying MPC End-to-End

- Generality vs. Performance
- Privacy vs. Transparency
<table>
<thead>
<tr>
<th>Generality vs. Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Generality vs. Performance</td>
</tr>
<tr>
<td>----------------------------</td>
</tr>
<tr>
<td>Option 1: Specialized Solution</td>
</tr>
<tr>
<td>Generality vs. Performance</td>
</tr>
<tr>
<td>-----------------------------</td>
</tr>
<tr>
<td>Option 1: Specialized Solution</td>
</tr>
<tr>
<td>Generality vs. Performance</td>
</tr>
<tr>
<td>-----------------------------</td>
</tr>
<tr>
<td><strong>Option 1: Specialized Solution</strong></td>
</tr>
</tbody>
</table>
| Use a hand-tuned MPC protocol for a specific learning task. | }
<table>
<thead>
<tr>
<th>Generality vs. Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Option 1: Specialized Solution</strong></td>
</tr>
<tr>
<td>Use a hand-tuned MPC protocol for a specific learning task.</td>
</tr>
<tr>
<td>SecureML [MZ17]</td>
</tr>
<tr>
<td>Gazelle [JVC18]</td>
</tr>
<tr>
<td>XONN [RSCLLK19]</td>
</tr>
<tr>
<td>Spindle [FTPSSBH21]</td>
</tr>
<tr>
<td>Privacy-preserving ridge regression [NWIJBT13]</td>
</tr>
</tbody>
</table>
# Generality vs. Performance

<table>
<thead>
<tr>
<th>Option 1: Specialized Solution</th>
<th>Option 2: Generic Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use a hand-tuned MPC protocol for a specific learning task.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Generality vs. Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Option 1: Specialized Solution</strong></td>
</tr>
<tr>
<td>Use a hand-tuned MPC protocol for a specific learning task.</td>
</tr>
<tr>
<td>✓ Can be very efficient</td>
</tr>
<tr>
<td>Generality vs. Performance</td>
</tr>
<tr>
<td>---------------------------</td>
</tr>
<tr>
<td><strong>Option 1: Specialized Solution</strong></td>
</tr>
<tr>
<td>Use a hand-tuned MPC protocol for a specific learning task.</td>
</tr>
<tr>
<td>✓ Can be very efficient</td>
</tr>
<tr>
<td>✗ Not generalizable</td>
</tr>
<tr>
<td>Generality vs. Performance</td>
</tr>
<tr>
<td>----------------------------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Generality vs. Performance</td>
</tr>
<tr>
<td>-----------------------------</td>
</tr>
<tr>
<td><strong>Option 1: Specialized Solution</strong></td>
</tr>
<tr>
<td>Use a hand-tuned MPC protocol for a specific learning task.</td>
</tr>
<tr>
<td>Can be very efficient</td>
</tr>
</tbody>
</table>
| Not generalizable | }
<table>
<thead>
<tr>
<th>Generality vs. Performance</th>
<th>Option 1: Specialized Solution</th>
<th>Option 2: Generic Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Use a hand-tuned MPC protocol for a specific learning task.</td>
<td>Utilize a generic MPC framework.</td>
</tr>
<tr>
<td></td>
<td>Can be very efficient</td>
<td>SPDZ [DPSZ11]</td>
</tr>
<tr>
<td></td>
<td>Not generalizable</td>
<td>AG-MPC [WKR17]</td>
</tr>
</tbody>
</table>
### Generality vs. Performance

<table>
<thead>
<tr>
<th>Option 1: Specialized Solution</th>
<th>Option 2: Generic Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use a hand-tuned MPC protocol for a specific learning task.</td>
<td>Utilize a generic MPC framework.</td>
</tr>
<tr>
<td>Can be very efficient</td>
<td>MASCOT [KOS16]</td>
</tr>
<tr>
<td>Not generalizable</td>
<td>Vectorized [MZ17]</td>
</tr>
<tr>
<td></td>
<td>Overdrive [KPR18]</td>
</tr>
<tr>
<td>Generality vs. Performance</td>
<td></td>
</tr>
<tr>
<td>----------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Option 1: Specialized Solution</strong></td>
<td><strong>Option 2: Generic Solution</strong></td>
</tr>
<tr>
<td>Use a hand-tuned MPC protocol for a specific learning task.</td>
<td>Utilize a generic MPC framework.</td>
</tr>
<tr>
<td>- Can be very efficient</td>
<td></td>
</tr>
<tr>
<td>- Not generalizable</td>
<td></td>
</tr>
</tbody>
</table>
## Generality vs. Performance

<table>
<thead>
<tr>
<th>Option 1: Specialized Solution</th>
<th>Option 2: Generic Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use a hand-tuned MPC protocol for a specific learning task.</td>
<td>Utilize a generic MPC framework.</td>
</tr>
<tr>
<td>- Can be very efficient</td>
<td>- Generic functionality</td>
</tr>
<tr>
<td>- Not generalizable</td>
<td></td>
</tr>
<tr>
<td>Generality vs. Performance</td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Option 1: Specialized Solution</strong></td>
<td><strong>Option 2: Generic Solution</strong></td>
</tr>
<tr>
<td>Use a hand-tuned MPC protocol for a specific learning task.</td>
<td>Utilize a generic MPC framework.</td>
</tr>
<tr>
<td>✅ Can be very efficient</td>
<td>✅ Generic functionality</td>
</tr>
<tr>
<td>❌ Not generalizable</td>
<td>❌ Difficult to use efficiently</td>
</tr>
</tbody>
</table>
Challenges in Deploying MPC End-to-End

• Generality vs. Performance
Challenges in Deploying MPC End-to-End

- Generality vs. Performance
- Privacy vs. Transparency
Challenges in Deploying MPC End-to-End

- Generality vs. Performance
- Privacy vs. Transparency
- Policies
Challenges in Deploying MPC End-to-End

• Generality vs. Performance
• Privacy vs. Transparency
• Policies
• Auditing
Before Release: Policies
Before Release: Policies

- What if my competitors benefit more than me?
Before Release: Policies

- What if my competitors benefit more than me?
Before Release: Policies

• What if my competitors benefit more than me?
Before Release: Policies

- What if my competitors benefit more than me?
Before Release: Policies

- What if my competitors benefit more than me?
Before Release: Policies

- What if my competitors benefit more than me?
Before Release: Policies

- What if my competitors benefit more than me?
Before Release: Policies

- What if my competitors benefit more than me?
Before Release: Policies

- What if my competitors benefit more than me?
Before Release: Policies

• What if my competitors benefit more than me?
Before Release: Policies

• What if my competitors benefit more than me?
Before Release: Policies

- What if my competitors benefit significantly more than me?
Before Release: Policies

- What if my competitors benefit significantly more than me?
- What happens if the released model leaks information about the underlying training data?
Before Release: Policies

• What if my competitors benefit significantly more than me?

• What happens if the released model leaks information about the underlying training data?

• What happens if the released model does not meet a minimum performance threshold?
After Release: Auditing

77%

77%

77%
After Release: Auditing

77%
After Release: Auditing
After Release: Auditing
After Release: Auditing
After Release: Auditing

77%

77%

STOP

77%

Not a Stop Sign

77%
After Release: Auditing
After Release: Auditing

77%
How do we trace bad data to the source?
Cerebro
Cerebro

End-to-end platform for secure collaborative learning
Cerebro

End-to-end platform for secure collaborative learning

Threat Model
Cerebro

End-to-end platform for secure collaborative learning

Threat Model

- N-party dishonest majority.
Cerebro

End-to-end platform for secure collaborative learning

Threat Model

• N-party dishonest majority.

• Protects against semi-honest and malicious adversaries.
Challenges
Challenges

Generality vs. Performance
Challenges

Generality vs. Performance

Privacy vs. Transparency
Cerebro’s Techniques

Generality vs. Performance

Domain Specific Language (DSL)

MPC Compiler

Privacy vs. Transparency

Release Policies

Auditing Framework
Cerebro’s Techniques

Generality vs. Performance

Domain Specific Language (DSL)

MPC Compiler

Privacy vs. Transparency

Release Policies

Auditing Framework
Cerebro’s MPC Compiler

MPC Compiler
Cerebro’s MPC Compiler

MPC Compiler → Program

Q → Program
Cerebro’s MPC Compiler
Cerebro’s MPC Compiler

MPC Compiler

Program

Q

Local precompute

Q1
Cerebro’s MPC Compiler

MPC Compiler

Program

Q

Q1

Local precompute

Q2

Global compute
Cerebro’s MPC Compiler

MPC Compiler

Program

Q

Local precompute

Q1

Global compute

Q2

Optimizing based on network layout
Cerebro’s MPC Compiler

MPC Compiler

Program

Local precompute

Q1

Global compute

Q2

Optimizing based on network layout

vs.
Cerebro’s MPC Compiler

MPC Compiler

Program

Local precompute

Q1

Global compute

Q2

Optimizing based on network layout

vs.

vs.
Optimizing based on physical layout

Example: Semi-Honest Triple Generation [DPSZ11]
Optimizing based on physical layout

Example: Semi-Honest Triple Generation [DPSZ11]
Optimizing based on physical layout

Example: Semi-Honest Triple Generation [DPSZ11]
Optimizing based on physical layout

Example: Semi-Honest Triple Generation [DPSZ11]
Optimizing based on physical layout

Example: Semi-Honest Triple Generation [DPSZ11]
Optimizing based on physical layout

Example: Semi-Honest Triple Generation [DPSZ11]
Optimizing based on physical layout

Example: Semi-Honest Triple Generation [DPSZ11]
Optimizing based on physical layout

Example: Semi-Honest Triple Generation [DPSZ11]

\[ \text{Enc}(r_1 + r_2 + r_3 + r_4) = \]
Optimizing based on physical layout

Example: Semi-Honest Triple Generation [DPSZ11]

$$\text{Enc}(r_1 + r_2 + r_3 + r_4) = \text{Enc}\left(\sum_{i=1}^{4} r_i\right)$$
Optimizing based on physical layout

Example: Semi-Honest Triple Generation [DPSZ11]
Optimizing based on physical layout

\[ \text{Enc}(r_1) \]
\[ \text{Enc}(r_2) \]
\[ \text{Enc}(r_3) \]
\[ \text{Enc}(r_4) \]
Optimizing based on physical layout

Seattle $\text{Enc}(r_2)$

Los Angeles $\text{Enc}(r_3)$

Aggregator

London $\text{Enc}(r_1)$

Paris $\text{Enc}(r_4)$
Optimizing based on physical layout

Seattle: \(\text{Enc}(r_2)\)

Los Angeles: \(\text{Enc}(r_3)\)

London: \(\text{Enc}(r_1)\)

Paris: \(\text{Enc}(r_4)\)

Aggregator

2Gbps
Optimizing based on physical layout

Seattle

Enc(\(r_2\))

Enc(\(r_3\))

Los Angeles

2Gbps

Enc(\(r_1\))

London

Enc(\(r_4\))

Paris

Aggregator

2Gbps
Optimizing based on physical layout

Seattle

Enc(r_2)

Aggregator

2Gbps

Los Angeles

Enc(r_3)

Enc(r_1)

London

Enc(r_4)

Paris

100Mbps

100Mbps
Optimizing based on physical layout

Seattle

Enc($r_3$)

Los Angeles

Enc($r_1$) Enc($r_2$)

Enc($r_4$)

London

Paris

2Gbps

100Mbps

2Gbps
Optimizing based on physical layout

Seattle

Enc$(r_1)$

Enc$(r_2)$

Enc$(r_3)$

Los Angeles

100Mbps

Enc$(r_4)$

London

2Gbps

Paris

2Gbps
Optimizing based on physical layout

Seattle

Los Angeles

2Gbps

Aggregator

Enc\(r_1\)

Enc\(r_2\)

Enc\(r_3\)

Enc\(r_4\)

London

2Gbps

100Mbps

Paris
Optimizing based on physical layout

Regional Aggregator

Seattle

Enc\((r_2)\)

2Gbps

Los Angeles

Enc\((r_3)\)

Regional/Global Aggregator

London

Enc\((r_1)\)

Paris

Enc\((r_4)\)

100Mbps
Optimizing based on physical layout

Regional Aggregator

Seattle

Enc(r_2)

Enc(r_3)

2Gbps

Regional/Global Aggregator

London

Enc(r_1)

Los Angeles

Enc(r_1)

Enc(r_4)

Paris

100Mbps
Optimizing based on physical layout

Seattle

Enc(r_2)
Enc(r_3)

Los Angeles

100Mbps

Regional Aggregator

Enc(r_1)
Enc(r_4)

Regional/Global Aggregator

London

Paris
Optimizing based on physical layout

\[ \text{Enc}(r_2 + r_3) \]

\[ \text{Enc}(r_1 + r_4) \]

2Gbps

100Mbps

Los Angeles

Seattle

Regional Aggregator

Regional/Global Aggregator

Paris

London
Optimizing based on physical layout

Regional Aggregator
Seattle
2Gbps

Regional/Global Aggregator
London
2Gbps

Regional/Global Aggregator
Paris
2Gbps

Regional Aggregator
Los Angeles
2Gbps

Enc(r_1 + r_4)
Enc(r_2 + r_3)

100Mbps
Optimizing based on physical layout

Regional Aggregator

Seattle

2Gbps

Regional/Global Aggregator

Los Angeles

2Gbps

Enc \left( \sum_{i=1}^{4} r_i \right)

London

2Gbps

Enc \left( \sum_{i=1}^{4} r_i \right)

Paris

100Mbps
Optimizing based on physical layout

2Gbps

Regional Aggregator

Enc(\sum_{i=1}^{4} r_i)

Seattle

2Gbps

Regional/Global Aggregator

Enc(\sum_{i=1}^{4} r_i)

Los Angeles

100Mbps

Enc(\sum_{i=1}^{4} r_i)

Paris

Enc(\sum_{i=1}^{4} r_i)
Optimizing based on physical layout

Regional Aggregator

Seattle

2Gbps

Enc\left(\sum_{i=1}^{4} r_i\right)

Regional/Global Aggregator

London

2Gbps

Enc\left(\sum_{i=1}^{4} r_i\right)

Los Angeles

Enc\left(\sum_{i=1}^{4} r_i\right)

Enc\left(\sum_{i=1}^{4} r_i\right)

Enc\left(\sum_{i=1}^{4} r_i\right)

100Mbps

Enc\left(\sum_{i=1}^{4} r_i\right)

Paris
Optimizing based on physical layout

Regional Aggregator

Seattle

2Gbps

Los Angeles

Enc\left(\sum_{i=1}^{4} r_i\right)

Regional/Global Aggregator

London

Enc\left(\sum_{i=1}^{4} r_i\right)

Enc\left(\sum_{i=1}^{4} r_i\right)

Enc\left(\sum_{i=1}^{4} r_i\right)

Enc\left(\sum_{i=1}^{4} r_i\right)

100Mbps

Enc\left(\sum_{i=1}^{4} r_i\right)

Enc\left(\sum_{i=1}^{4} r_i\right)

Enc\left(\sum_{i=1}^{4} r_i\right)

Enc\left(\sum_{i=1}^{4} r_i\right)
Evaluation
Impact of Cerebro’s Physical Layout Optimizations
Impact of Cerebro’s Physical Layout Optimizations

• 12 parties, 2 regions
Impact of Cerebro’s Physical Layout Optimizations

- 12 parties, 2 regions
- 9-3 split
Impact of Cerebro’s Physical Layout Optimizations

• 12 parties, 2 regions
  • 9-3 split
  • 2Gbps connection within each region, varied total bandwidth across regions.
Impact of Cerebro’s Physical Layout Optimizations

- 12 parties, 2 regions
- 9-3 split
- 2Gbps connection within each region, varied total bandwidth across regions.
Impact of Cerebro’s Physical Layout Optimizations

- 12 parties, 2 regions
- 9-3 split
- 2Gbps connection within each region, varied total bandwidth across regions.
Impact of Cerebro’s Physical Layout Optimizations
Impact of Cerebro’s Physical Layout Optimizations

- Flat
- 2-Layer

# regular mult/s

- 0
- 100
- 17500
- 35000
- 52500
- 70000

cross-region bandwidth (Mbps)

- 100
- 300
- 600
- 900
- 1200
Impact of Cerebro’s Physical Layout Optimizations

![Graph showing the impact of Cerebro’s Physical Layout Optimizations with two layers compared to flat layout. The x-axis represents cross-region bandwidth (Mbps) ranging from 0 to 1200 Mbps, and the y-axis represents the number of regular multiplies (mul/s) ranging from 0 to 70000.]
Impact of Cerebro’s Physical Layout Optimizations

![Impact of Cerebro's Physical Layout Optimizations](image_url)
Impact of Cerebro’s Physical Layout Optimizations

- Regular mult/s
- Cross-region bandwidth (Mbps)
- Flat
- 2-Layer

Up to 3.5x performance improvement
Effectiveness of Planner Decision Making
Effectiveness of Planner Decision Making

- Performance improvement of Cerebro’s physical planning when compared to the worst case secure plan generated without Cerebro.
Effectiveness of Planner Decision Making
Effectiveness of Planner Decision Making

- Arithmetic
- Boolean
- Planner

10-layer decision tree inference

Total time (s)

# parties in 2Gbps network
Effectiveness of Planner Decision Making

- Arithmetic
- Boolean
- Planner

Number of parties in 2Gbps network vs. total time (s)

- 10-layer decision tree inference
Effectiveness of Planner Decision Making

- Arithmetic
- Boolean
- Planner

# parties in 2Gbps network
10-layer decision tree inference
Effectiveness of Planner Decision Making

![Graph showing the relationship between the number of parties in a 2Gbps network and the total time (s) for 10-layer decision tree inference. The graph compares arithmetic, boolean, and planner decision making.](image-url)
Effectiveness of Planner Decision Making

- Arithmetic
- Boolean
- Planner

# parties in 2Gbps network

10-layer decision tree inference

Graph showing the total time (s) vs. the number of parties in a 2Gbps network for different decision making methods.
Effectiveness of Planner Decision Making

- **Arithmetic**
- **Boolean**
- **Planner**

**# parties in 2Gbps network**

**10-layer decision tree inference**

**total time (s)**
Effectiveness of Planner Decision Making

![Graph showing total time (s) vs. # parties in 2Gbps network. The graph includes three lines: Arithmetic (pink), Planner (black), and Boolean (blue). The x-axis represents the number of parties in the 2Gbps network, ranging from 2 to 12. The y-axis represents the total time in seconds, ranging from 0 to 40. The graph illustrates the impact of the number of parties on the total time for 10-layer decision tree inference.](image)
Effectiveness of Planner Decision Making

10-layer decision tree inference

# parties in 2Gbps network

total time (s)
Effectiveness of Planner Decision Making

- **Arithmetic**
- **Boolean**
- **Planner**

Total time (s)

# parties in 2Gbps network

10-layer decision tree inference

- 5.77x
- 1.48x
Conclusion
Conclusion

- Cerebro is a programmable secure collaborative learning platform.
Conclusion

• Cerebro is a programmable secure collaborative learning platform.

• Addresses the challenges that arise when deploying secure multiparty computation in practice.
Conclusion

• Cerebro is a programmable secure collaborative learning platform.

• Addresses the challenges that arise when deploying secure multiparty computation in practice.

• Open source: https://github.com/mc2-project/cerebro
Conclusion

• Cerebro is a programmable secure collaborative learning platform.

• Addresses the challenges that arise when deploying secure multiparty computation in practice.

• Open source: https://github.com/mc2-project/cerebro

• Contact: ryan.deng@berkeley.edu